

Method of Moments Estimation and Identifiability of Semiparametric Nonlinear Errors-in-Variables Models

Liqun Wang

Department of Statistics, University of Manitoba

Winnipeg, Manitoba, Canada R3T 2N2

e-mail: liqun_wang@umanitoba.ca

Cheng Hsiao

Department of Economics, University of Southern California

Los Angeles, CA 90089-0253, USA

February 11, 2009

Abstract

This paper deals with a nonlinear errors-in-variables model where the distributions of the unobserved predictor variables and of the measurement errors are nonparametric. Using the instrumental variable approach, we propose method of moments estimators for the unknown parameters and simulation-based estimators to overcome the possible computational difficulty of minimizing an objective function which involves multiple integrals. Both estimators are consistent and asymptotically normally distributed under fairly general regularity conditions. Moreover, root-n consistent semiparametric estimators and a rank condition for model identifiability are derived using the combined methods of nonparametric technique and Fourier deconvolution.

JEL SUBJECT CLASSIFICATION: C13, C14, C15.

KEYWORDS: Fourier deconvolution, identifiability, instrumental variables, measurement error, method of moments, root-n consistency, semiparametric estimator, simulation-based estimator.

1 Introduction

Measurement error occurs frequently (e.g. Aigner et al. (1984), Fuller (1987) and Hsiao (1992)). If a model is linear in variables, the issue of random measurement error 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 estimators of the unknown

parameters when the variables are subject to random measurement errors (e.g. Y. Amemiya (1985, 1990) and Hsiao (1989)).

To obtain consistent estimators for nonlinear measurement error models, some researchers assume that the measurement error variances tend to zero as sample size increases to infinity (e.g., and Wolter and Fuller (1982), Amemiya (1985, 1990), Stefanski and Carroll (1985), and Amemiya and Fuller (1988)). Alternatively, other researchers assume that the conditional distribution of the unobserved predictor variable given its observed proxy is known up to a finite-dimensional parameter (e.g. Hsiao (1989, 1992)). Recently Li (2002) and Schennach (2004) studied models with replicate observations, while Schennach (2007) used instrumental variable approach. Besides, various special nonlinear models are investigated, e.g., polynomial models with a scalar predictor variable (Cheng and Schneeweiss (1998), Hausman et al. (1991), Hausman, Newey and Powell (1995), Huang and Huwang (2001)), and limited dependent variable models (Weiss (1993), Wang (1998, 2002) and Wang and Hsiao (2007)). Another stream of investigation consists of non- or semi-parametric methods with the assumption that the measurement error is univariate and its distribution is either completely known or is normal with an unknown variance parameter (e.g., Fan and Truong (1993), Carroll, Maca and Ruppert (1999), Taupin (2001), Carroll et al. (2004), and Delaigle (2007)).

In this paper, we consider the method of moments estimation of general nonlinear measurement error model. Specifically, we consider the model

$$Y = g(X; \theta_0) + \varepsilon, \quad (1.1)$$

where $Y \in \mathbb{R}$, $X \in \mathbb{R}^k$, ε is the random error and $\theta_0 \in \mathbb{R}^p$ is a vector of unknown parameters. In general, $g(x; \theta_0)$ is nonlinear in x . Suppose that X is unobservable, instead we observe

$$Z = X + \delta, \quad (1.2)$$

where δ is a random measurement error. Further, we assume that an instrumental variable $W \in \mathbb{R}^l$ exists and is related to X through

$$X = \Gamma_0 W + U, \quad (1.3)$$

where Γ_0 is a $k \times l$ matrix of unknown parameters which has rank k and U is independent of W with $E(U) = 0$. The random errors in (1.1) and (1.2) are supposed to satisfy $E(\varepsilon | X, Z, W) = 0$ and $E(\delta | X, W) = 0$. The functional forms of the distributions of X , ε and δ are unknown. In this sense model (1.1)-(1.3) is semiparametric. In this model, the observed variables are (Y, Z, W) . Our primary interest is to estimate θ_0 , Γ_0 and the distribution F_U of U .

Model (1.1) - (1.1) was considered by these and other authors before. Wang and Hsiao (1995) derived a rank condition for identifiability and proposed a semiparametric estimator under the condition that $g(x; \theta_0)$ is integrable. Later, this integrability condition was relaxed by Schennach (2007) who used generalized function technique and achieved more general identifiability conditions. In addition, assuming the model to be identifiable, Newey (2001) derived a consistent estimator

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. In this paper, we use the approach of Wang and Hsiao (1995) and extend their results to general $g(x; \theta_0)$ which is not necessarily integrable.

In particular, for the case of a parametric distribution $f_U(u; \phi_0)$ we propose method of moments estimators for θ and ϕ which are shown to be consistent and asymptotically normally distributed under fairly general regularity conditions. Simulation-based estimators are also considered to overcome the possible computational difficulty of minimizing an objective function which involves multiple integrals. For the case of nonparametric distribution $F_U(u)$, we combine the nonparametric technique with Fourier deconvolution to obtain a root-n consistent estimator for θ and a kernel-based estimator for the density of U . Moreover, this approach results in a surprisingly simple condition for the identifiability of a nonlinear errors-in-variables model.

The paper is organized as follows. In Section 2 we introduce the method of moments estimators and derive their consistency and asymptotic normality. In Section 3 we construct simulation-based estimators and show their asymptotic properties. In Section 4 we propose a nonparametric estimator for the density of U and a root-n consistent semiparametric estimator for θ . Section 5 gives a rank condition for model identifiability and illustrative examples. Finally, conclusions and discussion are contained in Section 6, whereas proofs are given in Section 7.

2 Method of Moments Estimator

In this section we propose a method of moments estimator for nonlinear errors-in-variables model under the assumption that the distribution $f_U(u; \phi_0)$ of U is known up to a vector of unknown parameters $\phi_0 \in \Phi \subset \mathbb{R}^q$. The case where the distribution of U is nonparametric is treated in Sections 4 and 5.

First, substituting (1.3) into (1.2) results in a usual linear regression equation

$$E(Z | W) = \Gamma_0 W. \quad (2.1)$$

Therefore Γ_0 can be consistently estimated by the least squares fitting of Z on W . Moreover, by model assumptions we have

$$E(Y | W) = \int g(\Gamma_0 W + u; \theta_0) f_U(u; \phi_0) du \quad (2.2)$$

and

$$E(YZ | W) = \int (\Gamma_0 W + u) g(\Gamma_0 W + u; \theta_0) f_U(u; \phi_0) du. \quad (2.3)$$

Throughout the paper, all integrals are taken over the space \mathbb{R}^k . It follows that θ_0 and ϕ_0 can be estimated using nonlinear least squares method, given that they are identifiable by (2.2) and (2.3).

Since it is straightforward to estimate Γ_0 , in the following we focus on the estimation of θ_0 and ϕ_0 . First, we use some examples to demonstrate that θ_0 and ϕ_0 may indeed be estimated using

(2.2) and (2.3). To simplify notation, we consider the case where $\Gamma_0 = 1$, all variables are scalars and $U \sim N(0, \phi)$. For the same reason, we suppress the subscript zero in θ_0 and denote it as θ .

EXAMPLE 2.1 Linear model $g(x; \theta) = \theta x$. For this model, it is easy to find $E(Y | W) = \theta W$ and $E(YZ | W) = \theta\phi + \theta W^2$, from which both θ and ϕ can be consistently estimated by nonlinear least squares method.

EXAMPLE 2.2 Polynomial model $g(x; \theta) = \theta_1 + \theta_2 x^2$. In this case, we have $E(Y | W) = (\theta_1 + \theta_2\phi) + \theta_2 W^2$ and $E(YZ | W) = (\theta_1 + 3\theta_2\phi)W + \theta_2 W^3$. Again, it is clear that θ_2 , $\theta_1 + \theta_2\phi$ and $\theta_1 + 3\theta_2\phi$ can be consistently estimated and, therefore, so do θ_1 and ϕ .

EXAMPLE 2.3 Exponential model $g(x; \theta) = \theta_1 \exp(\theta_2 x)$, where $\theta_1 \theta_2 \neq 0$. For this model, we have $E(Y | W) = \theta_1 \exp(\theta_2 W + \theta_2^2 \phi / 2)$ and $E(YZ | W) = \theta_1 (\theta_2 \phi + W) \exp(\theta_2 W + \theta_2^2 \phi / 2)$. Now θ_2 and $\theta_1 \exp(\theta_2^2 \phi / 2)$ can be consistently estimated from the first equation, and $\theta_1 \theta_2 \phi \exp(\theta_2^2 \phi / 2)$ from the second. It follows that θ_1 and ϕ can be consistently estimated too.

Let $\psi = (\theta', \phi)'$ and $\Psi = \Theta \times \Phi$, which is assumed to be compact in \mathbb{R}^{p+q} . The true parameter value of the model is denoted by $\psi_0 \in \Psi$. To simplify notation, let $\tilde{Z} = (1, Z)'$ and $\tilde{x} = (1, x)'$. Then through variable substitution, (2.2) and (2.3) can be written together as

$$E(Y\tilde{Z}|W) = \int \tilde{x}g(x; \theta_0)f_U(x - \Gamma_0 W; \phi_0)dx. \quad (2.4)$$

For every $v \in \mathbb{R}^k$ and $\psi \in \Psi$, define

$$m(v; \psi) = \int \tilde{x}g(x; \theta)f_U(x - v; \phi)dx. \quad (2.5)$$

Then it is clear that $m(\Gamma_0 W; \psi_0) = E(Y\tilde{Z}|W)$.

Suppose $(Y_j, Z_j, W_j), j = 1, 2, \dots, n$, is an *i.i.d.* random sample with finite moments $EY^2 < \infty$, $E\|YZ\|^2 < \infty$ and nonsingular $EW W'$, where $\|\cdot\|$ denotes the Euclidian norm. Further, let $\hat{\rho}_j(\psi) = Y_j \tilde{Z}_j - m(\hat{\Gamma} W_j; \psi)$, where

$$\hat{\Gamma} = \left(\sum_{j=1}^n Z_j W_j' \right) \left(\sum_{j=1}^n W_j W_j' \right)^{-1} \quad (2.6)$$

is the least squares estimator of Γ_0 . Then the method of moments estimator (MME) for ψ is defined as $\hat{\psi}_n = \operatorname{argmin}_{\psi \in \Psi} Q_n(\psi)$, where

$$Q_n(\psi) = \sum_{j=1}^n \hat{\rho}_j'(\psi) A_j \hat{\rho}_j(\psi), \quad (2.7)$$

and $A_j = A(W_j)$ is a nonnegative definite matrix which may depend on W_j .

Throughout the paper, let $\gamma = \operatorname{vec}\Gamma$ denote the vector consisting of the columns of Γ , where vec is the so-called vectorization operator. We also assume that the parameter space of γ is a compact

subset of \mathbb{R}^{kl} containing the true value $\gamma_0 = \text{vec}\Gamma_0$. The consistency of $\hat{\theta}_n$ can be derived in traditional fashion by establishing the uniform convergence of $Q_n(\psi)/n$ to a nonstochastic function $Q(\psi)$ which has unique minimizer $\psi_0 \in \Psi$. To achieve this, we assume the following regularity conditions, where μ denotes Lebesgue measure.

ASSUMPTION 1 $g(x; \theta)$ is a measurable function of x for each $\theta \in \Theta$ and is continuous in $\theta \in \Theta$ (a.e. μ). $f_U(u; \phi)$ is continuously differentiable with respect to (w.r.t.) u for each $\phi \in \Phi$ and is continuous in $\phi \in \Phi$ (a.e. μ). Furthermore, $E \|A(W)\| (|Y|^2 + \|YZ\|^2) < \infty$,

$$E \|A(W)\| \left(\int \sup_{\psi, \gamma} |g(x; \theta) f_U(x - \Gamma W; \phi)| (\|x\| + 1) dx \right)^2 < \infty \quad (2.8)$$

and

$$E \|A(W)\| \left(\|W\| \int \sup_{\psi, \gamma} \left\| g(x; \theta) \frac{\partial f_U(x - \Gamma W; \phi)}{\partial u'} \right\| (\|x\| + 1) dx \right)^2 < \infty, \quad (2.9)$$

where the supremum is taken within the compact parameter spaces of ψ and γ .

ASSUMPTION 2 $E[\rho(\psi) - \rho(\psi_0)]' A(W) [\rho(\psi) - \rho(\psi_0)] = 0$ if and only if $\psi = \psi_0$, where $\rho(\psi) = Y\tilde{Z} - m(\Gamma_0 W; \psi)$.

ASSUMPTIONS 1 and 2 are common in the literature of nonlinear regression. ASSUMPTION 2 is a high level condition for identifiability. Some sufficient conditions for the identifiability are given in Section 5. ASSUMPTION 1 ensures that the objective function $Q_n(\psi)$ is continuous and converges uniformly in ψ . The following example shows that (2.8) and (2.9) are generally satisfied, e.g., when $g(x; \theta)$ is a polynomial in x and U has a normal distribution.

EXAMPLE 2.4 Suppose $g(x; \theta) = \theta x$, $U \sim N(0, \phi)$, all variables are scalars and the parameter spaces are compact intervals $[\theta_{min}, \theta_{max}]$, $[\phi_{min}, \phi_{max}]$ and $[\Gamma_{min}, \Gamma_{max}]$. Then, for every $x, w \in \mathbb{R}$ and θ, ϕ, Γ in their respective parameter spaces,

$$\begin{aligned} |g(x; \theta) f_U(x - \Gamma w; \phi)| (\|x\| + 1) &= \left| \frac{\theta x}{\sqrt{2\pi\phi}} \exp\left(-\frac{(x - \Gamma w)^2}{2\phi}\right) \right| (\|x\| + 1) \\ &\leq \frac{\theta_{max} |x| (\|x\| + 1)}{\sqrt{2\pi\phi_{min}}} \exp\left(-\frac{x^2 + \Gamma_{min}^2 w^2}{2\phi_{max}}\right) \left[\exp\left(\frac{\Gamma_{max} x w}{\phi_{max}}\right) + \exp\left(\frac{\Gamma_{min} x w}{\phi_{max}}\right) \right], \end{aligned}$$

which clearly satisfies (2.8) if, e.g., $A(W)$ is an identity matrix. Similarly, it is easy to see that (2.9) is satisfied too. \square

THEOREM 2.1 Under ASSUMPTIONS 1 - 2, $\hat{\psi}_n \xrightarrow{a.s.} \psi_0$, as $n \rightarrow \infty$.

To derive the asymptotic normality for $\hat{\psi}_n$, we assume further regularity conditions as follows.

ASSUMPTION 3 *There exist open subsets $\theta_0 \in \Theta_0 \subset \Theta$ and $\phi_0 \in \Phi_0 \subset \Phi$, in which $g(x; \theta)$ is twice continuously differentiable w.r.t. θ (a.e. μ) and $f_U(u; \phi)$ is twice continuously differentiable w.r.t. ϕ (a.e. μ). Furthermore, γ_0 has an open neighborhood, such that the first two derivatives of $g(x; \theta)f_U(x - \Gamma w; \phi)$ w.r.t. ψ are uniformly bounded by a function $\eta(x, w)$, which satisfies $E \|A(W)\| \left(\int \eta(x, W)(\|x\| + 1)dx \right)^2 < \infty$.*

ASSUMPTION 4 *For $\psi_0 \in \Psi$,*

$$E \|A(W)\| \left(\|W\| \int_{\sup_{\gamma}} \left\| \frac{\partial f_U(x - \Gamma W; \phi_0)}{\partial u'} \right\| \left\| \frac{\partial g(x; \theta_0)}{\partial \theta} \right\| (\|x\| + 1)dx \right)^2 < \infty \quad (2.10)$$

and

$$E \|A(W)\| \left(\|W\| \int_{\sup_{\gamma}} \left\| \frac{\partial^2 f_U(x - \Gamma W; \phi_0)}{\partial \phi \partial u'} \right\| |g(x; \theta_0)| (\|x\| + 1)dx \right)^2 < \infty, \quad (2.11)$$

where the supremum is taken within the open subset of γ .

ASSUMPTION 5 *The matrix*

$$B = E \left[\frac{\partial \rho'(\psi_0)}{\partial \psi} A(W) \frac{\partial \rho(\psi_0)}{\partial \psi'} \right]$$

is nonsingular, where

$$\frac{\partial \rho(\psi)}{\partial \theta'} = - \int \tilde{x} \frac{\partial g(x; \theta)}{\partial \theta'} f_U(x - \Gamma_0 W; \phi) dx \quad (2.12)$$

and

$$\frac{\partial \rho(\psi)}{\partial \phi'} = - \int \tilde{x} g(x; \theta) \frac{\partial f_U(x - \Gamma_0 W; \phi)}{\partial \phi'} dx. \quad (2.13)$$

Again, ASSUMPTIONS 3 - 5 are commonly used regularity conditions that are sufficient for the asymptotic normality of method of moments estimators. Together with the Dominated Convergence Theorem (DCT), ASSUMPTION 3 implies that the first derivative of $Q_n(\psi)$ admits the first-order Taylor expansion and that the second derivative of $Q_n(\psi)$ converges uniformly. Moreover, it ensures that the first derivative $\partial \rho(\psi)/\partial \psi'$ exists and is given by (2.12) and (2.13), while ASSUMPTION 4 implies that the first derivative $\partial \rho(\psi)/\partial \gamma'$ exists and is given by

$$\frac{\partial \rho(\psi)}{\partial \gamma'} = \int \tilde{x} g(x; \theta) \frac{\partial f_U(x - \Gamma_0 W; \phi)}{\partial u'} dx (W \otimes I_k)', \quad (2.14)$$

where \otimes stands for the Kronecker product (Magnus and Neudecker (1988), p.30). Finally, ASSUMPTION 5 and the DCT imply that the second derivative of $Q_n(\psi)$ has a non-singular limiting matrix. Again, it is easy to see that ASSUMPTIONS 3 - 5 are satisfied for the polynomial model $g(x; \theta)$ and the normal random error U .

