Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Click on the ‘Q’ link to go to the location in the proof.

<table>
<thead>
<tr>
<th>Location in article</th>
<th>Query / remark click on the Q link to go</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Uncited references: This section comprises references that occur in the reference list but not in the body of the text. Please cite each reference in the text or, alternatively, delete it. Any reference not dealt with will be retained in this section.</td>
</tr>
</tbody>
</table>

Thank you for your assistance.
Method of moments estimation and identifiability of semiparametric nonlinear errors-in-variables models

Liqun Wang, Cheng Hsiao

1. Introduction

Measurement error occurs frequently (e.g., Aigner et al., 1984; Fuller, 1987; 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., Amemiya, 1985, 1990; 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., Wolter and Fuller, 1982; Amemiya, 1985, 1990; Stefanski and Carroll, 1985; 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). Later Li (2002) and Schennach (2004) studied models with replicate observations, while Schennach (2007) used the 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, 1995; Huang and Huwung, 2001), and limited dependent variable models (Weiss, 1993; Wang, 1998, 2002; 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 et al., 1999; Taupin, 2001; Carroll et al., 2004; Delaigle, 2007).

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

\[ Y = g(X; \theta_0) + \varepsilon, \tag{1.1} \]

where \( Y \in \mathbb{R}, X \in \mathbb{R}^k, \varepsilon \) 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

\[ Z = X + \delta, \tag{1.2} \]

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

\[ X = \Gamma_0 W + U, \tag{1.3} \]

where \( U \) is a vector of independent and identically distributed (i.i.d.) random errors with zero mean and finite variance.

© 2011 Elsevier B.V. All rights reserved.
where $\Gamma_0$ 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 $E(e | X, Z, W) = 0$ and $E(\delta | X, W) = 0$. The functional forms of the distributions of $X$, $\varepsilon$ and $\delta$ 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 $\theta_0$, $\Gamma_0$ and the distribution $F_0$ of $U$.

Model (1.1)–(1.3) 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, the integrability condition was relaxed by Schennach (2007) who used a generalized function technique and achieved more general identifiability conditions. In addition, assuming the model to be identifiable, Newey (2001) derived a consistent estimator when $F_0(u)$ belongs to a parametric family and a consistent semiparametric estimator when $F_0(u)$ 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 the general $g(x; \theta_0)$ which is not necessarily integrable.

In particular, for the case of a parametric distribution $F_0(u; \phi_0)$ we propose method of moments estimators for $\theta$ and $\phi$ 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 $F_0, (u)$, we combine the nonparametric technique with Fourier deconvolution to obtain a root-$n$ consistent estimator for $\theta$ 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 $\theta$. Section 5 gives a rank condition for model identifiability and illustrative examples. Finally, conclusions and discussions 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 a nonlinear errors-in-variables model under the assumption that the distribution $f_0(u; \phi_0)$ of $U$ is known up to a vector of unknown parameters $\phi_0 \in \Phi \subset \mathbb{R}^m$. 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 $\Gamma_0$ can be consistently estimated by the least squares fitting of $Z$ on $W$. Moreover, by model assumptions we have

$$E(Y | W) = \int g(\Gamma_0 W + u; \theta_0)f_0(u; \phi_0)du \quad (2.2)$$

and

$$E(YZ | W) = \int (\Gamma_0 W + u)g(\Gamma_0 W + u; \theta_0)f_0(u; \phi_0)du. \quad (2.3)$$

Throughout the paper, all integrals are taken over the space $\mathbb{R}^k$. It follows that $\theta_0$ and $\phi_0$ can be estimated using a nonlinear least squares method, given that they are identifiable by (2.2) and (2.3).

Since it is straightforward to estimate $\Gamma_0$, in the following we focus on the estimation of $\theta_0$ and $\phi_0$. First, we use some examples to demonstrate that $\theta_0$ and $\phi_0$ may indeed be estimated using (2.2) and (2.3). To simplify notation, we consider the case where $\Gamma_0 = I$, all variables are scalars and $U \sim N(0, \phi)$. For the same reason, we suppress the subscript zero in $\theta_0$ and denote it as $\theta$.

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

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

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 $E(Y | W) = \theta_1 \exp(\theta_2 W + \theta_2 W^2)$ and $E(YZ | W) = \theta_1 \exp(\theta_2 W + \theta_2 W^2) + \theta_2 W^3$. Then, $\theta_1$ and $\theta_2$ can be consistently estimated from the first equation, and $\theta_1, \theta_2 \exp(\theta_2 W^2)$ from the second. It follows that $\theta_1$ and $\phi$ can be consistently estimated too.

Let $\psi = \langle \theta^T, \phi \rangle^T$ and $\Psi = \theta \times \phi$, which is assumed to be compact in $\mathbb{R}^{m+1}$. The true parameter value of the model is denoted by $\psi_0 \in \Psi$. To simplify notation, let $Z = (1, Z')^T$ and $x = (1, x')^T$. Then through variable substitution, (2.2) and (2.3) can be written together as

$$E(YZ | W) = \int \tilde{x} \tilde{g}(x; \theta_0) \tilde{f}_0(x - \Gamma_0 W; \phi_0)dx. \quad (2.4)$$

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

$$m(v; \psi) = \int \tilde{x} \tilde{g}(x; \theta_0) \tilde{f}_0(x - v; \phi)dx. \quad (2.5)$$

Then it is clear that $m(\Gamma_0 W; \psi_0) = E(YZ | W)$.