THEOREM 2.2 *Under ASSUMPTIONS 1 - 5, as $n \rightarrow \infty$, $\sqrt{n}(\hat{\psi}_n - \psi_0) \xrightarrow{L} N(0, B^{-1}DCD'B^{-1})$, where*

$$D = \left[I_{p+q}, E \left(\frac{\partial \rho'(\psi_0)}{\partial \psi} A(W) \frac{\partial \rho(\psi_0)}{\partial \gamma'} \right) (EWW' \otimes I_k)^{-1} \right],$$

$$C = \begin{pmatrix} C_{11} & C'_{21} \\ C_{21} & C_{22} \end{pmatrix},$$

$$C_{11} = E \left[\frac{\partial \rho'(\psi_0)}{\partial \psi} A(W) \rho(\psi_0) \rho'(\psi_0) A(W) \frac{\partial \rho(\psi_0)}{\partial \psi'} \right],$$

$$C_{21} = E \left[(W \otimes (Z - \Gamma_0 W)) \rho'(\psi_0) A(W) \frac{\partial \rho(\psi_0)}{\partial \psi'} \right]$$

and $C_{22} = E [WW' \otimes (Z - \Gamma_0 W)(Z - \Gamma_0 W)']$. Furthermore,

$$B = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \frac{\partial \hat{\rho}'_j(\hat{\psi}_n)}{\partial \psi} A_j \frac{\partial \hat{\rho}_j(\hat{\psi}_n)}{\partial \psi'}$$

and

$$DCD' = \text{plim}_{n \rightarrow \infty} \frac{1}{4n} \frac{\partial Q_n(\hat{\psi}_n)}{\partial \psi} \frac{\partial Q_n(\hat{\psi}_n)}{\partial \psi'},$$

where

$$\frac{\partial Q_n(\psi)}{\partial \psi} = 2 \sum_{j=1}^n \frac{\partial \hat{\rho}'_j(\psi)}{\partial \psi} A_j \hat{\rho}_j(\psi).$$

The asymptotic covariance of $\hat{\psi}_n$ depends on the weight $A(W)$. A natural question is how to choose $A(W)$ to obtain the most efficient estimator. To answer this question, we first write $D = (I_{p+q}, G)$, so that $DCD' = C_{11} + GC_{21} + C'_{21}G' + GC_{22}G'$. It is easy to see that the last three terms in DCD' are due to the least squares estimation of Γ . To simplify discussion, assume for the moment that Γ_0 is known, so that these three terms do not appear in DCD' . The following discussion remains valid, when Γ_0 is unknown and estimated using a subset of the sample (Y_j, Z_j, W_j) , $j = 1, 2, \dots, n$, while $Q_n(\psi)$ is constructed using the rest of the sample points. Then the independence of the sample points implies that $C_{21} = 0$. Since $\partial \rho'(\psi_0)/\partial \psi$ depends on W only, matrix C_{11} can be written as

$$C_{11} = E \left[\frac{\partial \rho'(\psi_0)}{\partial \psi} A(W) H A(W) \frac{\partial \rho(\psi_0)}{\partial \psi'} \right],$$

where $H = H(W) = E[\rho(\psi_0)\rho'(\psi_0)|W]$. Then, analogous to the weighted (nonlinear) least squares estimation, we have

$$B^{-1}C_{11}B^{-1} \geq E \left[\frac{\partial \rho'(\psi_0)}{\partial \psi} H^{-1} \frac{\partial \rho(\psi_0)}{\partial \psi'} \right]^{-1} \quad (2.15)$$

(in the sense that the difference of the left-hand and right-hand sides is nonnegative definite), and the lower bound is attained for $A(W) = H^{-1}$ in both B and C_{11} (Hansen (1982), Abarin and Wang (2006)).

In practice, however, H depends on unknown parameters and therefore needs to be estimated. This suggests the following two-stage procedure of estimation. First, minimize $Q_n(\psi)$ with identity matrix $A(W) = I_{k+1}$ to obtain the first-stage estimator $\hat{\psi}_n$. Secondly, estimate $H = H(W)$ by a nonparametric method such as kernel estimator or by $\hat{H} = \frac{1}{n} \sum_{j=1}^n \hat{\rho}_j(\hat{\psi}_n) \hat{\rho}'_j(\hat{\psi}_n)$ for models where H does not depend on W , and then minimize $Q_n(\psi)$ again with $A(W) = \hat{H}^{-1}$ to obtain

the second-stage estimator $\hat{\psi}_n$. Since \hat{H} is consistent for H , the asymptotic covariance of $\hat{\psi}_n$ is given by the right-hand side of (2.15). Consequently $\hat{\psi}_n$ is asymptotically more efficient than the first-stage estimator $\hat{\psi}_n$. More detailed discussions about the so-called feasible generalized least squares estimators can be found in, e.g., Amemiya (1974) and Gallant (1987, Chapter 5).

3 Simulation-based Estimator

The numerical computation of MME $\hat{\psi}_n$ or adaptive generalized method of moments estimator is straightforward if the explicit form of $m(v; \psi)$ can be obtained. However, explicit forms of the integrals in (2.5) can be difficult or impossible to derive (for instance, if g is logistic and f_U is normal). In this case, one may use a simulation-based approach to approximate the multiple integrals in which they are simulated by Monte Carlo methods such as importance sampling.

First, choose a known density $h(x)$ and generate an *i.i.d.* random sample $\{x_{js}, s = 1, 2, \dots, 2S, j = 1, 2, \dots, n\}$ from $h(x)$. Then approximate $m(\Gamma W_j; \psi)$ by Monte Carlo simulators

$$m_S(\Gamma W_j; \psi) = \frac{1}{S} \sum_{s=1}^S \frac{\tilde{x}_{js} g(x_{js}; \theta) f_U(x_{js} - \Gamma W_j; \phi)}{h(x_{js})}$$

and

$$m_{2S}(\Gamma W_j; \psi) = \frac{1}{S} \sum_{s=S+1}^{2S} \frac{\tilde{x}_{js} g(x_{js}; \theta) f_U(x_{js} - \Gamma W_j; \phi)}{h(x_{js})},$$

where $\tilde{x}_{js} = (1, x'_{js})'$. Finally, the simulation-based estimator (SBE) for ψ is defined by $\hat{\psi}_{n,S} = \operatorname{argmin}_{\psi \in \Psi} Q_{n,S}(\psi)$, where

$$Q_{n,S}(\psi) = \sum_{j=1}^n \hat{\rho}'_{j,S}(\psi) A_j \hat{\rho}_{j,2S}(\psi), \quad (3.1)$$

and $\hat{\rho}_{j,S}(\psi) = Y_j \tilde{Z}_j - m_S(\hat{\Gamma} W_j; \psi)$ and $\hat{\rho}_{j,2S}(\psi) = Y_j \tilde{Z}_j - m_{2S}(\hat{\Gamma} W_j; \psi)$.

It is easy to see that $m_S(\Gamma W_j; \psi)$ and $m_{2S}(\Gamma W_j; \psi)$ are unbiased simulators for $m(\Gamma W_j; \psi)$, because by construction, $E[m_S(\Gamma W_j; \psi) | W_j] = E[m_{2S}(\Gamma W_j; \psi) | W_j] = m(\Gamma W_j; \psi)$. In addition, using two independent sets of simulated points in $\hat{\rho}_{j,S}$ and $\hat{\rho}_{j,2S}$ guarantees $Q_{n,S}(\psi)$ to be an unbiased simulator for $Q_n(\psi)$ in the sense that they have the same conditional expectation given the data $(Y_j, Z_j, W_j), j = 1, 2, \dots, n$. This "simulation-by-parts" has an important consequence that the following consistency and asymptotic normality of $\hat{\psi}_{n,S}$ hold for a fixed S . In contrast, most simulation-based methods in the literature require that $S \rightarrow \infty$.

Since $Q_{n,S}(\psi)$ does not involve integrals any more, it is continuous in and differentiable with respect to ψ , as long as functions $g(x; \theta)$ and $f_U(u; \phi)$ have these properties. In particular, the first derivative $\partial \rho'_{j,S}(\psi) / \partial \psi$ consists of

$$\frac{\partial \rho'_{j,S}(\psi)}{\partial \theta} = -\frac{1}{S} \sum_{s=1}^S \frac{\partial g(x_{js}; \theta)}{\partial \theta} \frac{\tilde{x}'_{js} f_U(x_{js} - \Gamma W_j; \phi)}{h(x_{js})},$$

$$\frac{\partial \rho'_{j,S}(\psi)}{\partial \phi} = -\frac{1}{S} \sum_{s=1}^S \frac{\partial f_U(x_{js} - \Gamma W_j; \phi)}{\partial \phi} \frac{\tilde{x}'_{js} g(x_{js}; \theta)}{h(x_{js})}$$

and the first derivative $\partial \rho'_{j,2S}(\psi)/\partial \psi$ is given similarly.

THEOREM 3.1 *Suppose that the support of $h(x)$ covers the support of $|g(x; \theta)| f_U(x - v; \phi)$ for all $v \in \mathbb{R}^k$ and $\psi \in \Psi$. Then the simulation estimator $\hat{\psi}_{n,S}$ has the following properties:*

1. Under ASSUMPTIONS 1 - 2, $\hat{\psi}_{n,S} \xrightarrow{a.s.} \psi_0$, as $n \rightarrow \infty$.
2. Under ASSUMPTIONS 1 - 5, $\sqrt{n}(\hat{\psi}_{n,S} - \psi_0) \xrightarrow{L} N(0, B^{-1} D C_S D' B^{-1})$, where

$$C_S = \begin{pmatrix} C_{S,11} & C'_{21} \\ C_{21} & C_{22} \end{pmatrix}$$

and

$$\begin{aligned} 2C_{S,11} &= E \left[\frac{\partial \rho'_{1,S}(\psi_0)}{\partial \psi} A_1 \rho_{1,2S}(\psi_0) \rho'_{1,2S}(\psi_0) A_1 \frac{\partial \rho_{1,S}(\psi_0)}{\partial \psi'} \right] \\ &+ E \left[\frac{\partial \rho'_{1,S}(\psi_0)}{\partial \psi} A_1 \rho_{1,2S}(\psi_0) \rho'_{1,S}(\psi_0) A_1 \frac{\partial \rho_{1,2S}(\psi_0)}{\partial \psi'} \right]. \end{aligned}$$

Furthermore,

$$D C_S D' = \text{plim}_{n \rightarrow \infty} \frac{1}{4n} \frac{\partial Q_{n,S}(\hat{\psi}_n)}{\partial \psi} \frac{\partial Q_{n,S}(\hat{\psi}_n)}{\partial \psi'},$$

where

$$\frac{\partial Q_{n,S}(\psi)}{\partial \psi} = \sum_{j=1}^n \left[\frac{\partial \hat{\rho}'_{j,S}(\psi)}{\partial \psi} A_j \hat{\rho}_{j,2S}(\psi) + \frac{\partial \hat{\rho}'_{j,2S}(\psi)}{\partial \psi} A_j \hat{\rho}_{j,S}(\psi) \right].$$

Although $\hat{\psi}_{n,S}$ is feasible in general, the simulation approximation of $\rho_j(\psi)$ by $\rho_{j,S}(\psi)$ and $\rho_{j,2S}(\psi)$ may cause efficiency loss. The following Corollary shows that the efficiency loss due to simulation is of magnitude $O(1/S)$, the proof of which is completely analogous to that of Corollary 4 of Wang (2004) and hence is omitted.

COROLLARY 3.2 *Under the conditions of Theorem 3.1,*

$$\begin{aligned} C_{S,11} &= C_{11} + \frac{1}{2S} E \left[\frac{\partial(\rho_{11} - \rho_1)' A_1 \rho_1}{\partial \psi} \frac{\partial \rho'_1 A_1 (\rho_{11} - \rho_1)}{\partial \psi'} \right] \\ &+ \frac{1}{4S^2} E \left[\frac{\partial(\rho_{11} - \rho_1)' A_1 (\rho_{12} - \rho_1)}{\partial \psi} \frac{\partial(\rho_{12} - \rho_1)' A_1 (\rho_{11} - \rho_1)}{\partial \psi'} \right], \end{aligned} \quad (3.2)$$

where $\rho_1 = \rho_1(\psi_0)$ and $\rho_{js} = Y_j \tilde{Z}_j - \tilde{x}_{js} g(x_{js}; \theta) f_U(x_{js} - \Gamma_0 W_j; \psi_0) / h(x_{js})$ is the summand in $\rho_{j,S}(\psi_0) = \sum_{s=1}^S \rho_{js} / S$.

Asymptotically, the importance density $h(x)$ has no effect on the efficiency of $\hat{\psi}_{n,S}$, as long as it satisfies the condition of Theorem 3.1. In practice, however, the choice of $h(x)$ will affect the finite sample variances of the Monte Carlo estimators $m_S(\Gamma W_j; \psi)$ and $m_{2S}(\Gamma W_j; \psi)$. Theoretically, the best choice of $h(x)$ is proportional to the absolute value of the integrand $\|\tilde{x} g(x; \theta) f_U(x - \Gamma W; \psi)\|$. Practically, a density close to being proportional to the integrand is a good choice.

4 Semiparametric Estimator

In this and next section, we relax the parametric restriction on the distribution of U and instead assume that F_U is nonparametric. We derive a semiparametric estimator for θ and a kernel-based nonparametric estimator for F_U using moment equations (2.2) and (2.3) which become

$$E(Y | W) = \int g(\Gamma_0 W + u; \theta_0) dF_U(u) \quad (4.1)$$

and

$$E(YZ | W) = \int (\Gamma_0 W + u)g(\Gamma_0 W + u; \theta_0) dF_U(u). \quad (4.2)$$

The basic idea is to apply Fourier deconvolution to (4.1) or (4.2) to separate θ and F_U . This approach is based on the following assumptions.

ASSUMPTION 6 *The distribution of W is absolutely continuous with respect to Lebesgue measure and has support \mathbb{R}^l .*

ASSUMPTION 7 *$g(x; \theta_0)(\|x\| + 1) \in L^1(\mathbb{R}^k)$, the space of all absolutely integrable functions on \mathbb{R}^k . Furthermore, the set $\mathcal{T} = \{t \in \mathbb{R}^k : \tilde{g}(t; \theta_0) \neq 0\}$ is dense in \mathbb{R}^k , where $\tilde{g}(t; \theta_0) = \int e^{-it'x} g(x; \theta_0) dx$ is the Fourier transform of $g(x; \theta_0)$ and $i = \sqrt{-1}$.*

The integrability of $g(x; \theta_0)$ in ASSUMPTION 7 implies the existence of the Fourier transform $\tilde{g}(t; \theta_0)$. Roughly speaking, the second part of the assumption means that the zeros of $\tilde{g}(t; \theta_0)$ are isolated points in \mathbb{R}^k . The examples given at the end of next section show that this condition is fairly general. Further discussion and possible generalization of this condition is given in REMARK 5.1. For every $v \in \mathbb{R}^k$, let

$$m_1(v) = \int g(v + u; \theta_0) dF_U(u). \quad (4.3)$$

Since Γ_0 has full rank, ASSUMPTION 6 implies that $m_1(\Gamma_0 W) = E(Y | W)$ is fully observable on \mathbb{R}^k . Moreover, ASSUMPTION 7 implies that $m_1(v) \in L^1(\mathbb{R}^k)$. Taking Fourier transformation on both sides of (4.3) yields

$$\begin{aligned} \tilde{m}_1(t) &= \int e^{-it'v} m_1(v) dv \\ &= \int e^{-it'x} g(x; \theta_0) dx \cdot \int e^{it'u} dF_U(u) \\ &= \tilde{g}(t; \theta_0) \tilde{f}_U(t), \end{aligned} \quad (4.4)$$

where $\tilde{f}_U(t)$ is the characteristic function of U . Here we have slightly abused notation by using $\tilde{f}_U(t)$ to denote the Fourier inverse transform, which applies to $\tilde{f}_U(t)$ only throughout this article. It follows from (4.4) that, for any $t \in \mathcal{T}$, $\tilde{f}_U(t)$ is uniquely determined by

$$\tilde{f}_U(t) = \tilde{m}_1(t) / \tilde{g}(t; \theta_0). \quad (4.5)$$

Further, because any characteristic function is uniformly continuous in \mathbb{R}^k , ASSUMPTION 7 implies that the value of $\tilde{f}_U(t)$ at any zero of $\tilde{g}(t, \theta_0)$ is also uniquely determined. If, in addition, $\tilde{f}_U(t) \in L^1(\mathbb{R}^k)$, then the density of U exists and is given by

$$f_U(u; \theta_0) = \frac{1}{(2\pi)^k} \int e^{it'u} \frac{\tilde{m}_1(t)}{\tilde{g}(t; \theta_0)} dt. \quad (4.6)$$

This expression can be substituted into (4.1) and (4.2), so that the method of moments estimator for θ can be obtained by minimizing an objective function similar to (2.7). Details of this construction is given below.

First, let $\hat{\Gamma}$ denote the least squares estimator in (2.6) and $V_j = \hat{\Gamma}W_j, j = 1, 2, \dots, n$. Then the density function $f_V(v)$ of $V = \Gamma_0W$ and the conditional mean function $m_1(v)$ are estimated by

$$\hat{f}_V(v) = \frac{1}{na_n^k} \sum_{j=1}^n K\left(\frac{v - V_j}{a_n}\right) \quad (4.7)$$

and

$$\hat{m}_1(v) = \frac{1}{na_n^k} \sum_{j=1}^n Y_j K\left(\frac{v - V_j}{a_n}\right) / \hat{f}_V(v), \quad (4.8)$$

where $K(\cdot)$ is a kernel function and a_n is the bandwidth satisfying $0 < a_n \rightarrow 0$ as $n \rightarrow \infty$.

Second, let $\hat{\hat{m}}_1(t) = \int_{B_n} e^{-it'v} \hat{m}_1(v) dv$, where $B_n = \{v \in \mathbb{R}^k : |\hat{f}_V(v)| \geq b_n\}$, $0 < b_n \rightarrow 0$ as $n \rightarrow \infty$, and define, for each $\theta \in \Theta$,

$$\hat{f}_U(u; \theta) = \frac{1}{(2\pi)^k} \int_{C_n} e^{it'u} \frac{\hat{\hat{m}}_1(t)}{\tilde{g}(t; \theta)} dt,$$

where $C_n = \{t \in \mathbb{R}^k : \|t\| \leq 1/c_n, |\tilde{g}(t; \theta)| \geq c_n\}$ and $0 < c_n \rightarrow 0$, as $n \rightarrow \infty$.

Finally, the semiparametric estimator (SPE) for θ is defined as $\hat{\theta}_n = \operatorname{argmin}_{\theta \in \Theta} Q_n(\theta)$, where

$$Q_n(\theta) = \sum_{j=1}^n \hat{\rho}'_j(\theta) A_j \hat{\rho}_j(\theta) \quad (4.9)$$

with $\hat{\rho}_j(\theta) = Y_j \tilde{Z}_j - \int \tilde{x}g(x; \theta) \hat{f}_U(x - \hat{\Gamma}W_j; \theta) dx$. The consistency of $\hat{\theta}_n$ can be derived similarly as for the MME $\hat{\psi}_n$. However, as in many cases, e.g. Robinson (1988), the derivation becomes much more complicated because of the presence of the first-stage nonparametric estimators in $Q_n(\theta)$, which have convergence rates lower than \sqrt{n} . To achieve the \sqrt{n} -consistency, usually higher order kernels are used and combined with certain smoothness conditions for the density and conditional mean functions.

ASSUMPTION 8 *There exists an integer $d \geq 1$, such that $f_V(v)$, $m_1(v)f_V(v)$ and their partial derivatives of order 1 through d are continuous and uniformly bounded on \mathbb{R}^k .*

ASSUMPTION 9 *The kernel function $K(v)$ is bounded on \mathbb{R}^k and, for the integer d in ASSUMPTION 8, satisfies: (1) $\int K(v) dv = 1$; $\int v_1^{d_1} v_2^{d_2} \cdots v_k^{d_k} K(v) dv = 0$, for $d_j \geq 0$ and $1 \leq \sum_{j=1}^k d_j \leq d - 1$; and $\int |v_1^{d_1} v_2^{d_2} \cdots v_k^{d_k} K(v)| dv < \infty$, for $d_j \geq 0$ and $\sum_{j=1}^k d_j = d$; (2) for every $1 \leq j \leq k$, $\sup_{v \in \mathbb{R}^k} \|\partial K(v) / \partial v_j\| (\|v\| + 1) < \infty$; and (3) $K(v) \in L^1(\mathbb{R}^k)$ and $\int e^{it'v} K(v) dv \in L^1(\mathbb{R}^k)$.*

ASSUMPTION 10 $\sup_{\Theta} |g(x; \theta)| (\|x\| + 1) \in L^1(\mathbb{R}^k)$ and $\sup_{\Theta} |\tilde{m}_1(t)/\tilde{g}(t; \theta)| (\|t\| + 1) \in L^1(\mathbb{R}^k)$.

ASSUMPTION 11 $g(x; \theta)$ is a measurable function of x for each $\theta \in \Theta$ and is continuous in $\theta \in \Theta$ (a.e. μ). Furthermore, $E \|A(W)\| (\|W\|^2 + 1) < \infty$, $E \|A(W)\| (|Y|^2 + \|YZ\|^2) < \infty$ and

$$E \|A(W)\| \left(\int_{\Theta} \sup_{\Theta} |g(x; \theta) f_U(x - \Gamma_0 W; \theta)| (\|x\| + 1) dx \right)^2 < \infty. \quad (4.10)$$

ASSUMPTION 12 $E[\rho(\theta) - \rho(\theta_0)]' A(W) [\rho(\theta) - \rho(\theta_0)] = 0$ if and only if $\theta = \theta_0$, where $\rho(\theta) = Y\tilde{Z} - \int \tilde{x}g(x; \theta) f_U(x - \Gamma_0 W; \theta) dx$.

ASSUMPTIONS 8 and 9 have been used by Robinson (1988) and Andrews (1995) to achieve uniform convergence for their kernel estimators of the conditional mean functions. ASSUMPTION 10 guarantees that the Fourier transform $\tilde{g}(t; \theta)$ exists for all $\theta \in \Theta$ and that the density $f_U(u; \theta)$ exists and is given by (4.6). This assumption may be weakened to the condition that the density $f_U(u; \theta)$ exists and is piecewise continuous, in which case $f_U(u; \theta)$ may be defined by the usual inversion formula or the so-called principal value of the integral on the right-hand side of (4.6) (Walker (1988)). Similarly to ASSUMPTION 2, ASSUMPTION 12, is a high-level assumption for identifiability, which is implied by the conditions of THEOREM 5.1 in the next section.

THEOREM 4.1 Suppose ASSUMPTIONS 6 - 12 hold and, as $n \rightarrow \infty$, $\sqrt{n} a_n^{k+1} b_n^3 c_n^{k+1} \rightarrow \infty$, $a_n^d b_n^{-3} c_n^{-k-1} \rightarrow 0$ and $c_n^{-k-1} \int_{B_n^c} |m_1(v)| dv \xrightarrow{P} 0$, where B_n^c is the complement of B_n in \mathbb{R}^k . Then, as $n \rightarrow \infty$, (1) $\hat{\theta}_n \xrightarrow{P} \theta_0$; (2) $\sup_{u \in \mathbb{R}^k} \left| \hat{f}_U(u; \hat{\theta}_n) - f_U(u; \theta_0) \right| \xrightarrow{P} 0$; and (3) for every $t \in \mathbb{R}^k$ such that $\tilde{g}(t; \theta_0) \neq 0$, $\hat{f}_U(t; \hat{\theta}_n) = \hat{m}_1(t)/\tilde{g}(t; \hat{\theta}_n) \xrightarrow{P} \tilde{f}_U(t; \theta_0)$.

The above semiparametric estimator involves three tuning parameters. In practice, these parameters can be chosen as follows. First, take the bandwidth $a_n = n^{-a}$ where $0 < a < 1/2(k+1)$ can be chosen according to a certain optimum criterion for the kernel estimators in (4.7) and (4.8). Second, the quantity $d_n = \int_{B_n^c} |m_1(v)| dv$ reflects the tail behavior of $m_1(v)$ as $v \rightarrow \infty$ which can be evaluated for the given model $g(x, \theta_0)$ and density $f_V(v)$. Suppose $d_n = o(n^{-\delta})$ for some $\delta > 0$. Then $c_n = n^{-c\delta/(k+1)}$, $0 < c < 1$ satisfies $c_n^{-k-1} d_n \rightarrow 0$. Finally, choose $b > 0$ and $0 < c < 1$ such that $3b + c\delta < \min\{ad, 1/2 - a(k+1)\}$. Then $b_n = n^{-b}$ satisfies $\sqrt{n} a_n^{k+1} b_n^3 c_n^{k+1} = n^{1/2 - a(k+1) - 3b - c\delta} \rightarrow \infty$ and $a_n^d b_n^{-3} c_n^{-k-1} = n^{-ad + 3b + c\delta} \rightarrow 0$.

Similar to the simulation-based estimator of Section 3, we can also construct a simulated version of the semiparametric estimator. Specifically, $\hat{\rho}_j(\theta)$ in (4.9) can be replaced by Monte Carlo simulators such as $\hat{\rho}_{j,S}(\theta)$ and $\hat{\rho}_{j,2S}(\theta)$ in (3.1). Then a simulation-based semiparametric estimator (SBSPE) $\hat{\theta}_{n,S}$ can be defined by minimizing the simulated version of $Q_n(\theta)$, i.e.,

$$Q_{n,S}(\theta) = \sum_{j=1}^n \hat{\rho}'_{j,S}(\theta) A_j \hat{\rho}_{j,2S}(\theta), \quad (4.11)$$

where $\hat{\rho}_{j,S}(\theta) = Y_j \tilde{Z}_j - \frac{1}{S} \sum_{s=1}^S \tilde{x}_{js} g(x_{js}; \theta) \hat{f}_U(x_{js} - \hat{\Gamma} W_j; \theta) / h(x_{js})$, and $\hat{\rho}_{j,2S}(\theta)$ is defined similarly using $\{x_{js}, s = S+1, S+2, \dots, 2S\}$. Moreover, from Section 3 it is easy to see that $\hat{\theta}_{n,S}$ has the same properties given in THEOREM 4.1 for the SPE $\hat{\theta}_n$. The asymptotic normality of $\hat{\theta}_{n,S}$ can also be established in a similar way, under the following further assumptions.

ASSUMPTION 13 Θ contains an open neighborhood Θ_0 of θ_0 such that (1) $g(x; \theta)$ is twice continuously differentiable w.r.t. θ in Θ_0 ; (2) the first two derivatives of $\hat{m}_1(t)/\tilde{g}(t; \theta)$ w.r.t. θ and its first derivative w.r.t. γ are uniformly bounded in Θ_0 and an open and bounded neighborhood of γ_0 by $\eta(t) > 0$ which satisfies $\int_{C_n} \eta(t) dt < \infty$; (3) $g(x; \theta) f_U(x - \Gamma_0 W; \theta)$ has the same properties given in ASSUMPTIONS 3 and 4 for $g(x; \theta) f_U(x - \Gamma_0 W; \phi)$.

ASSUMPTION 14 The kernel function $K(v)$ admits the second order partial derivatives and $\sup_{v \in \mathbb{R}^k} \|\partial^2 K(v) / \partial v \partial v'\| < \infty$.

ASSUMPTION 15 $\sup_u |f_U^c(u; \theta_0) - f_U(u; \theta_0)| = o_p(n^{-1/2})$, where

$$f_U^c(u; \theta_0) = \frac{1}{(2\pi)^k} \int_{C_n} \frac{e^{-it'u}}{\tilde{g}(t; \theta_0)} \int_{B_n} e^{-it'v} m_1(v) dv dt.$$

ASSUMPTION 16 The matrix

$$B = E \left(\frac{\partial \rho'(\theta_0)}{\partial \theta} A(W) \frac{\partial \rho(\theta_0)}{\partial \theta'} \right)$$

is nonsingular.

The asymptotic normality of our estimator relies on the asymptotic behavior of the empirical process $\xi_j = \left(\xi'_{1j}, \xi'_{2j}, \xi'_{3j}, \xi'_{4j} \right)'$, where

$$\xi_{1j} = \frac{\partial \rho'_{j,S}(\theta_0)}{\partial \theta} A_j \rho_{j,2S}(\theta_0), \quad (4.12)$$

$$\xi_{2j} = \frac{1}{(2\pi)^k} E_j \frac{\partial \rho'_0(\theta_0)}{\partial \theta} A_j \int \tilde{x} g(x; \theta_0) \int_{C_n} \frac{e^{-it'(x - \Gamma_0 W)}}{\tilde{g}(t; \theta_0)} \int \frac{e^{-it'v} \eta_j(v)}{f_V(v)} dv dt dx, \quad (4.13)$$

$$\xi'_{3j} = \frac{\rho'_0(\theta_0) A_j}{(2\pi)^k} E_j \frac{\partial}{\partial \theta} \left[\int \tilde{x} g(x; \theta_0) \int_{C_n} \frac{e^{-it'(x - \Gamma_0 W)}}{\tilde{g}(t; \theta_0)} \right]_{\theta=\theta_0} \int \frac{e^{-it'v} \eta_j(v)}{f_V(v)} dv dt dx \quad (4.14)$$

and $\xi_{4j} = W_j \otimes (Z_j - \Gamma_0 W_j)$. Here $\rho_0(\theta_0) = YZ - E(YZ | W)$ and E_j denotes the conditional expectation given

$$\eta_j(v) = \frac{Y_j - m_1(v)}{a_n^k} K \left(\frac{v - \Gamma_0 W_j}{a_n} \right) - E \left[\frac{Y_j - m_1(v)}{a_n^k} K \left(\frac{v - \Gamma_0 W_j}{a_n} \right) \right]. \quad (4.15)$$

Then, we have the following result.

THEOREM 4.2 *In addition to the conditions of THEOREM 4.1, suppose $\sqrt{na_n^{2k+1}b_n^2} \rightarrow \infty$ and $\sqrt{na_n^q b_n^{-2} c_n^{-k-1}} \rightarrow 0$, where d is as in ASSUMPTION 8. Then under ASSUMPTIONS 13 - 16, as $n \rightarrow \infty$, $\sqrt{n}(\hat{\theta}_{n,S} - \theta_0) \xrightarrow{L} N(0, B^{-1}DCD'B^{-1})$, where $C = \lim_{n \rightarrow \infty} E\xi_j \xi_j'$ and*

$$D = \left[I_p, I_p, I_p, E \left(\frac{\partial \rho_j'(\theta_0)}{\partial \theta} A_j \frac{\partial \rho_j(\theta_0)}{\partial \gamma'} \right) (EWW' \otimes I_k)^{-1} \right].$$

It is easy to see that $\xi_{1j} = \xi_{5j} + \xi_{6j}$, where

$$\xi_{5j} = \frac{\partial \rho_j'(\theta_0)}{\partial \theta} A_j \rho_j(\theta_0),$$

$$\xi_{6j} = \frac{\partial \Delta_{j,2S}'}{\partial \theta} A_j \rho_j(\theta_0) + \frac{\partial \rho_j'(\theta_0)}{\partial \theta} A_j \Delta_{j,S} + \frac{\partial \Delta_{j,2S}'}{\partial \theta} A_j \Delta_{j,S}$$

and

$$\Delta_{j,S} = E(Y_j | W_j) - \frac{1}{S} \sum_{s=1}^S \frac{\tilde{x}_{js} g(x_{js}; \theta_0) f_U(x_{js} - \Gamma_0 W_j; \theta_0)}{h(x_{js})}.$$

Moreover, the asymptotic covariance matrix of $\hat{\theta}_{n,S}$ consists of the approximation errors of $\hat{\Gamma}$ for Γ_0 , ξ_{4j} , \hat{f}_u for f_u , ξ_{2j} and ξ_{3j} , the sampling error ξ_{5j} and the simulation error ξ_{6j} . It is easy to see that, if Γ_0 is known, then $\xi_{4j} = 0$, and if f_U is known, then $\xi_{2j} = \xi_{3j} = 0$. Therefore, if Γ_0 and f_U are known, the asymptotic covariance matrix of our simulation estimator only depends on the sampling error ξ_{5j} and simulation error ξ_{6j} . Since $E\xi_{6j} \xi_{6j}' = O(S^{-1})$, the impact of simulation error can be reduced by increasing the simulation size S .

5 Identifiability

Identifiability is a long-standing and difficult problem in nonlinear errors-in-variables models. It has both theoretical and practical importance, but very few results have been obtained so far because of its mathematical complexity. In the literature, this problem has usually been avoided by assuming distributions of certain unobserved variables or random errors be known, or it has been completely ignored in applied work. For a model with errors-in-variables to be identifiable, additional information such as validation data, repeated measurements or instrumental variables is needed (Fuller (1987) and Carroll, Ruppert and Stefanski (1995)). Hausman et al. (1991) show that the polynomial model is identifiable using instrumental variables. Wang and Hsiao (1995) obtained identifiability for models with integrable $g(x; \theta_0)$. Further, Schennach (2007) showed that the identifiability holds for general models which is not necessarily integrable. In this section, we use the framework of the previous section to derive a rank condition for identifiability of model (1.1) – (1.3). We first present the rank condition of Wang and Hsiao (1995) for the integrable $g(x; \theta_0)$ and then extend the result to more general case.

First, Γ_0 is clearly identifiable by (2.1) and the least squares method. In the previous section, we have demonstrated that F_U is uniquely determined by θ_0 and Γ_0 through (4.5). In the following,

we study the identifiability of θ_0 using (4.1) and (4.2), given that Γ_0 is identified. Analogous to (4.3), for every $v \in \mathbb{R}^k$, let

$$m_2(v) = \int (v + u)g(v + u; \theta_0) dF_U(u). \quad (5.1)$$

Then $m_2(\Gamma_0 W) = E(YZ | W)$ and ASSUMPTION 7 implies that $m_2(v) \in L^1(\mathbb{R}^k)$. Now, integrating both sides of (4.3) and (5.1) and applying Fubini Theorem, we obtain

$$\int m_1(v) dv = \int g(x; \theta_0) dx := g_1(\theta_0), \quad (5.2)$$

$$\int m_2(v) dv = \int xg(x; \theta_0) dx := g_2(\theta_0). \quad (5.3)$$

The left-hand sides of (5.2) and (5.3) are observable and the closed forms of the integrals on the right-hand sides can be obtained, because the functional form of $g(x; \theta_0)$ is known. By the Rank Theorem (Zeidler 1986, page 178), a sufficient condition for (5.2) - (5.3) to have unique solution θ_0 in its neighborhood is that the Jacobian matrix

$$J(\theta_0) = \left(\frac{\partial g_1(\theta_0)}{\partial \theta}, \frac{\partial g_2(\theta_0)}{\partial \theta} \right) \quad (5.4)$$

has full rank. Thus, we have the following result.

THEOREM 5.1 *Under ASSUMPTIONS 6 - 7, a sufficient condition for θ_0 and F_U to be identifiable is $\text{rank}J(\theta_0) = p$.*

It is easy to see that a necessary condition for $\text{rank}J(\theta_0) = p$ is $p \leq k + 1$, because $J(\theta_0)$ has dimensions p by $k + 1$.

REMARK 5.1 If the second condition in ASSUMPTION 7 is violated but there exists a $1 \leq j \leq k$, such that the set of all zeros of $\tilde{g}_j(t; \theta_0) = \int e^{-it'x} x_j g(x; \theta_0) dx$ is dense in \mathbb{R}^k , where x_j is the j -th coordinate of $x = (x_1, x_2, \dots, x_k)$, then $\tilde{f}_U(t)$ can still be identified by using the j -th equation in (5.1). This is easy to see by taking Fourier transformation on both sides of (5.1), which yields $\tilde{m}_2(t) = \tilde{f}_U(t) \int e^{-it'x} xg(x; \theta_0) dx$. Moreover, if $\tilde{f}_U(t)$ is analytic, then the second condition in ASSUMPTION 7 can be further weakened to the assumption that $\tilde{g}(t; \theta_0) \neq 0$ from some $t \in \mathbb{R}^k$. This follows from the facts that the continuity of $\tilde{g}(t; \theta_0)$ implies that $\tilde{g}(t; \theta_0) \neq 0$ in an open neighborhood, and that any analytic function is uniquely determined by its values on a finite segment of the complex plane. Note that any distribution admitting a moment generating function has an analytic characteristic function (Lukacs 1970, p197-198). Examples of such distributions include uniform, normal, double-exponential and many discrete distributions.

From a practical point of view, integrability of $g(x; \theta_0)$ in ASSUMPTION 7 is not as restrictive as it appears, because in many real problems, the possible values of X are bounded. In this sense a truncated model which vanishes outside a sufficiently large compact set can be used which satisfies ASSUMPTION 7. From a theoretical point of view, the integrability of $g(x; \theta_0)$ may be weakened to the following assumption.

ASSUMPTION 17 $E|g(X; \theta_0)|(\|X\| + 1) < \infty$.

To see this, let $g_n(x; \theta_0) = g(x; \theta_0)\mathbf{1}(\|x\| < T_n)$, where $\mathbf{1}(\cdot)$ is the indicator function and $T_n \rightarrow \infty$ (e.g., $T_n = cn^a$, for some $c > 0$ and $a > 0$), and modify $m_1(v)$ and $m_2(v)$ in (4.3) and (5.1) as

$$\begin{aligned} m_{1,n}(v) &= \int g_n(v+u; \theta_0) dF_U(u), \\ m_{2,n}(v) &= \int (v+u)g_n(v+u; \theta_0) dF_U(u). \end{aligned}$$

Then $m_{1,n}(\Gamma_0 W)$ and $m_{2,n}(\Gamma_0 W)$ approximate $E(Y | W)$ and $E(YZ | W)$ respectively in the following sense.

THEOREM 5.2 *Under ASSUMPTION 17, it holds*

$$\lim_{n \rightarrow \infty} E |m_{1,n}(\Gamma_0 W) - E(Y|W)| = 0 \quad (5.5)$$

and

$$\lim_{n \rightarrow \infty} E \|m_{2,n}(\Gamma_0 W) - E(YZ|W)\| = 0. \quad (5.6)$$

Since $E(Y|W)$ and $E(YZ|W)$ can be consistently estimated by nonparametric methods, together with Assumptions 6, THEOREM 5.2 implies that $m_{1,n}(\Gamma_0 W)$ and $m_{2,n}(\Gamma_0 W)$ are (asymptotically) observable on \mathbb{R}^k . Therefore results of THEOREM 5.1 hold with $g(x; \theta_0)$ replaced by $g_n(x; \theta_0)$ in ASSUMPTION 7 and in the Jacobian matrix $J(\theta_0)$.

In the rest of this section, we use some examples to illustrate how to apply THEOREM 5.1 to check model identifiability. Again, we consider cases where all variables are scalars and $\Gamma_0 = 1$. In this case, we need only to verify ASSUMPTIONS 7 or 17.

EXAMPLE 5.1 Exponential model $g(x; \theta) = e^{-\theta x^2}$, $\theta > 0$. Clearly $g(x; \theta)$ is integrable. Further, the second part of ASSUMPTION 7 is satisfied because $\tilde{g}(t; \theta) = \sqrt{\pi/\theta} e^{-t^2/4\theta}$. To check the rank condition, we calculate $g_1(\theta) = \int e^{-\theta x^2} dx = \sqrt{\pi/\theta}$ and $g_2(\theta) = \int x e^{-\theta x^2} dx = 0$. It follows that $J(\theta) = \left(-\sqrt{\pi}/(2\theta\sqrt{\theta}), 0\right)$ which has rank one. Therefore by THEOREM 5.1 the model is identifiable.

EXAMPLE 5.2 Linear model $g(x; \theta) = \theta x$. For this model, ASSUMPTION 17 is satisfied if $E\|X\|^2 < \infty$. Further, since for any $T > 0$,

$$\tilde{g}(t; \theta) = \int_{-T}^T e^{-it'x} \theta x dx = 2i\theta \left(\frac{T \cos(tT)}{t} - \frac{\sin(tT)}{t^2} \right),$$

the second condition in ASSUMPTION 7 is satisfied. To check the rank condition, we calculate $g_1(\theta) = \theta \int_{-T}^T x dx = 0$ and $g_2(\theta) = \theta \int_{-T}^T x^2 dx = 2\theta T^3/3$. Therefore $J(\theta) = (0, 2T^3/3)$ which has rank one. Hence by THEOREM 5.1 the model is identifiable.

EXAMPLE 5.3 Polynomial model $g(x; \theta) = \theta_1 x + \theta_2 x^2$. In this case, ASSUMPTION 17 is satisfied if $E \|X\|^3 < \infty$. Further, because for any $T > 0$,

$$\tilde{g}(t; \theta) = 2T \cos(tT) \left(\frac{i\theta_1}{t} + \frac{\theta_2}{t^2} \right) + 2 \sin(tT) \left(\frac{\theta_2 T^2}{t} - \frac{i\theta_1}{t^2} - \frac{2\theta_2}{t^3} \right),$$

the second condition in ASSUMPTION 7 is clearly satisfied. Again, it is straightforward to calculate $g_1(\theta) = 2\theta_2 T^3/3$ and $g_2(\theta) = 2\theta_1 T^3/3$. Hence

$$J(\theta) = \frac{2T^3}{3} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

which has rank two. Therefore the model is identifiable.

EXAMPLE 5.4 Exponential $g(x; \theta) = \exp(\theta x)$, $\theta \neq 0$. For this model, ASSUMPTION 17 becomes $\int e^{\theta x} dF_X(x) < \infty$ and $\int \|x\| e^{\theta x} dF_X(x) < \infty$, which are satisfied if X has a normal distribution. Since for any $T > 0$ and $t \neq -i\theta$,

$$\tilde{g}(t; \theta) = \frac{e^{(\theta-it)T} - e^{-(\theta-it)T}}{\theta - it},$$

the second condition in ASSUMPTION 7 is satisfied. Moreover, since for any θ , $g_1(\theta) = (e^{\theta T} - e^{-\theta T})/\theta$, and

$$g_2(\theta) = \frac{T(e^{\theta T} + e^{-\theta T})}{\theta} - \frac{e^{\theta T} - e^{-\theta T}}{\theta^2},$$

Jacobian matrix $J(\theta)$ is of rank one, which implies that the model is identifiable.

EXAMPLE 5.5 Consider $g(x; \theta) = \theta_1 + \theta_2 x^{\theta_3}$, where $\theta_2 \theta_3 \neq 0$. Clearly this model would be identifiable if X were observable. Since now the rank of $J(\theta)$ is at most two while $p = 3$, the model cannot be identified by (4.1) and (4.2).

EXAMPLE 5.6 Let $g(x; \theta) = (\theta_1 + \theta_2 x)^2$, $\theta_2 \neq 0$. Again, θ_1 and θ_2 would be identifiable if X were observable. However, θ_1 and θ_1 are not identifiable by (4.1) and (4.2) if $\theta_1^2 = \theta_2^2$, because now $J(\theta) = (\theta_1^2 - \theta_2^2) = 0$. Note that if the prior restriction $\theta_1 = \theta_2$ or $\theta_1 = -\theta_2$ is imposed, then the model can again be identifiable. However, these restrictions imply very different model specifications.

6 Conclusions and Discussion

Consistent estimation and identifiability of general nonlinear errors-in-variables models with multivariate predictor variables and possibly non-normal random errors have been challenging problems for decades. Most researchers rely on restrictive conditions to achieve consistent estimation, or treat more general models at the expense of the accuracy of estimation (e.g., approximately consistent approach). Moreover, most methods in the literature are designed for the case where either validation or replicate data are available.

In this paper, we use the instrumental variable approach to study a general model, where the predictor variable is multivariate and the distributions of the measurement error and the random error in the regression equation are nonparametric. Root- n consistent parametric and semiparametric estimators for the model are developed using the method of moments. A rank condition for model identifiability is derived by combining the nonparametric technique and Fourier deconvolution.

It is possible to generalize the prediction equation (1.3) to a nonlinear one, say, $X = \Gamma(W) + U$. All the results of this paper should be obtained analogously, provided function $\Gamma(\cdot)$ can be consistently estimated with convergence rate \sqrt{n} . The latter is generally satisfied if $\Gamma(\cdot)$ is parametric and estimated by the usual nonlinear least squares method. The independence between W and U is stronger than the usual instrumental variable assumption that they are uncorrelated. As pointed out by a referee, this assumption can be relaxed through parametric modeling of conditional distribution $f_{U|W}(u|w; \phi_0)$ instead of the marginal distribution $f_{U|W}(u; \phi_0)$. However, it is not clear, and deserves future research, how such an extension is possible for the nonparametric case. Another issue that should be investigated in the future research is the finite sample properties of the proposed estimators, which may be done through extensive and carefully designed simulation studies.

7 Proofs

7.1 Proof of Theorem 2.1

Since $f_U(u; \phi)$ is continuously differentiable with respect to u (ASSUMPTION 1), by (2.9) and the Dominated Convergence Theorem (DCT), $\hat{\rho}_j(\psi)$ and hence $Q_n(\psi)$ are continuously differentiable with respect to $\gamma = \text{vec}\Gamma$. For sufficiently large n , therefore, $Q_n(\psi)$ has the first-order Taylor expansion about $\gamma_0 = \text{vec}\Gamma_0$:

$$Q_n(\psi) = \sum_{j=1}^n \rho'_j(\psi) A_j \rho_j(\psi) + 2 \sum_{j=1}^n \rho'_j(\psi, \tilde{\gamma}) A_j \frac{\partial \rho_j(\psi, \tilde{\gamma})}{\partial \gamma'} (\tilde{\gamma} - \gamma_0), \quad (7.1)$$

where $\rho_j(\psi) = Y_j \tilde{Z}_j - m(\Gamma_0 W_j; \psi)$, $\rho_j(\psi, \tilde{\gamma}) = Y_j \tilde{Z}_j - m(\tilde{\Gamma} W_j; \psi)$,

$$\frac{\partial \rho_j(\psi, \tilde{\gamma})}{\partial \gamma'} = \int \tilde{x} g(x; \theta) \frac{\partial f_U(x - \tilde{\Gamma} W_j; \psi)}{\partial u'} dx (W_j \otimes I_k)'$$

and $\tilde{\gamma} = \text{vec}\tilde{\Gamma}$ satisfies $\|\tilde{\gamma} - \gamma_0\| \leq \|\hat{\gamma} - \gamma_0\|$. Further, since $g(x; \theta)$ and $f_U(u; \phi)$ are continuous in θ and ϕ respectively, by (2.8) and the DCT, $\rho_j(\psi)$ is continuous in ψ and, moreover,

$$\begin{aligned} E \sup_{\psi} |\rho'_1(\psi) A_1 \rho_1(\psi)| &\leq E \|A_1\| \sup_{\psi} \|\rho_1(\psi)\|^2 \\ &\leq 2E \|A_1\| (|Y_1|^2 + \|Y_1 Z_1\|^2) + \\ &\quad 2E \|A_1\| \left(\int \sup_{\psi} |g(x; \theta) f_U(x - \Gamma_0 W_1; \phi)| (\|x\| + 1) dx \right)^2 \\ &< \infty. \end{aligned}$$

It follows from the uniform law of large numbers (ULLN, Jennrich 1969, Theorem 2) that the first term on the right-hand side of (7.1) satisfies

$$\sup_{\psi} \left| \frac{1}{n} \sum_{j=1}^n \rho'_j(\psi) A_j \rho_j(\psi) - Q(\psi) \right| \xrightarrow{a.s.} 0, \quad (7.2)$$

where $Q(\psi) = E \rho'_1(\psi) A_1 \rho_1(\psi)$. Similarly, since by Cauchy-Schwarz inequality and (2.9),

$$\begin{aligned} & \left(E \sup_{\psi, \gamma} \left\| \rho'_1(\psi, \gamma) A_1 \frac{\partial \rho_1(\psi, \gamma)}{\partial \gamma'} \right\| \right)^2 \\ & \leq \left(E \|A_1\| \sup_{\psi, \gamma} \|\rho_1(\psi, \gamma)\| \left\| \frac{\partial \rho_1(\psi, \gamma)}{\partial \gamma'} \right\| \right)^2 \\ & \leq E \|A_1\| \sup_{\psi, \gamma} \|\rho_1(\psi, \gamma)\|^2 E \|A_1\| \sup_{\psi, \gamma} \left\| \frac{\partial \rho_1(\psi, \gamma)}{\partial \gamma'} \right\|^2 \\ & \leq k E \|A_1\| \sup_{\psi, \gamma} \|\rho_1(\psi, \gamma)\|^2 E \|A_1\| \|W_1\|^2 \left(\int \sup_{\psi, \gamma} \left\| g(x; \theta) \frac{\partial f_U(x - \Gamma W_1; \phi)}{\partial u'} \right\| (\|x\| + 1) dx \right)^2 \\ & < \infty, \end{aligned}$$

again by the ULLN we have

$$\sup_{\psi, \gamma} \left\| \frac{1}{n} \sum_{j=1}^n \rho'_j(\psi, \gamma) A_j \frac{\partial \rho_j(\psi, \gamma)}{\partial \gamma'} \right\| = O(1) \quad (a.s.)$$

and therefore

$$\begin{aligned} & \sup_{\psi} \left\| \frac{1}{n} \sum_{j=1}^n \rho'_j(\psi, \tilde{\gamma}) A_j \frac{\partial \rho_j(\psi, \tilde{\gamma})}{\partial \gamma'} (\tilde{\gamma} - \gamma_0) \right\| \\ & \leq \sup_{\psi, \gamma} \left\| \frac{1}{n} \sum_{j=1}^n \rho'_j(\psi, \gamma) A_j \frac{\partial \rho_j(\psi, \gamma)}{\partial \gamma'} \right\| \|\tilde{\gamma} - \gamma_0\| \xrightarrow{a.s.} 0. \end{aligned} \quad (7.3)$$

It follows from (7.1) - (7.3) that

$$\sup_{\psi} \left| \frac{1}{n} Q_n(\psi) - Q(\psi) \right| \xrightarrow{a.s.} 0. \quad (7.4)$$

Now we show that $Q(\psi)$ attains a unique minimum at $\psi_0 \in \Psi$. Because $E(\rho_1(\psi_0)|W_1) = 0$ and $\rho_1(\psi) - \rho_1(\psi_0)$ depends on W_1 only, we have

$$E[\rho'_1(\psi_0) A_1 (\rho_1(\psi) - \rho_1(\psi_0))] = E[E(\rho'_1(\psi_0)|W_1) A_1 (\rho_1(\psi) - \rho_1(\psi_0))] = 0,$$

which implies $Q(\psi) = Q(\psi_0) + E[(\rho_1(\psi) - \rho_1(\psi_0))' A_1 (\rho_1(\psi) - \rho_1(\psi_0))]$. By ASSUMPTION 2, $Q(\psi) \geq Q(\psi_0)$ and equality holds if and only if $\psi = \psi_0$. Thus, it follows from Amemiya (1973, Lemma 3) that $\hat{\psi}_n \xrightarrow{a.s.} \psi_0$.

7.2 Proof of Theorem 2.2

By ASSUMPTION 3 and the DCT, the first derivative $\partial Q_n(\psi)/\partial\psi$ exists and has the first-order Taylor expansion in the open neighborhood $\Psi_0 \subset \Psi$ of ψ_0 . Since $\partial Q_n(\hat{\psi}_n)/\partial\psi = 0$ and $\hat{\psi}_n \xrightarrow{a.s.} \psi_0$, for sufficiently large n we have

$$\frac{\partial Q_n(\psi_0)}{\partial\psi} + \frac{\partial^2 Q_n(\tilde{\psi})}{\partial\psi\partial\psi'}(\hat{\psi}_n - \psi_0) = 0, \quad (7.5)$$

where $\|\tilde{\psi} - \psi_0\| \leq \|\hat{\psi}_n - \psi_0\|$. The first derivative of $Q_n(\psi)$ in (7.5) is given by

$$\frac{\partial Q_n(\psi)}{\partial\psi} = 2 \sum_{j=1}^n \frac{\partial \hat{\rho}'_j(\psi)}{\partial\psi} A_j \hat{\rho}_j(\psi), \quad (7.6)$$

where $\partial \hat{\rho}'_j(\psi)/\partial\psi$ consists of

$$\frac{\partial \hat{\rho}'_j(\psi)}{\partial\theta} = - \int \frac{\partial g(x; \theta)}{\partial\theta} \tilde{x}' f_U(x - \hat{\Gamma}W_j; \phi) dx \quad (7.7)$$

and

$$\frac{\partial \hat{\rho}'_j(\psi)}{\partial\phi} = - \int \frac{\partial f_U(x - \hat{\Gamma}W_j; \phi)}{\partial\phi} \tilde{x}' g(x; \theta) dx. \quad (7.8)$$

The second derivative in (7.5) is given by

$$\frac{\partial^2 Q_n(\psi)}{\partial\psi\partial\psi'} = 2 \sum_{j=1}^n \left[\frac{\partial \hat{\rho}'_j(\psi)}{\partial\psi} A_j \frac{\partial \hat{\rho}_j(\psi)}{\partial\psi'} + (\hat{\rho}'_j(\psi) A_j \otimes I_{p+q}) \frac{\partial \text{vec}(\partial \hat{\rho}'_j(\psi)/\partial\psi)}{\partial\psi'} \right],$$

where $\partial \text{vec}(\partial \hat{\rho}'_j(\psi)/\partial\psi)/\partial\psi'$ consists of

$$\frac{\partial \text{vec}(\partial \hat{\rho}'_j(\psi)/\partial\theta)}{\partial\theta'} = - \int \tilde{x} \otimes \frac{\partial^2 g(x; \theta)}{\partial\theta\partial\theta'} f_U(x - \hat{\Gamma}W_j; \phi) dx,$$

$$\frac{\partial \text{vec}(\partial \hat{\rho}'_j(\psi)/\partial\theta)}{\partial\phi'} = - \int \tilde{x} \otimes \frac{\partial g(x; \theta)}{\partial\theta} \frac{\partial f_U(x - \hat{\Gamma}W_j; \phi)}{\partial\phi'} dx,$$

$$\frac{\partial \text{vec}(\partial \hat{\rho}'_j(\psi)/\partial\phi)}{\partial\theta'} = - \int \tilde{x} \otimes \frac{\partial f_U(x - \hat{\Gamma}W_j; \phi)}{\partial\phi} \frac{\partial g(x; \theta)}{\partial\theta'} dx,$$

and

$$\frac{\partial \text{vec}(\partial \hat{\rho}'_j(\psi)/\partial\phi)}{\partial\phi'} = - \int \tilde{x} \otimes \frac{\partial^2 f_U(x - \hat{\Gamma}W_j; \phi)}{\partial\phi\partial\phi'} g(x; \theta) dx.$$

It follows from ASSUMPTION 3 that

$$\begin{aligned}
& E \sup_{\psi, \gamma} \left\| \frac{\partial \rho'_1(\psi, \gamma)}{\partial \psi} A_1 \frac{\partial \rho_1(\psi, \gamma)}{\partial \psi'} \right\| \\
& \leq E \|A_1\| \sup_{\psi, \gamma} \left\| \frac{\partial \rho_1(\psi, \gamma)}{\partial \psi'} \right\|^2 \\
& = E \|A_1\| \sup_{\psi, \gamma} \left(\left\| \frac{\partial \rho_1(\psi, \gamma)}{\partial \theta'} \right\|^2 + \left\| \frac{\partial \rho_1(\psi, \gamma)}{\partial \phi'} \right\|^2 \right) \\
& \leq E \|A_1\| \left(\int \sup_{\psi, \gamma} \left\| \frac{\partial g(x; \theta)}{\partial \theta} f_U(x - \Gamma W_1; \phi) \right\| (\|x\| + 1) dx \right)^2 \\
& + E \|A_1\| \left(\int \sup_{\psi, \gamma} \left\| \frac{\partial f_U(x - \Gamma W_j; \phi)}{\partial \phi} g(x; \theta) dx \right\| (\|x\| + 1) dx \right)^2 \\
& < \infty.
\end{aligned}$$

Similarly, by ASSUMPTION 3 we have

$$\begin{aligned}
& \left(E \sup_{\psi, \gamma} \left\| (\rho'_1(\psi, \gamma) A_1 \otimes I_{p+q}) \frac{\partial \text{vec}(\partial \rho'_1(\psi, \gamma) / \partial \psi)}{\partial \psi'} \right\| \right)^2 \\
& \leq (p+q) \left(E \|A_1\| \sup_{\psi, \gamma} \|\rho_1(\psi, \gamma)\| \left\| \frac{\partial \text{vec}(\partial \rho'_1(\psi, \gamma) / \partial \psi)}{\partial \psi'} \right\| \right)^2 \\
& \leq (p+q) E \|A_1\| \sup_{\psi, \gamma} \|\rho_1(\psi, \gamma)\|^2 E \|A_1\| \sup_{\psi, \gamma} \left\| \frac{\partial \text{vec}(\partial \rho'_1(\psi, \gamma) / \partial \psi)}{\partial \psi'} \right\|^2 \\
& < \infty.
\end{aligned}$$

It follows from the ULLN and Amemiya (1973, Lemma 4) that

$$\begin{aligned}
\frac{1}{2n} \frac{\partial^2 Q_n(\tilde{\psi})}{\partial \psi \partial \psi'} & \xrightarrow{a.s.} E \left[\frac{\partial \rho'_1(\psi_0)}{\partial \psi} A_1 \frac{\partial \rho_1(\psi_0)}{\partial \psi'} + (\rho'_1(\psi_0) A_1 \otimes I_{p+q}) \frac{\partial \text{vec}(\partial \rho'_1(\psi_0) / \partial \psi)}{\partial \psi'} \right] \\
& = B,
\end{aligned} \tag{7.9}$$

where the last equality holds because $\partial \text{vec}(\partial \rho'_1(\psi_0) / \partial \psi) / \partial \psi'$ depends on W_1 only and therefore

$$\begin{aligned}
& E \left[(\rho'_1(\psi_0) A_1 \otimes I_{p+q}) \frac{\partial \text{vec}(\partial \rho'_1(\psi_0) / \partial \psi)}{\partial \psi'} \right] \\
& = E \left[(E(\rho'_1(\psi_0) | W_1) A_1 \otimes I_{p+q}) \frac{\partial \text{vec}(\partial \rho'_1(\psi_0) / \partial \psi)}{\partial \psi'} \right] \\
& = 0.
\end{aligned}$$

Further, by ASSUMPTION 4 and the DCT, $\partial Q_n(\psi_0) / \partial \psi$ is continuously differentiable with respect to γ and hence, for sufficiently large n , has the first-order Taylor expansion about γ_0 :

$$\frac{\partial Q_n(\psi_0)}{\partial \psi} = 2 \sum_{j=1}^n \frac{\partial \rho'_j(\psi_0)}{\partial \psi} A_j \rho_j(\psi_0) + \frac{\partial^2 \tilde{Q}_n(\psi_0)}{\partial \psi \partial \gamma'} (\hat{\gamma} - \gamma_0), \tag{7.10}$$

where

$$\begin{aligned} \frac{\partial^2 \tilde{Q}_n(\psi_0)}{\partial \psi \partial \gamma'} &= 2 \sum_{j=1}^n \left[\frac{\partial \rho'_j(\psi_0, \tilde{\gamma})}{\partial \psi} A_j \frac{\partial \rho_j(\psi_0, \tilde{\gamma})}{\partial \gamma'} + (\rho'_j(\psi_0, \tilde{\gamma}) A_j \otimes I_{p+q}) \frac{\partial \text{vec}(\partial \rho'_j(\psi_0, \tilde{\gamma}) / \partial \psi)}{\partial \gamma'} \right], \\ \frac{\partial \rho_j(\psi_0, \tilde{\gamma})}{\partial \gamma'} &= \int \tilde{x} g(x; \theta_0) \frac{\partial f_U(x - \tilde{\Gamma} W_j; \phi_0)}{\partial u'} dx (W_j \otimes I_k)', \\ \frac{\partial \text{vec}(\partial \rho'_j(\psi_0, \tilde{\gamma}) / \partial \theta)}{\partial \gamma'} &= \int \tilde{x} \otimes \frac{\partial g(x; \theta_0)}{\partial \theta} \frac{\partial f_U(x - \tilde{\Gamma} W_j; \phi_0)}{\partial u'} dx (W_j \otimes I_k)', \\ \frac{\partial \text{vec}(\partial \rho'_j(\psi_0, \tilde{\gamma}) / \partial \phi)}{\partial \gamma'} &= \int g(x; \theta_0) \tilde{x} \otimes \frac{\partial^2 f_U(x - \tilde{\Gamma} W_j; \phi_0)}{\partial \phi \partial u'} dx (W_j \otimes I_k)' \end{aligned}$$

and $\tilde{\gamma} = \text{vec} \tilde{\Gamma}$ satisfies $\|\tilde{\gamma} - \gamma_0\| \leq \|\hat{\gamma} - \gamma_0\|$. Similarly to (7.9), by ASSUMPTION 4 we can show that

$$\frac{1}{2n} \frac{\partial^2 \tilde{Q}_n(\psi_0)}{\partial \psi \partial \gamma'} \xrightarrow{a.s.} E \left[\frac{\partial \rho'_1(\psi_0)}{\partial \psi} A_1 \frac{\partial \rho_1(\psi_0)}{\partial \gamma'} \right]. \quad (7.11)$$

By definition (2.6), $\hat{\Gamma} - \Gamma_0 = \left[\sum_{j=1}^n (Z_j - \Gamma_0 W_j) W_j' \right] \left(\sum_{j=1}^n W_j W_j' \right)^{-1}$, which can be written as

$$\hat{\gamma} - \gamma_0 = \text{vec}(\hat{\Gamma} - \Gamma_0) = \left(\sum_{j=1}^n W_j W_j' \otimes I_k \right)^{-1} \sum_{j=1}^n W_j \otimes (Z_j - \Gamma_0 W_j)$$

(Magnus and Neudecker (1988), p.30). Hence (7.10) can be written as

$$\frac{\partial Q_n(\psi_0)}{\partial \psi} = 2D_n \sum_{j=1}^n T_j,$$

where

$$D_n = \left(I_{p+q}, \frac{1}{2} \frac{\partial^2 \tilde{Q}_n(\psi_0)}{\partial \psi \partial \gamma'} \left(\sum_{j=1}^n W_j W_j' \otimes I_k \right)^{-1} \right)$$

and

$$T_j = \begin{pmatrix} \frac{\partial \rho'_j(\psi_0)}{\partial \psi} A_j \rho_j(\psi_0) \\ W_j \otimes (Z_j - \Gamma_0 W_j) \end{pmatrix}.$$

By the Law of Large Numbers, $n \left(\sum_{j=1}^n W_j W_j' \otimes I_k \right)^{-1} \xrightarrow{a.s.} (E W_1 W_1' \otimes I_k)^{-1}$, which together with (7.11) implies

$$D_n \xrightarrow{a.s.} \left(I_{p+q}, E \left[\frac{\partial \rho'_1(\psi_0)}{\partial \psi} A_1 \frac{\partial \rho_1(\psi_0)}{\partial \gamma'} \right] (E W_1 W_1' \otimes I_k)^{-1} \right) = D,$$

where D is given in Theorem 2.2. Moreover, since $T_j, j = 1, 2, \dots, n$ are *i.i.d.*, by the Central Limit Theorem, $\frac{1}{\sqrt{n}} \sum_{j=1}^n T_j \xrightarrow{L} N(0, C)$, where $C = E(T_1 T_1')$. Therefore, by Slutsky's Theorem, we have

$$\frac{1}{2\sqrt{n}} \frac{\partial Q_n(\psi_0)}{\partial \psi} \xrightarrow{L} N(0, D C D'). \quad (7.12)$$

Finally, the theorem follows from (7.5), (7.9) and (7.12).

7.3 Proof of Theorem 3.1

We first prove Theorem 3.1.1. By ASSUMPTION 1 $Q_{n,S}(\psi)$ has the first-order Taylor expansion about γ_0 :

$$\begin{aligned} Q_{n,S}(\psi) &= \sum_{j=1}^n \rho'_{j,S}(\psi) A_j \rho_{j,2S}(\psi) \\ &+ \sum_{j=1}^n \left[\rho'_{j,S}(\psi, \tilde{\gamma}) A_j \frac{\partial \rho_{j,2S}(\psi, \tilde{\gamma})}{\partial \gamma'} + \rho'_{j,2S}(\psi, \tilde{\gamma}) A_j \frac{\partial \rho_{j,S}(\psi, \tilde{\gamma})}{\partial \gamma'} \right] (\hat{\gamma} - \gamma_0), \end{aligned} \quad (7.13)$$

where $\rho_{j,S}(\psi) = Y_j \tilde{Z}_j - m_S(\Gamma_0 W_j; \psi)$, $\rho_{j,S}(\psi, \tilde{\gamma}) = Y_j \tilde{Z}_j - m_S(\tilde{\Gamma} W_j; \psi)$,

$$\frac{\partial \rho_{j,S}(\psi, \tilde{\gamma})}{\partial \gamma'} = \frac{1}{S} \sum_{s=1}^S \frac{\tilde{x}_{js} g(x_{js}; \theta)}{h(x_{js})} \frac{\partial f_U(x_{js} - \tilde{\Gamma} W_j; \phi)}{\partial u'} (W_j \otimes I_k)',$$

$\|\tilde{\gamma} - \gamma_0\| \leq \|\hat{\gamma} - \gamma_0\|$ and $\rho_{j,2S}(\psi, \tilde{\gamma})$, $\partial \rho_{j,2S}(\psi, \tilde{\gamma}) / \partial \gamma'$ are given similarly. Further, because

$$\begin{aligned} &E \left(\sup_{\psi} \|\rho_{1,S}(\psi)\| \mid W_1, Y_1, Z_1 \right) \\ &\leq \|Y_1 \tilde{Z}_1\| + E \left(\sup_{\psi} \|m_S(\Gamma_0 W_1; \psi)\| \mid W_1 \right) \\ &\leq \|Y_1 \tilde{Z}_1\| + \frac{1}{S} \sum_{s=1}^S E \left(\sup_{\psi} \left\| \frac{\tilde{x}_{1s} g(x_{1s}; \theta) f_U(x_{1s} - \Gamma_0 W_1; \phi)}{h(x_{1s})} \right\| \mid W_1 \right) \\ &\leq \|Y_1 \tilde{Z}_1\| + \int \sup_{\psi} |g(x; \theta) f_U(x - \Gamma_0 W_1; \phi)| (\|x\| + 1) dx, \end{aligned}$$

and similarly

$$E \left(\sup_{\psi} \|\rho_{1,2S}(\psi)\| \mid W_1, Y_1, Z_1 \right) \leq \|Y_1 \tilde{Z}_1\| + \int \sup_{\psi} |g(x; \theta) f_U(x - \Gamma_0 W_1; \phi)| (\|x\| + 1) dx,$$

and because $\rho_{1,S}(\psi)$ and $\rho_{1,2S}(\psi)$ are conditionally independent given (W_1, Y_1, Z_1) , it follows from ASSUMPTIONS 1 and (2.8) that

$$\begin{aligned} &E \sup_{\psi} |\rho'_{1,S}(\psi) A_1 \rho_{1,2S}(\psi)| \\ &\leq E \|A_1\| E \left(\sup_{\psi} \|\rho_{1,S}(\psi)\| \mid W_1, Y_1, Z_1 \right) E \left(\sup_{\psi} \|\rho_{1,2S}(\psi)\| \mid W_1, Y_1, Z_1 \right) \\ &\leq E \|A_1\| \left(\|Y_1 \tilde{Z}_1\| + \int \sup_{\psi} |g(x; \theta) f_U(x - \Gamma_0 W_1; \phi)| (\|x\| + 1) dx \right)^2 \\ &\leq 2E \|A_1\| (\|Y_1\|^2 + \|Y_1 Z_1\|^2) + 2E \|A_1\| \left(\int \sup_{\psi} |g(x; \theta) f_U(x - \Gamma_0 W_1; \phi)| (\|x\| + 1) dx \right)^2 \\ &< \infty. \end{aligned}$$

Therefore by the ULLN the first term on the right-hand side of (7.13) satisfies

$$\sup_{\psi} \left| \frac{1}{n} \sum_{j=1}^n \rho'_{j,S}(\psi) A_j \rho_{j,2S}(\psi) - E \rho'_{1,S}(\psi) A_1 \rho_{1,2S}(\psi) \right| \xrightarrow{a.s.} 0, \quad (7.14)$$

where

$$\begin{aligned} E \rho'_{1,S}(\psi) A_1 \rho_{1,2S}(\psi) &= E [E(\rho'_{1,S}(\psi) | W_1, Y_1, Z_1) A_1 E(\rho_{1,2S}(\psi) | W_1, Y_1, Z_1)] \\ &= E \rho'_1(\psi) A_1 \rho_1(\psi) \\ &= Q(\psi). \end{aligned}$$

The second term on the right-hand side of (7.13) satisfies

$$\begin{aligned} & E \sup_{\psi, \gamma} \left\| \rho'_{1,S}(\psi, \gamma) A_1 \frac{\partial \rho_{1,2S}(\psi, \gamma)}{\partial \gamma'} \right\| \\ & \leq E \|A_1\| \sup_{\psi, \gamma} \left\| \rho'_{1,S}(\psi, \gamma) \right\| \left\| \frac{\partial \rho_{1,2S}(\psi, \gamma)}{\partial \gamma'} \right\| \\ & \leq E \|A_1\| E \left(\sup_{\psi, \gamma} \|\rho_{1,S}(\psi, \gamma)\| |W_1, Y_1, Z_1 \right) E \left(\sup_{\psi, \gamma} \left\| \frac{\partial \rho_{1,2S}(\psi, \gamma)}{\partial \gamma'} \right\| |W_1, Y_1, Z_1 \right). \end{aligned}$$

Since

$$\begin{aligned} & E \left(\sup_{\psi, \gamma} \left\| \frac{\partial \rho_{1,2S}(\psi, \gamma)}{\partial \gamma'} \right\| |W_1, Y_1, Z_1 \right) \\ & \leq \frac{1}{S} \sum_{s=S+1}^{2S} E \left(\sup_{\psi, \gamma} \left\| \frac{\tilde{x}_{1s} g(x_{1s}; \theta)}{h(x_{1s})} \frac{\partial f_U(x_{1s} - \Gamma W_1; \phi)}{\partial u'} (W_1 \otimes I_k)' \right\| |W_1 \right) \\ & \leq \sqrt{k} \|W_1\| \int \sup_{\psi, \gamma} \left\| g(x; \theta) \frac{\partial f_U(x - \Gamma W_1; \phi)}{\partial u'} \right\| (\|x\| + 1) dx, \end{aligned}$$

by Cauchy-Schwarz inequality and ASSUMPTION 1,

$$\begin{aligned} & \left(E \sup_{\psi, \gamma} \left\| \rho'_{1,S}(\psi, \gamma) A_1 \frac{\partial \rho_{1,2S}(\psi, \gamma)}{\partial \gamma'} \right\| \right)^2 \\ & \leq \left[E \|A_1\| \left(\left\| Y_1 \tilde{Z}_1 \right\| + \int \sup_{\psi, \gamma} |g(x; \theta) f_U(x - \Gamma W_1; \phi)| (\|x\| + 1) dx \right) \times \right. \\ & \quad \left. \left(\sqrt{k} \|W_1\| \int \sup_{\psi, \gamma} \left\| g(x; \theta) \frac{\partial f_U(x - \Gamma W_1; \phi)}{\partial u'} \right\| (\|x\| + 1) dx \right) \right]^2 \\ & \leq k E \|A_1\| \left(\left\| Y_1 \tilde{Z}_1 \right\| + \int \sup_{\psi, \gamma} |g(x; \theta) f_U(x - \Gamma W_1; \phi)| (\|x\| + 1) dx \right)^2 \times \\ & \quad E \|A_1\| \left(\|W_1\| \int \sup_{\psi, \gamma} \left\| g(x; \theta) \frac{\partial f_U(x - \Gamma W_1; \phi)}{\partial u'} \right\| (\|x\| + 1) dx \right)^2 \\ & < \infty. \end{aligned}$$

Thus by the ULLN we have

$$\begin{aligned} & \sup_{\psi} \left\| \frac{1}{n} \sum_{j=1}^n \rho'_{j,2S}(\psi, \tilde{\gamma}) A_j \frac{\partial \rho_{j,S}(\psi, \tilde{\gamma})}{\partial \gamma'} (\hat{\gamma} - \gamma_0) \right\| \\ & \leq \sup_{\psi, \gamma} \left\| \frac{1}{n} \sum_{j=1}^n \rho'_{j,2S}(\psi, \gamma) A_j \frac{\partial \rho_{j,S}(\psi, \gamma)}{\partial \gamma'} \right\| \|(\hat{\gamma} - \gamma_0)\| \xrightarrow{a.s.} 0. \end{aligned} \quad (7.15)$$

It follows from (7.13) - (7.15) that

$$\sup_{\psi} \left| \frac{1}{n} Q_{n,S}(\psi) - Q(\psi) \right| \xrightarrow{a.s.} 0. \quad (7.16)$$

It has been shown in the proof of THEOREM 2.1 that $Q(\psi)$ attains a unique minimum at $\psi_0 \in \Psi$. Therefore $\hat{\psi}_{n,S} \xrightarrow{a.s.} \psi_0$ follows from Amemiya (1973, Lemma 3).

Next we prove Theorem 3.1.2. First, by ASSUMPTION 3 $\partial Q_{n,S}(\psi)/\partial \psi$ has the first-order Taylor expansion in an open neighborhood $\Psi_0 \subset \Psi$ of ψ_0 :

$$\frac{\partial Q_{n,S}(\psi_0)}{\partial \psi} + \frac{\partial^2 Q_{n,S}(\tilde{\psi})}{\partial \psi \partial \psi'} (\hat{\psi}_{n,S} - \psi_0) = 0, \quad (7.17)$$

where $\|\tilde{\psi} - \psi_0\| \leq \|\hat{\psi}_{n,S} - \psi_0\|$. The first derivative of $Q_{n,S}(\psi)$ in (7.17) is given by

$$\frac{\partial Q_{n,S}(\psi)}{\partial \psi} = \sum_{j=1}^n \left[\frac{\partial \hat{\rho}'_{j,S}(\psi)}{\partial \psi} A_j \hat{\rho}_{j,2S}(\psi) + \frac{\partial \hat{\rho}'_{j,2S}(\psi)}{\partial \psi} A_j \hat{\rho}_{j,S}(\psi) \right], \quad (7.18)$$

where $\partial \hat{\rho}'_{j,S}(\psi)/\partial \psi$ consists of

$$\frac{\partial \hat{\rho}'_{j,S}(\psi)}{\partial \theta} = -\frac{1}{S} \sum_{s=1}^S \frac{\partial g(x_{js}; \theta)}{\partial \theta} \frac{\tilde{x}'_{js} f_U(x_{js} - \hat{\Gamma} W_j; \phi)}{h(x_{js})}$$

and

$$\frac{\partial \hat{\rho}'_{j,S}(\psi)}{\partial \phi} = -\frac{1}{S} \sum_{s=1}^S \frac{\partial f_U(x_{js} - \hat{\Gamma} W_j; \phi)}{\partial \phi} \frac{\tilde{x}'_{js} g(x_{js}; \theta)}{h(x_{js})}.$$

The second derivative in (7.17) is given by

$$\begin{aligned} \frac{\partial^2 Q_{n,S}(\psi)}{\partial \psi \partial \psi'} &= \sum_{j=1}^n \left[\frac{\partial \hat{\rho}'_{j,S}(\psi)}{\partial \psi} A_j \frac{\partial \hat{\rho}_{j,2S}(\psi)}{\partial \psi'} + (\hat{\rho}'_{j,2S}(\psi) A_j \otimes I_{p+q}) \frac{\partial \text{vec}(\partial \hat{\rho}'_{j,S}(\psi)/\partial \psi)}{\partial \psi'} \right] \\ &+ \sum_{j=1}^n \left[\frac{\partial \hat{\rho}'_{j,2S}(\psi)}{\partial \psi} A_j \frac{\partial \hat{\rho}_{j,S}(\psi)}{\partial \psi'} + (\hat{\rho}'_{j,S}(\psi) A_j \otimes I_{p+q}) \frac{\partial \text{vec}(\partial \hat{\rho}'_{j,2S}(\psi)/\partial \psi)}{\partial \psi'} \right], \end{aligned}$$

where $\partial \text{vec}(\partial \hat{\rho}'_{j,S}(\psi)/\partial \psi)/\partial \psi'$ consists of

$$\frac{\partial \text{vec}(\partial \hat{\rho}'_{j,S}(\psi)/\partial \theta)}{\partial \theta'} = -\frac{1}{S} \sum_{s=1}^S \tilde{x}_{js} \otimes \frac{\partial^2 g(x_{js}; \theta)}{\partial \theta \partial \theta'} \frac{f_U(x_{js} - \hat{\Gamma} W_j; \phi)}{h(x_{js})},$$

$$\frac{\partial \text{vec}(\partial \hat{\rho}'_{j,S}(\psi)/\partial \theta)}{\partial \phi'} = -\frac{1}{S} \sum_{s=1}^S \tilde{x}_{js} \otimes \frac{\partial g(x_{js}; \theta)}{\partial \theta} \frac{\partial f_U(x_{js} - \hat{\Gamma}W_j; \phi)}{\partial \phi'} \frac{1}{h(x_{js})},$$

$$\frac{\partial \text{vec}(\partial \hat{\rho}'_{j,S}(\psi)/\partial \phi)}{\partial \theta'} = -\frac{1}{S} \sum_{s=1}^S \tilde{x}_{js} \otimes \frac{\partial f_U(x_{js} - \hat{\Gamma}W_j; \phi)}{\partial \phi} \frac{\partial g(x_{js}; \theta)}{\partial \theta'} \frac{1}{h(x_{js})},$$

and

$$\frac{\partial \text{vec}(\partial \hat{\rho}'_{j,S}(\psi)/\partial \phi)}{\partial \phi'} = -\frac{1}{S} \sum_{s=1}^S \tilde{x}_{js} \otimes \frac{\partial^2 f_U(x_{js} - \hat{\Gamma}W_j; \phi)}{\partial \phi \partial \phi'} \frac{g(x_{js}; \theta)}{h(x_{js})}.$$

The elements in $\partial \hat{\rho}'_{j,2S}(\psi)/\partial \psi$ and $\partial \text{vec}(\partial \hat{\rho}'_{j,2S}(\psi)/\partial \psi)/\partial \psi'$ are given similarly. Completely analogous to (7.9), we can show that $\frac{1}{n} \frac{\partial^2 Q_{n,S}(\psi)}{\partial \psi \partial \psi'}$ converges *a.s.* to

$$E \left[\frac{\partial \rho'_{1,S}(\psi_0)}{\partial \psi} A_1 \frac{\partial \rho_{1,2S}(\psi_0)}{\partial \psi'} + (\rho'_{1,2S}(\psi_0) A_1 \otimes I_{p+q}) \frac{\partial \text{vec}(\partial \rho'_{1,S}(\psi_0)/\partial \psi)}{\partial \psi'} \right] +$$

$$E \left[\frac{\partial \rho'_{1,2S}(\psi_0)}{\partial \psi} A_1 \frac{\partial \rho_{1,S}(\psi_0)}{\partial \psi'} + (\rho'_{1,S}(\psi_0) A_1 \otimes I_{p+q}) \frac{\partial \text{vec}(\partial \rho'_{1,2S}(\psi_0)/\partial \psi)}{\partial \psi'} \right]$$

uniformly for all $\psi \in \Psi$. Since

$$E \left[\frac{\partial \rho'_{1,S}(\psi_0)}{\partial \psi} A_1 \frac{\partial \rho_{1,2S}(\psi_0)}{\partial \psi'} \right] = E \left[E \left(\frac{\partial \rho'_{1,S}(\psi_0)}{\partial \psi} | W_1 \right) A_1 E \left(\frac{\partial \rho_{1,2S}(\psi_0)}{\partial \psi'} | W_1 \right) \right]$$

$$= E \left[\frac{\partial \rho'_1(\psi_0)}{\partial \psi} A_1 \frac{\partial \rho_1(\psi_0)}{\partial \psi'} \right]$$

$$= B$$

and

$$E \left[(\rho'_{1,2S}(\psi_0) A_1 \otimes I_{p+q}) \frac{\partial \text{vec}(\partial \rho'_{1,S}(\psi_0)/\partial \psi)}{\partial \psi'} \right]$$

$$= E \left[(E(\rho'_{1,2S}(\psi_0) | W_1) A_1 \otimes I_{p+q}) \frac{\partial \text{vec}(\partial \rho'_{1,S}(\psi_0)/\partial \psi)}{\partial \psi'} \right]$$

$$= 0,$$

we have

$$\frac{1}{n} \frac{\partial^2 Q_{n,S}(\tilde{\psi})}{\partial \psi \partial \psi'} \xrightarrow{a.s.} 2B. \quad (7.19)$$

Again, by ASSUMPTION 4 $\partial Q_{n,S}(\psi_0)/\partial \psi$ has the first-order Taylor expansion about γ_0 :

$$\frac{\partial Q_{n,S}(\psi_0)}{\partial \psi} = \sum_{j=1}^n \left[\frac{\partial \rho'_{j,S}(\psi_0)}{\partial \psi} A_j \rho_{j,2S}(\psi_0) + \frac{\partial \rho'_{j,2S}(\psi_0)}{\partial \psi} A_j \rho_{j,S}(\psi_0) \right] \quad (7.20)$$

$$+ \frac{\partial^2 \tilde{Q}_{n,S}(\psi_0)}{\partial \psi \partial \gamma'} (\hat{\gamma} - \gamma_0),$$

where

$$\begin{aligned} & \frac{\partial^2 \tilde{Q}_{n,S}(\psi_0)}{\partial \psi \partial \gamma'} \\ &= \sum_{j=1}^n \left[\frac{\partial \rho'_{j,S}(\psi_0, \tilde{\gamma})}{\partial \psi} A_j \frac{\partial \rho_{j,2S}(\psi_0, \tilde{\gamma})}{\partial \gamma'} + (\rho'_{j,2S}(\psi_0, \tilde{\gamma}) A_j \otimes I_{p+q}) \frac{\partial \text{vec}(\partial \rho'_{j,S}(\psi_0, \tilde{\gamma}) / \partial \psi)}{\partial \gamma'} \right] \\ &+ \sum_{j=1}^n \left[\frac{\partial \rho'_{j,2S}(\psi_0, \tilde{\gamma})}{\partial \psi} A_j \frac{\partial \rho_{j,S}(\psi_0, \tilde{\gamma})}{\partial \gamma'} + (\rho'_{j,S}(\psi_0, \tilde{\gamma}) A_j \otimes I_{p+q}) \frac{\partial \text{vec}(\partial \rho'_{j,2S}(\psi_0, \tilde{\gamma}) / \partial \psi)}{\partial \gamma'} \right], \end{aligned}$$

$$\frac{\partial \text{vec}(\partial \rho'_{j,S}(\psi_0, \tilde{\gamma}) / \partial \theta)}{\partial \gamma'} = \frac{1}{S} \sum_{s=1}^S \tilde{x}_{js} \otimes \frac{\partial g(x_{js}; \theta_0)}{\partial \theta} \frac{\partial f_U(x_{js} - \tilde{\Gamma} W_j; \phi_0)}{\partial u'} \frac{(W_j \otimes I_k)'}{h(x_{js})},$$

$$\frac{\partial \text{vec}(\partial \rho'_{j,S}(\psi_0, \tilde{\gamma}) / \partial \phi)}{\partial \gamma'} = \frac{1}{S} \sum_{s=1}^S g(x_{js}; \theta_0) \tilde{x}_{js} \otimes \frac{\partial^2 f_U(x_{js} - \tilde{\Gamma} W_j; \phi_0)}{\partial \phi \partial u'} \frac{(W_j \otimes I_k)'}{h(x_{js})},$$

and $\|\tilde{\gamma} - \gamma_0\| \leq \|\hat{\gamma} - \gamma_0\|$. Now rewrite (7.20) as

$$\frac{\partial Q_{n,S}(\psi_0)}{\partial \psi} = 2D_{n,S} \sum_{j=1}^n T_{j,S},$$

where

$$D_{n,S} = \left(I_{p+q}, \frac{1}{2} \frac{\partial^2 \tilde{Q}_{n,S}(\psi_0)}{\partial \psi \partial \gamma'} \left(\sum_{j=1}^n W_j W_j' \otimes I_k \right)^{-1} \right)$$

and

$$T_{j,S} = \frac{1}{2} \left(\begin{array}{c} \frac{\partial \rho'_{j,S}(\psi_0)}{\partial \psi} A_j \rho_{j,2S}(\psi_0) + \frac{\partial \rho'_{j,2S}(\psi_0)}{\partial \psi} A_j \rho_{j,S}(\psi_0) \\ 2W_j \otimes (Z_j - \Gamma_0 W_j) \end{array} \right).$$

Then, analogous to (7.11), by ASSUMPTION 4 we can show that

$$\frac{1}{n} \frac{\partial^2 \tilde{Q}_{n,S}(\psi_0)}{\partial \psi \partial \gamma'} \xrightarrow{a.s.} 2E \left[\frac{\partial \rho'_1(\psi_0)}{\partial \psi} A_1 \frac{\partial \rho_1(\psi_0)}{\partial \gamma'} \right]$$

and hence

$$D_{n,S} \xrightarrow{a.s.} \left(I_{p+q}, E \left[\frac{\partial \rho'_1(\psi_0)}{\partial \psi} A_1 \frac{\partial \rho_1(\psi_0)}{\partial \gamma'} \right] (E W_1 W_1' \otimes I_k)^{-1} \right) = D. \quad (7.21)$$

Further, by the Central Limit Theorem we have

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n T_{j,S} \xrightarrow{L} N(0, C_S), \quad (7.22)$$

where

$$C_S = E T_{1,S} T_{1,S}' = \begin{pmatrix} C_{S,11} & C'_{S,21} \\ C_{S,21} & C_{S,22} \end{pmatrix},$$

$$\begin{aligned}
C_{S,11} &= \frac{1}{4} E \left[\left(\frac{\partial \rho'_{1,S}(\psi_0)}{\partial \psi} A_1 \rho_{1,2S}(\psi_0) + \frac{\partial \rho'_{1,2S}(\psi_0)}{\partial \psi} A_1 \rho_{1,S}(\psi_0) \right) \times \right. \\
&\quad \left. \left(\rho'_{1,2S}(\psi_0) A_1 \frac{\partial \rho_{1,S}(\psi_0)}{\partial \psi'} + \rho'_{1,S}(\psi_0) A_1 \frac{\partial \rho_{1,2S}(\psi_0)}{\partial \psi'} \right) \right] \\
&= \frac{1}{2} E \left[\frac{\partial \rho'_{1,S}(\psi_0)}{\partial \psi} A_1 \rho_{1,2S}(\psi_0) \rho'_{1,2S}(\psi_0) A_1 \frac{\partial \rho_{1,S}(\psi_0)}{\partial \psi'} \right] \\
&\quad + \frac{1}{2} E \left[\frac{\partial \rho'_{1,S}(\psi_0)}{\partial \psi} A_1 \rho_{1,2S}(\psi_0) \rho'_{1,S}(\psi_0) A_1 \frac{\partial \rho_{1,2S}(\psi_0)}{\partial \psi'} \right],
\end{aligned}$$

$$\begin{aligned}
C_{S,21} &= \frac{1}{2} E \left[(W_1 \otimes (Z_1 - \Gamma_0 W_1)) \left(\rho'_{1,2S}(\psi_0) A_1 \frac{\partial \rho_{1,S}(\psi_0)}{\partial \psi'} + \rho'_{1,S}(\psi_0) A_1 \frac{\partial \rho_{1,2S}(\psi_0)}{\partial \psi'} \right) \right] \\
&= E \left[(W_1 \otimes (Z_1 - \Gamma_0 W_1)) \rho'_1(\psi_0) A_1 \frac{\partial \rho_1(\psi_0)}{\partial \psi'} \right] \\
&= C_{21}
\end{aligned}$$

and

$$\begin{aligned}
C_{S,22} &= E \left[(W_1 \otimes (Z_1 - \Gamma_0 W_1)) (W_1 \otimes (Z_1 - \Gamma_0 W_1))' \right] \\
&= E \left[W_1 W_1' \otimes (Z_1 - \Gamma_0 W_1) (Z_1 - \Gamma_0 W_1)' \right] \\
&= C_{22}.
\end{aligned}$$

It follows from (7.21) and (7.22) we have

$$\frac{1}{2\sqrt{n}} \frac{\partial Q_{n,S}(\psi_0)}{\partial \psi} \xrightarrow{L} N(0, DC_S D'). \quad (7.23)$$

Finally, Theorem 4.1.2 follows from (7.17), (7.19) and (7.23).

7.4 Proof of Theorem 4.1

First, by Andrews (1995, Theorem 2), under ASSUMPTIONS 8 and 9, the kernel estimators in (4.7) and (4.8) satisfy

$$\sup_{v \in \mathbb{R}^k} \left| \hat{f}_V(v) - f_V(v) \right| = O_p \left(n^{-1/2} a_n^{-k-1} \right) + O_p \left(a_n^d \right) \quad (7.24)$$

and

$$\sup_{B_n} |\hat{m}_1(v) - m_1(v)| = O_p \left(n^{-1/2} a_n^{-k-1} b_n^{-2} \right) + O_p \left(a_n^d b_n^{-2} \right). \quad (7.25)$$

Further, for any $\theta \in \Theta$ and $u \in \mathbb{R}^k$,

$$\begin{aligned}
& (2\pi)^k \left| \hat{f}_U(u; \theta) - f_U(u; \theta) \right| \tag{7.26} \\
& \leq \left| \int_{C_n} e^{it'u} \frac{\hat{m}_1(t) - \tilde{m}_1(t)}{\tilde{g}(t; \theta)} dt \right| + \left| \int_{C_n^c} e^{it'u} \frac{\tilde{m}_1(t)}{\tilde{g}(t; \theta)} dt \right| \\
& \leq \int_{C_n} \left| \frac{\hat{m}_1(t) - \tilde{m}_1(t)}{\tilde{g}(t; \theta)} \right| dt + \int_{C_n^c} \left| \frac{\tilde{m}_1(t)}{\tilde{g}(t; \theta)} \right| dt \\
& \leq \frac{1}{c_n} \int_{C_n} \left| \hat{m}_1(t) - \tilde{m}_1(t) \right| dt + \int_{C_n^c} \sup_{\Theta} \left| \frac{\tilde{m}_1(t)}{\tilde{g}(t; \theta)} \right| dt \\
& \leq \frac{2^k}{c_n^{k+1}} \left(\int_{B_n} |\hat{m}_1(v) - m_1(v)| dv + \int_{B_n^c} |m_1(v)| dv \right) + \int_{C_n^c} \sup_{\Theta} \left| \frac{\tilde{m}_1(t)}{\tilde{g}(t; \theta)} \right| dt.
\end{aligned}$$

Since, by (7.24), $\lim_{n \rightarrow \infty} P(\inf_{B_n} |f_V(v)| \geq b_n/2) = 1$, with probability approaching one, we have

$$\begin{aligned}
\int_{B_n} |\hat{m}_1(v) - m_1(v)| dv & \leq \frac{2}{b_n} \int_{B_n} |\hat{m}_1(v) - m_1(v)| f_V(v) dv \\
& \leq \frac{2}{b_n} \sup_{B_n} |\hat{m}_1(v) - m_1(v)| = o_p(c_n^{k+1}),
\end{aligned}$$

where the last equality follows by (7.25) and condition of THEOREM 4.1. It follows from (7.26) and ASSUMPTION 10, that

$$\sup_{\theta \in \Theta} \sup_u \left| \hat{f}_U(u; \theta) - f_U(u; \theta) \right| = o_p(1). \tag{7.27}$$

Now, we write

$$\begin{aligned}
Q_n(\theta) & = \sum_{j=1}^n \rho_j(\theta)' A_j \rho_j(\theta) + 2 \sum_{j=1}^n \rho_j'(\theta) A_j (\hat{\rho}_j(\theta) - \rho_j(\theta)) \tag{7.28} \\
& \quad + \sum_{j=1}^n (\hat{\rho}_j(\theta) - \rho_j(\theta))' A_j (\hat{\rho}_j(\theta) - \rho_j(\theta)),
\end{aligned}$$

where $\rho_j(\theta) = Y_j \tilde{Z}_j - \int \tilde{x} g(x; \theta) f_U(x - \Gamma_0 W_j; \theta) dx$. Then, since

$$\begin{aligned}
\left| f_U(x - \Gamma_0 W_j; \theta) - f_U(x - \hat{\Gamma} W_j; \theta) \right| & = \left| \frac{1}{(2\pi)^k} \int \left(e^{-it' \Gamma_0 W_j} - e^{-it' \hat{\Gamma} W_j} \right) e^{it'x} \frac{\tilde{m}_1(t)}{\tilde{g}(t; \theta)} dt \right| \\
& \leq \frac{1}{(2\pi)^k} \int \left| e^{-it' \Gamma_0 W_j} - e^{-it' \hat{\Gamma} W_j} \right| \left| \frac{\tilde{m}_1(t)}{\tilde{g}(t; \theta)} \right| dt \\
& \leq \frac{2 \left\| \hat{\Gamma} - \Gamma_0 \right\| \|W_j\|}{(2\pi)^k} \int \|t\| \left| \frac{\tilde{m}_1(t)}{\tilde{g}(t; \theta)} \right| dt,
\end{aligned}$$

we have

$$\begin{aligned}
\|\hat{\rho}_j(\theta) - \rho_j(\theta)\| &\leq \int \|\tilde{x}g(x; \theta)\| \left| f_U(x - \Gamma_0 W_j; \theta) - \hat{f}_U(x - \hat{\Gamma} W_j; \theta) \right| dx \\
&\leq \int \|\tilde{x}g(x; \theta)\| \left| f_U(x - \Gamma_0 W_j; \theta) - f_U(x - \hat{\Gamma} W_j; \theta) \right| dx \\
&+ \int \|\tilde{x}g(x; \theta)\| \left| f_U(x - \hat{\Gamma} W_j; \theta) - \hat{f}_U(x - \hat{\Gamma} W_j; \theta) \right| dx \\
&\leq \frac{2 \|\hat{\Gamma} - \Gamma_0\| \|W_j\|}{(2\pi)^k} \int \|t\| \sup_{\Theta} \left| \frac{\tilde{m}_1(t)}{\tilde{g}(t; \theta)} \right| dt \int \sup_{\Theta} |g(x; \theta)| (\|x\| + 1) dx \\
&+ \sup_{\theta \in \Theta} \sup_u \left| \hat{f}_U(u; \theta) - f_U(u; \theta) \right| \int \sup_{\Theta} |g(x; \theta)| (\|x\| + 1) dx \\
&= \alpha_1 \|\hat{\Gamma} - \Gamma_0\| \|W_j\| + \alpha_2 \sup_{\theta \in \Theta} \sup_u \left| \hat{f}_U(u; \theta) - f_U(u; \theta) \right|,
\end{aligned}$$

where α_1 and α_2 are positive constants. Further, by Cauchy-Schwarz inequality,

$$\begin{aligned}
\left(E \sup_{\Theta} \|A_1\| (\|W_1\| + 1) \|\rho_1(\theta)\| \right)^2 &\leq E \|A_1\| (\|W_1\| + 1)^2 E \|A_1\| \sup_{\Theta} \|\rho_1(\theta)\|^2 \\
&\leq 2E \|A_1\| (\|W_1\|^2 + 1) \left[E \|A_1\| (\|Y_1\|^2 + \|Y_1 Z_1\|^2) \right. \\
&\quad \left. + E \|A_1\| \left(\int \sup_{\Theta} \|\tilde{x}g(x; \theta) f_U(x - \Gamma_0 W_1; \theta)\| dx \right)^2 \right] \\
&< \infty,
\end{aligned}$$

which implies $\frac{1}{n} \sum_{j=1}^n \|A_j\| (\|W_j\| + 1) \sup_{\Theta} \|\rho_j(\theta)\| = O_p(1)$. Therefore we have

$$\begin{aligned}
&\sup_{\Theta} \left| \frac{1}{n} \sum_{j=1}^n \rho_j(\theta)' A_j (\hat{\rho}_j(\theta) - \rho_j(\theta)) \right| \tag{7.29} \\
&\leq \frac{1}{n} \sum_{j=1}^n \|A_j\| \sup_{\Theta} \|\rho_j(\theta)\| \|\hat{\rho}_j(\theta) - \rho_j(\theta)\| \\
&\leq \left(\alpha_1 \|\hat{\Gamma} - \Gamma_0\| + \alpha_2 \sup_{\theta \in \Theta} \sup_{u \in \mathbb{R}^k} \left| \hat{f}_U(u; \theta) - f_U(u; \theta) \right| \right) \frac{1}{n} \sum_{j=1}^n \|A_j\| (\|W_j\| + 1) \sup_{\Theta} \|\rho_j(\theta)\| \\
&= o_p(1),
\end{aligned}$$

and analogously,

$$\begin{aligned}
&\sup_{\Theta} \left| \frac{1}{n} \sum_{j=1}^n (\hat{\rho}_j(\theta) - \rho_j(\theta))' A_j (\hat{\rho}_j(\theta) - \rho_j(\theta)) \right| \tag{7.30} \\
&\leq \frac{1}{n} \sum_{j=1}^n \|A_j\| \sup_{\Theta} \|\hat{\rho}_j(\theta) - \rho_j(\theta)\|^2 \\
&\leq \left(2\alpha_1^2 \|\hat{\Gamma} - \Gamma_0\|^2 + 2\alpha_2^2 \sup_{\theta \in \Theta} \sup_{u \in \mathbb{R}^k} \left| \hat{f}_U(u; \theta) - f_U(u; \theta) \right|^2 \right) \frac{1}{n} \sum_{j=1}^n \|A_j\| (\|W_j\|^2 + 1) \\
&= o_p(1).
\end{aligned}$$

Since

$$\begin{aligned}
E \sup_{\Theta} |\rho_1'(\theta) A_1 \rho_1(\theta)| &\leq E \|A_1\| \sup_{\Theta} \|\rho_1(\theta)\|^2 \\
&\leq 2E \|A_1\| (\|Y_1\|^2 + \|Y_1 Z_1\|^2) \\
&+ E \|A_1\| \left(\int \sup_{\Theta} \|\tilde{x}g(x; \theta) f_U(x - \Gamma_0 W_1; \theta)\| dx \right)^2 \\
&< \infty,
\end{aligned}$$

by the ULLN we have

$$\sup_{\Theta} \left| \frac{1}{n} \sum_{j=1}^n \rho_j(\theta)' A_j \rho_j(\theta) - Q(\theta) \right| = o_p(1), \quad (7.31)$$

where $Q(\theta) = E \rho_1(\theta)' A_1 \rho_1(\theta)$. It follows from (7.28) - (7.31) that $\frac{1}{n} Q_n(\theta) \xrightarrow{P} Q(\theta)$ uniformly in $\theta \in \Theta$. Analogous to the proof of THEOREM 2.1, by ASSUMPTION 12 we can show that $Q(\theta)$ attains a unique minimum at $\theta_0 \in \Theta$. Therefore $\hat{\theta}_n \xrightarrow{P} \theta_0$ follows from Amemiya (1973, Lemma 3).

Furthermore, THEOREM 4.1.2 follows from (7.27) and Amemiya (1973, Lemma 4).

To prove THEOREM 4.1.3, note that for any $t \in \mathbb{R}^k$, since $\tilde{g}(t; \theta_0) \neq 0$ and $\tilde{g}(t; \theta)$ is continuous in θ , there exists an open neighborhood Θ_t of θ_0 in Θ , such that $\inf_{\theta \in \Theta_t} |\tilde{g}(t; \theta)| \geq c > 0$. Therefore, for any $\theta \in \Theta_t$,

$$\left| \frac{\hat{m}_1(t) - \tilde{m}_1(t)}{\tilde{g}(t; \theta)} \right| \leq \frac{1}{c} \int_{B_n} |\hat{m}_1(v) - m_1(v)| dv + \frac{1}{c} \int_{B_n^c} |m_1(v)| dv,$$

which implies that $\sup_{\theta \in \Theta_t} |\hat{f}_U(t; \theta) - \tilde{f}_U(t; \theta)| = o_p(1)$. The result follows then from Amemiya (1973, Lemma 4).

7.5 Proof of Theorem 4.2

By ASSUMPTION 13 the first derivative $\partial Q_{n,S}(\theta)/\partial\theta$ exists and has the first order Taylor expansion in a neighborhood of θ_0 . Since $\partial Q_{n,S}(\hat{\theta}_n)/\partial\theta = 0$ and $\hat{\theta}_{n,S} \xrightarrow{P} \theta_0$, for sufficiently large n , we have

$$0 = \frac{\partial Q_{n,S}(\theta_0)}{\partial\theta} + \frac{\partial^2 Q_{n,S}(\tilde{\theta})}{\partial\theta\partial\theta'} (\hat{\theta}_{n,S} - \theta_0), \quad (7.32)$$

where $\tilde{\theta}$ satisfies $\|\tilde{\theta} - \theta_0\| \leq \|\hat{\theta}_{n,S} - \theta_0\|$. The first derivative in (7.32) is given by

$$\frac{\partial Q_{n,S}(\theta)}{\partial\theta} = \sum_{j=1}^n \left[\frac{\partial \hat{\rho}'_{j,S}(\theta)}{\partial\theta} A_j \hat{\rho}_{j,2S}(\theta) + \frac{\partial \hat{\rho}'_{j,2S}(\theta)}{\partial\theta} A_j \hat{\rho}_{j,S}(\theta) \right],$$

where

$$\frac{\partial \hat{\rho}'_{j,S}(\theta)}{\partial\theta} = -\frac{1}{S} \sum_{j=1}^S \frac{\tilde{x}'_{j_s}}{h(x_{j_s})} \left[\frac{\partial g(x_{j_s}; \theta)}{\partial\theta} \hat{f}_U(x_{j_s} - \hat{\Gamma} W_j; \theta) + g(x_{j_s}; \theta) \frac{\partial \hat{f}_U(x_{j_s} - \hat{\Gamma} W_j; \theta)}{\partial\theta} \right]$$

and

$$\frac{\partial \hat{f}_U(x_{js} - \hat{\Gamma}W_j; \theta)}{\partial \theta} = -\frac{1}{(2\pi)^k} \int_{C_n} e^{it'(x_{js} - \hat{\Gamma}W_j)} \frac{\hat{m}_1(t)}{\hat{g}^2(t; \theta)} \frac{\partial \tilde{g}(t; \theta)}{\partial \theta} dt.$$

The second derivative in (7.32) is given by

$$\begin{aligned} \frac{\partial^2 Q_{n,S}(\theta)}{\partial \theta \partial \theta'} &= \sum_{j=1}^n \left[\frac{\partial \hat{\rho}'_{j,S}(\theta)}{\partial \theta} A_j \frac{\partial \hat{\rho}_{j,2S}(\theta)}{\partial \theta'} + (\hat{\rho}'_{j,2S}(\theta) A_j \otimes I_p) \frac{\partial \text{vec}(\partial \hat{\rho}'_{j,S}(\theta) / \partial \theta)}{\partial \theta'} \right] \\ &+ \sum_{j=1}^n \left[\frac{\partial \hat{\rho}'_{j,2S}(\theta)}{\partial \theta} A_j \frac{\partial \hat{\rho}_{j,S}(\theta)}{\partial \theta'} + (\hat{\rho}'_{j,S}(\theta) A_j \otimes I_p) \frac{\partial \text{vec}(\partial \hat{\rho}'_{j,2S}(\theta) / \partial \theta)}{\partial \theta'} \right], \end{aligned}$$

where $\partial \text{vec}(\partial \hat{\rho}'_{j,S}(\tilde{\theta}) / \partial \theta) / \partial \theta'$ is given by

$$\begin{aligned} &-\frac{1}{S} \sum_{j=1}^S \left[\left(\tilde{x}_{js} \otimes \frac{\partial^2 g(x_{js}; \theta)}{\partial \theta \partial \theta'} \right) \hat{f}_U(x_{js} - \hat{\Gamma}W_j; \theta) + \left(\tilde{x}_{js} \otimes \frac{\partial g(x_{js}; \theta)}{\partial \theta} \right) \frac{\partial \hat{f}_U(x_{js} - \hat{\Gamma}W_j; \theta)}{\partial \theta'} \right. \\ &\left. + \left(\tilde{x}_{js} \otimes \frac{\partial \hat{f}_U(x_{js} - \hat{\Gamma}W_j; \theta)}{\partial \theta} \right) \frac{\partial g(x_{js}; \theta)}{\partial \theta'} + \left(\tilde{x}_{js} \otimes \frac{\partial^2 \hat{f}_U(x_{js} - \hat{\Gamma}W_j; \theta)}{\partial \theta \partial \theta'} \right) g(x_{js}; \theta) \right] \frac{1}{h(x_{js})} \end{aligned}$$

and $\partial^2 \hat{f}_U(x_{js} - \hat{\Gamma}W_j; \theta) / \partial \theta \partial \theta'$ is given by

$$-\frac{1}{(2\pi)^k} \int_{C_n} e^{it'(x_{js} - \hat{\Gamma}W_j)} \frac{\hat{m}_1(t)}{\hat{g}^2(t; \theta)} \left[\frac{\partial^2 g(x_{js}; \theta)}{\partial \theta \partial \theta'} - \frac{2}{\hat{g}(t; \theta)} \frac{\partial \tilde{g}(t; \theta)}{\partial \theta} \frac{\partial \tilde{g}(t; \theta)}{\partial \theta'} \right] dt.$$

Since $\hat{\rho}_{j,S}(\theta)$ and its derivatives are continuous in θ by ASSUMPTION 13, analogous to the proof of (7.19) it can be shown by ASSUMPTION 13 that

$$\begin{aligned} \frac{1}{n} \frac{\partial^2 Q_{n,S}(\tilde{\theta})}{\partial \theta \partial \theta'} &\xrightarrow{P} E \left[\frac{\partial \rho'_{j,S}(\theta_0)}{\partial \theta} A_j \frac{\partial \rho_{j,2S}(\theta_0)}{\partial \theta'} + (\rho'_{j,2S}(\theta_0) A_j \otimes I_p) \frac{\partial \text{vec}(\partial \rho'_{j,S}(\theta_0) / \partial \theta)}{\partial \theta'} \right] \\ &= E \left[\frac{\partial \rho'_j(\theta_0)}{\partial \theta} A_j \frac{\partial \rho_j(\theta_0)}{\partial \theta'} \right] = B, \end{aligned} \quad (7.33)$$

In the following let $\bar{f}_V(v), \bar{m}_1(v)$ be defined similarly as $\hat{f}_V(v), \hat{m}_1(v)$ in (4.7) and (4.8), but with $\hat{\Gamma}$ substituted by Γ_0 . Similarly, $\bar{f}_U(u; \theta)$ and $\bar{\rho}_{j,S}(\theta)$ are used for the same situation. Then, by ASSUMPTION 13, $\partial Q_{n,S}(\theta_0) / \partial \theta$ has the first-order Taylor expansion about γ_0 :

$$\begin{aligned} \frac{\partial Q_{n,S}(\theta_0)}{\partial \theta} &= \sum_{j=1}^n \left[\frac{\partial \bar{\rho}'_{j,S}(\theta_0)}{\partial \theta} A_j \bar{\rho}_{j,2S}(\theta_0) + \frac{\partial \bar{\rho}'_{j,2S}(\theta_0)}{\partial \theta} A_j \bar{\rho}_{j,S}(\theta_0) \right] \\ &+ \frac{\partial^2 Q_{n,S}(\theta_0)}{\partial \theta \partial \gamma'} (\hat{\gamma} - \gamma_0) + o_p(\sqrt{n}), \end{aligned} \quad (7.34)$$

where

$$\begin{aligned} \frac{\partial^2 Q_{n,S}(\theta)}{\partial \theta \partial \gamma'} &= \sum_{j=1}^n \left[\frac{\partial \bar{\rho}'_{j,S}(\theta)}{\partial \theta} A_j \frac{\partial \bar{\rho}_{j,2S}(\theta)}{\partial \gamma'} + (\bar{\rho}'_{j,2S}(\theta) A_j \otimes I_{p+q}) \frac{\partial \text{vec}(\partial \bar{\rho}'_{j,S}(\theta) / \partial \theta)}{\partial \gamma'} \right] \\ &+ \sum_{j=1}^n \left[\frac{\partial \bar{\rho}'_{j,2S}(\theta)}{\partial \theta} A_j \frac{\partial \bar{\rho}_{j,S}(\theta)}{\partial \gamma'} + (\bar{\rho}'_{j,S}(\theta) A_j \otimes I_{p+q}) \frac{\partial \text{vec}(\partial \bar{\rho}'_{j,2S}(\theta) / \partial \theta)}{\partial \gamma'} \right], \end{aligned}$$

$$\frac{\partial \bar{\rho}'_{j,S}(\theta)}{\partial \gamma'} = -\frac{1}{S} \sum_{s=1}^S \frac{\tilde{x}_{js} g(x_{js}; \theta)}{h(x_{js})} \frac{\partial \bar{f}_U(x_{js} - \Gamma_0 W_j; \theta)}{\partial \gamma'},$$

and

$$\frac{\partial \bar{f}_U(x_{js} - \Gamma_0 W_j; \theta)}{\partial \gamma'} = \frac{1}{(2\pi)^k} \int_{C_n} \frac{e^{it'(x_{js} - \Gamma_0 W_j)}}{\tilde{g}(t; \theta)} \left[\frac{\partial \tilde{m}_1(t)}{\partial \gamma'} - i \tilde{m}_1(t) (W_j \otimes t)' \right] dt.$$

Here

$$\begin{aligned} \frac{\partial \tilde{m}_1(t)}{\partial \gamma'} &= \int_{B_n} e^{-it'v} \frac{\partial \bar{m}_1(v)}{\partial \gamma'} dv, \\ \frac{\partial \bar{m}_1(v)}{\partial \gamma'} &= \frac{1}{\bar{f}_V(v) n a_n} \sum_{j=1}^n (\bar{m}_1(v) - Y_j) \frac{\partial K_a(v - V_j)}{\partial v'} (W_j \otimes I_k)' \end{aligned}$$

and

$$K_a(v - V_j) = \frac{1}{a_n^k} K\left(\frac{v - V_j}{a_n}\right).$$

In the above we have slightly abused notation by using $V_j = \Gamma_0 W_j$. This will not cause confusion, since $\hat{\Gamma} W_j$ will not appear any more subsequently. Again, analogous to (7.11), by ASSUMPTIONS 13 and 14 we can show that

$$\frac{1}{n} \frac{\partial^2 Q_{n,S}(\theta_0)}{\partial \theta \partial \gamma'} \xrightarrow{P} 2E \left[\frac{\partial \rho'_j(\theta_0)}{\partial \theta} A_j \frac{\partial \rho_j(\theta_0)}{\partial \gamma'} \right].$$

In the following, because all functions of parameters are evaluated at θ_0 , we will omit it to further simplify notation. In addition, let $u_{js} = x_{js} - \Gamma_0 W_j$, $g_{js} = g(x_{js}; \theta_0)$, $h_{js} = h(x_{js})$ and \bar{f}_{js} be similarly defined. Now we express the first term of (7.34) in terms of $\xi_{1j}, \xi_{2j}, \xi_{3j}, \xi_{4j}$. To this end, we write

$$\begin{aligned} \sum_{j=1}^n \frac{\partial \bar{\rho}'_{j,S}}{\partial \theta} A_j \bar{\rho}_{j,2S} &= \sum_{j=1}^n \frac{\partial \rho'_{j,S}}{\partial \theta} A_j \rho_{j,2S} + \sum_{j=1}^n \frac{\partial \bar{\rho}'_{j,S}}{\partial \theta} A_j (\bar{\rho}_{j,2S} - \rho_{j,2S}) \\ &+ \sum_{j=1}^n \left(\frac{\partial \bar{\rho}'_{j,S}}{\partial \theta} - \frac{\partial \rho'_{j,S}}{\partial \theta} \right) A_j \rho_{j,2S} + \sum_{j=1}^n \left(\frac{\partial \bar{\rho}'_{j,S}}{\partial \theta} - \frac{\partial \rho'_{j,S}}{\partial \theta} \right) A_j (\bar{\rho}_{j,2S} - \rho_{j,2S}). \end{aligned} \quad (7.35)$$

Then by ASSUMPTION 15 we have

$$\bar{\rho}_{j,S} - \rho_{j,S} = -\frac{1}{S} \sum_{s=1}^S \frac{\tilde{x}_{js} g_{js}}{h_{js}} (\bar{f}_{js} - f_{js}^c) + o_p(\sqrt{n}),$$

where

$$\begin{aligned} \bar{f}_{js} - f_{js}^c &= \frac{1}{(2\pi)^k} \int_{C_n} \frac{e^{-it'u_{js}}}{\tilde{g}(t; \theta_0)} \int_{B_n} e^{-it'v} (\bar{m}_1(v) - m_1(v)) dv dt \\ &= \frac{1}{n(2\pi)^k} \sum_{\ell=1}^n \int_{C_n} \frac{e^{-it'u_{js}}}{\tilde{g}(t; \theta_0)} \int_{B_n} \frac{e^{-it'v}}{\bar{f}_V(v)} (Y_\ell - m_1(v)) K_a(v - V_\ell) dv dt \\ &= \frac{1}{n(2\pi)^k} \sum_{\ell=1}^n \int_{C_n} \frac{e^{-it'u_{js}}}{\tilde{g}(t; \theta_0)} \int_{B_n} \frac{e^{-it'v} \eta_\ell}{\bar{f}_V(v)} dv dt \\ &+ \frac{1}{n(2\pi)^k} \sum_{\ell=1}^n \int_{C_n} \frac{e^{-it'u_{js}}}{\tilde{g}(t; \theta_0)} \int_{B_n} \frac{e^{-it'v}}{\bar{f}_V(v)} E(Y_\ell - m_1(v)) K_a(v - V_\ell) dv dt \end{aligned} \quad (7.36)$$

and η_ℓ is defined in (4.15). Since, by (7.24), $\lim_{n \rightarrow \infty} P(\inf_{B_n} |\bar{f}_V(v)| \geq b_n/2) = 1$ and $\lim_{n \rightarrow \infty} P(\inf_{B_n} |f_V(v)| \geq b_n/2) = 1$, with probability approaching one, the absolute value of the second term of the last equation of (7.36) satisfies

$$\begin{aligned} & \left| \int_{C_n} \frac{e^{-it'u_{js}}}{\tilde{g}(t; \theta_0)} \int_{B_n} \frac{e^{-it'v}}{\bar{f}_V(v)} E(Y_\ell - m_1(v)) K_a(v - V_\ell) dv dt \right| \\ & \leq \int_{C_n} \frac{1}{|\tilde{g}(t; \theta_0)|} \int_{B_n} \frac{1}{|\bar{f}_V(v)|} |E(Y_\ell - m_1(v)) K_a(v - V_\ell)| dv dt \\ & \leq \frac{4}{b_n^2 c_n^{k+1}} \int |E(Y_\ell - m_1(v)) K_a(v - V_\ell)| f_V(v) dv \\ & = O_p(a_n^d b_n^{-2} c_n^{-k-1}) = o_p(n^{-1/2}), \end{aligned}$$

where the last two equalities follow from ASSUMPTIONS 13, 14 and theorem conditions. Furthermore, since $\sum \eta_\ell = O_p(\sqrt{n})$,

$$\frac{1}{n} \sum_{\ell=1}^n \int_{C_n} \frac{e^{-it'u_{js}}}{\tilde{g}(t; \theta_0)} \int_{B_n} e^{-it'v} \left(\frac{1}{\bar{f}_V(v)} - \frac{1}{f_V(v)} \right) \eta_\ell dv dt = o_p(n^{-1/2})$$

and, similarly,

$$\frac{1}{n} \sum_{\ell=1}^n \int_{C_n} \frac{e^{-it'u_{js}}}{\tilde{g}(t; \theta_0)} \int_{B_n^c} \frac{e^{-it'v}}{f_V(v)} \eta_\ell dv dt = o_p(n^{-1/2}).$$

It follows that

$$\bar{f}_{js} - f_{js}^c = \frac{1}{n(2\pi)^k} \sum_{\ell=1}^n \int_{C_n} \frac{e^{-it'u_{js}}}{\tilde{g}(t; \theta_0)} \int \frac{e^{-it'v} \eta_\ell}{f_V(v)} dv dt + o_p(n^{-1/2}).$$

Therefore, the second term on the right-hand side of (7.35) equals

$$\begin{aligned} & - \frac{1}{nS(2\pi)^k} \sum_{j=1}^n \frac{\partial \rho'_{j,2S}}{\partial \theta} A_j \sum_{s=1}^S \frac{\tilde{x}_{js} g_{js}}{h_{js}} \sum_{\ell=1}^n \int_{C_n} \frac{e^{-it'u_{js}}}{\tilde{g}(t; \theta_0)} \int \frac{e^{-it'v} \eta_\ell}{f_V(v)} dv dt + o_p(\sqrt{n}) \\ & = - \frac{1}{nS(2\pi)^k} \sum_{j=1}^n \sum_{\ell=1}^n \frac{\partial \rho'_{j,2S}}{\partial \theta} A_j \sum_{s=1}^S \frac{\tilde{x}_{js} g_{js}}{h_{js}} \int_{C_n} \frac{e^{-it'u_{js}}}{\tilde{g}(t; \theta_0)} \int \frac{e^{-it'v} \eta_\ell}{f_V(v)} dv dt + o_p(\sqrt{n}) \\ & = nU_n + o_p(\sqrt{n}), \end{aligned}$$

where U_n is a so-called U -Statistic of degree two. By Theorem 1 of Section 3.2.1 of Lee (1990, pp. 76), $nU_n = \sum_{j=1}^n \xi_{2j} + o_p(\sqrt{n})$. Further, note that in $\bar{\rho}_{j,2S}$ and $\rho_{j,2S}$ only $\tilde{x}_{js} g_{js}$ and $\tilde{g}(t; \theta_0)$ involve θ , completely analogously it can be shown that the third term on the right-hand side of (7.35) equals $\sum_{j=1}^n \xi_{3j} + o_p(\sqrt{n})$. Finally, it is obvious that the fourth term is $o_p(\sqrt{n})$. It follows that

$$\sum_{j=1}^n \frac{\partial \bar{\rho}'_{j,S}(\theta_0)}{\partial \theta} A_j \bar{\rho}_{j,2S}(\theta_0) = \sum_{j=1}^n (\xi_{2j} + \xi_{2j} + \xi_{3j}) + o_p(\sqrt{n}),$$

where $\xi_{1j}, \xi_{2j}, \xi_{3j}$ are given in (4.12) - (4.14). Then we can rewrite (7.34) as

$$\frac{\partial Q_{n,S}(\theta_0)}{\partial \theta} = 2D_{n,S} \sum_{j=1}^n \xi_j + o_p(\sqrt{n}),$$

where

$$D_{n,S} = \left(I_p, I_p, I_p, \frac{1}{n} \frac{\partial^2 Q_{n,S}(\theta_0)}{\partial \theta \partial \gamma'} \left[\left(\frac{1}{n} \sum_{j=1}^n W_j W_j' \right)^{-1} \otimes I_k \right] \right) \xrightarrow{P} D.$$

Since $E[W_j \otimes (Z_j - \Gamma_0 W_j)] = 0$, $E\eta_j(v) = 0$ and $E\left(\frac{\partial \rho'_{j,2S}}{\partial \theta} A_j \rho_{j,S}\right) = 0$, we have $E\xi_j = 0$. Furthermore, by model assumption, ASSUMPTION 13, the covariance matrix $C = E\xi_j \xi_j'$ is finite. It follows by Lindeberg-Levy CLT and Slutsky Theorem,

$$\frac{1}{2\sqrt{n}} \frac{\partial Q_{n,S}(\theta_0)}{\partial \theta} \xrightarrow{L} N(0, DCD'), \quad (7.37)$$

where D and C are as in Theorem 5.2. The theorem follows then from (7.32), (7.33), (7.37) and ASSUMPTION 16.

7.6 Proof of Theorem 5.2

Let F_W and F_X respectively denote the distribution function of W and X . Then by (4.1) and Fubini Theorem,

$$\begin{aligned} & E_W |m_{1,n}(\Gamma_0 W) - E(Y|W)| \\ &= \int |m_{1,n}(\Gamma_0 w) - E(Y|w)| dF_W(w) \\ &= \int \left| \int g(\Gamma_0 w + u; \theta_0) \mathbf{1}(\|\Gamma_0 w + u\| \geq T_n) dF_U(u) \right| dF_W(w) \\ &\leq \int \int |g(\Gamma_0 w + u; \theta_0)| \mathbf{1}(\|\Gamma_0 w + u\| \geq T_n) dF_U(u) dF_W(w) \\ &= \int \int |g(x; \theta_0)| \mathbf{1}(\|x\| \geq T_n) dF_U(x - \Gamma_0 w) dF_W(w) \\ &= \int |g(x; \theta_0)| \mathbf{1}(\|x\| \geq T_n) dF_X(x), \end{aligned}$$

where the last equality follows from (1.3). Since $E|g(X; \theta_0)| < \infty$, the right-hand side of the last equation tends to zero as $T_n \rightarrow \infty$, which implies (5.5). The proof of (5.6) is analogous.

Acknowledgement The authors are grateful to the anonymous referees for their helpful comments and suggestions. The main results in Sections 4 and 5 are circulated before as a working paper under the title "A simulated semiparametric estimation of nonlinear errors-in-variables models". These results were obtained while Liqun Wang was visiting the Department of Economics, University of Southern California, from 1993 to 1995. He wishes to thank the department for its hospitality. This work was partially supported by National Science Foundation grants SBR

94-09540 and SBR 96-19330, and by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

- [1] Abarin, T. and Wang, L. (2006). Comparison of GMM with second-order least squares estimator in nonlinear models. *Far East Journal of Theoretical Statistics*, 20, 179-196.
- [2] 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.
- [3] Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica*, 41, 997-1016.
- [4] Amemiya, T. (1974). The Nonlinear Two-Stage Least-Squares Estimator. *Journal of Econometrics*, 2, 105-110.
- [5] Amemiya, Y. (1985). Instrumental variable estimator for the nonlinear errors-in-variables model. *Journal of Econometrics*, 28, 273-289.
- [6] Amemiya, Y. (1990). Two-stage instrumental variable estimators for the nonlinear errors-in-variables model. *Journal of Econometrics*, 44, 311-332.
- [7] Amemiya, Y. and W. A. Fuller (1988). Estimation for the Nonlinear Functional Relationship. *Annals of Statistics*, 16, 147-160.
- [8] Andrews, D.W.K. (1995). Nonparametric Kernel Estimation for Semiparametric Models. *Econometric Theory*, 11, 560-596.
- [9] Carroll, R.J., Maca, J.D., and Ruppert, D. (1999). Nonparametric regression in the presence of measurement error. *Biometrika*, 86, 541-554.
- [10] Carroll, R.J., Ruppert, D. and Stefanski, L.A. (1995). *Measurement Error in Nonlinear Models*. Chapman & Hall, London.
- [11] Carroll, R.J., Ruppert, D., Crainiceanu, C.M., Tosteson, T.D. and Karagas, M.R. (2004). Nonlinear and nonparametric regression and instrumental variables. *Journal of American Statistical Association*, 99, 736-750.
- [12] Cheng, C. and Schneeweiss, H. (1998). Polynomial regression with errors in the variables. *Journal of Royal Statistical Society*, B60, 189-199.
- [13] Delaigle, A. (2007). Nonparametric density estimation from data with a mixture of Berkson and classical errors. *Canadian Journal of Statistics*, 35, 89-104.

- [14] Fan, J. and Truong, Y. K. (1993). Nonparametric regression with errors in variables. *Annals of Statistics*, 21, 1900-1925.
- [15] Fuller, A. W. (1987). *Measurement Error Models*. Wiley, New York.
- [16] Gallant, A. R. (1987). *Nonlinear Statistical Model*, Wiley, New York.
- [17] Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50, 1029-1054.
- [18] Hausman, J., W. Newey, H. Ichimura, and J. Powell (1991). Identification and estimation of polynomial errors-in-variables models. *Journal of Econometrics*, 50, 273-295.
- [19] Hausman, J., W. Newey, and J. Powell (1995). Nonlinear errors in variables: estimation of some Engel curves. *Journal of Econometrics*, 65, 205-233.
- [20] Hsiao, C. (1983). Identification. In Z. Griliches and M. D. Intriligator eds., *Handbook of Econometrics*, Vol. I. North-Holland, Amsterdam.
- [21] Hsiao, C. (1989). Consistent Estimation for Some Nonlinear Errors-Variables Models. *Journal of Econometrics*, 41, 159-185.
- [22] Hsiao, C. (1992). Nonlinear Latent Variable Models. In L. Matyas and P. Sevestre eds., *The Econometrics of Panel Data*, 242-261.
- [23] Huang, S., and Huwang, L. (2001). On the polynomial structural relationship. *The Canadian Journal of Statistics*, 29, 495-512.
- [24] Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.*, 40, 633-643.
- [25] Lee, A. J. (1990). *U-Statistics: Theory and Practice*. Marcel Dekker, New York.
- [26] Li, T. (2002). Robust and consistent estimation of nonlinear errors-in-variables models. *Journal of Econometrics*, 110, 1-26.
- [27] Lukacs, E. (1970). *Characteristic Functions*, Hafner Publishing Company, New York.
- [28] Magnus, J. R. and Neudecker, H. (1988). *Matrix Differential Calculus with Application in Statistics and Econometrics*. Wiley, New York.
- [29] Newey, (2001). Flexible simulated moment estimation of nonlinear errors-in-variables models. *Review of Economeics and Statistics*, 83, 616-627.
- [30] Robinson, P. M. (1988). Root-N-Consistent Semiparametric Regression. *Econometrica*, 56, 931-954.

- [31] Stefanski, L. A. and Carroll, R. J. (1985). Covariate measurement error in logistic regression. *Ann. Statist.*, 13, 1335-1351.
- [32] Schennach, S. M. (2004). Estimation of nonlinear models with measurement error. *Econometrica*, 72, 33-75.
- [33] Schennach, S. M. (2007). Instrumental variable estimation of nonlinear errors-in-variables models. *Econometrica*, 75, 201-239.
- [34] Taupin, M. (2001). Semi-parametric estimation in the nonlinear structural errors-in-variables model. *Ann. Statist.*, 29, 66-93.
- [35] Walker, J. S. (1988). *Fourier Analysis*. Oxford University Press, New York.
- [36] Wang, L. (1998). Estimation of censored linear errors-in-variables models. *Journal of Econometrics*, 84, 383-400.
- [37] Wang, L. (2002). A simple adjustment for measurement errors in some limited dependent variable models. *Statistics and Probability Letters*, 58, 427-433.
- [38] Wang, L. (2004). Estimation of Nonlinear Models with Berkson Measurement Errors. *Annals of Statistics*, 32, 2559-2579.
- [39] Wang, L. and Hsiao, C. (1995). A simulated semiparametric estimation of nonlinear errors-in-variables models. *Working Paper*, Department of Economics, University of Southern California.
- [40] Wang, L. and Hsiao, C. (2007). Two-stage estimation of limited dependent variable models with errors-in-variables. *Econometrics Journal*, 10, 426-438.
- [41] Weiss, A. A. (1993). Some aspects of measurement error in a censored regression model. *Journal of Econometrics*, 56, 169-188.
- [42] Wolter, K. M. and W. A. Fuller (1982). Estimation of Nonlinear Errors-in-Variables Models. *Annals of Statistics*, 10, 539-548.
- [43] Zeidler, E. (1986). *Nonlinear Functional Analysis and Its Applications I*. Springer-Verlag, New York.