Suppose $(Y_j, Z_j, W_j), j = 1, 2, \ldots, n$, is an i.i.d. random sample with finite moments $EY_j^2 < \infty, E(Z_j)^2 < \infty$ and nonsingular EWW’, where $\| \cdot \|$ denotes the Euclidean norm. Further, let $\tilde{r}_j(\psi) = Y_j Z_j - m(\Gamma_0 W_j; \psi)$, where

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

is the least squares estimator of $\Gamma_0$. Then the method of moments estimator (MME) for $\psi$ is defined as $\hat{\psi}_n = \text{argmin}_{\psi} Q_n(\psi)$, where

$$Q_n(\psi) = \sum_{j=1}^n \tilde{r}_j(\psi) A_j \tilde{r}_j(\psi), \quad (2.7)$$

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

Throughout the paper, let $\gamma = \text{vec} \Gamma$ denote the vector consisting of the columns of $\Gamma$, where vec is the so-called vectorization operator. We also assume that the parameter space of $\gamma$ is a compact subset of $\mathbb{R}^{k\ell}$ containing the true value $\gamma_0 = \text{vec} \Gamma_0$. The consistency of $\hat{\psi}_n$ can be derived in traditional fashion by establishing the uniform convergence of $Q_n(\psi)/n$ to a nonstochastic function $Q(\psi)$ which has a unique minimizer $\psi_0 \in \Psi$. To achieve this, we assume the following regularity conditions, where $\mu$ 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. \( \mu \)). \( f_j(u; \phi) \) is continuously differentiable with respect to (w.r.t.) \( u \) for each \( \phi \in \Phi \) and is continuous in \( \phi \in \Phi \) (a.e. \( \mu \)). Furthermore, \( E\|A(W)\|\|Y\|^2 + \|YZ\|^2 < \infty \),

\[
E\|A(W)\|\left( \int_Y \sup_{\gamma} |g(x; \theta)f_j(x - \Gamma W; \phi)|(|x| + 1)dx \right)^2 < \infty \tag{2.8}
\]

and

\[
E\|A(W)\|\left( \int_Y \sup_{\gamma} \left| \frac{\partial f_j(x - \Gamma W; \phi)}{\partial u'} \right| \right) \left( \int_Y |g(x; \theta)|(|x| + 1)dx \right)^2 < \infty, \tag{2.9}
\]

where the supremum is taken within the compact parameter spaces of \( \psi \) and \( \gamma \).

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

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 \( Q(\psi) \) is continuous and converges uniformly in \( \psi \). 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 \( W \) 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 \( [\theta_{\min}, \theta_{\max}], [\theta_{\min}, \phi_{\max}] \) and \( [\Gamma_{\min}, \Gamma_{\max}] \). Then, for every \( x, w \in \mathbb{R} \) and \( \theta, \phi, \Gamma \) in their respective parameter spaces,

\[
g(x; \theta)f_j(x - \Gamma W; \phi)(|x| + 1) \leq \frac{\theta x}{\sqrt{2\pi \phi}} \exp \left( -\frac{(x - \Gamma W)^2}{2\phi} \right)(|x| + 1) \leq \frac{\theta_{\max} |x|(|x| + 1)}{\sqrt{2\phi_{\min}}} \exp \left( -\frac{x^2 + \Gamma_{\max}^2 w^2}{2\phi_{\max}} \right)
\]

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 and 2, \( \tilde{\psi}_n \overset{\Delta}{\rightarrow} \psi_0 \) as \( n \rightarrow \infty \).

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

Assumption 3. There exist open subsets \( \theta_0 \subset \Theta \subset \Theta \) and \( \phi_0 \subset \Phi \subset \Phi \), in which \( g(x; \theta) \) and \( f_j(u; \phi) \) are twice continuously differentiable w.r.t. (a.e. \( \mu \)) and \( f_j(u; \phi) = f_j(u_0; \phi) \) is twice continuously differentiable w.r.t. (a.e. \( \mu \)). Furthermore, \( \gamma_0 \) has an open neighborhood, such that the first two derivatives of \( g(x; \theta) f_j(x - \Gamma W; \phi) \) w.r.t. \( \gamma \) are uniformly bounded by a function \( \eta(x, u) \), which satisfies \( E\|A(W)\| \left( \int_y \sup_{\gamma} |g(x; \theta)|(|x| + 1)dx \right)^2 < \infty \).

Assumption 4. For \( \psi_0 \in \Phi \),

\[
E\|A(W)\|\left( \int_Y \sup_{\gamma} \left| \frac{\partial f_j(x - \Gamma W; \phi_0)}{\partial u'} \right| \right) \left( \int_Y |g(x; \theta_0)|(|x| + 1)dx \right)^2 < \infty, \tag{2.10}
\]

and

\[
E\|A(W)\|i\|W\| \left( \int_Y \sup_{\gamma} \left| \frac{\partial^2 f_j(x - \Gamma W; \phi_0)}{\partial \phi \partial u'} \right| \right) \left( \int_Y |g(x; \theta_0)|(|x| + 1)dx \right)^2 < \infty, \tag{2.11}
\]

where the supremum is taken within the open subset of \( \gamma \).

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

\[
\frac{\partial \rho(\psi)}{\partial \psi} = -\int_X \Phi(x; \theta) f_j(x - \Gamma W; \phi) dx
\]

and

\[
\frac{\partial \rho(\psi)}{\partial \psi'} = -\int_X \Phi(x; \theta) \frac{\partial f_j(x - \Gamma W; \phi)}{\partial \psi'} dx. \tag{2.13}
\]

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 \( Q(\psi) \) admits the first-order Taylor expansion and that the second derivative of \( Q(\psi) \) converges uniformly. Moreover, it ensures that the first derivative \( \partial Q(\psi)/\partial \psi' \) exists and is given by (2.12) and (2.13), while Assumption 4 implies that the first derivative \( \partial Q(\psi)/\partial \psi' \) exists and is given by

\[
\frac{\partial Q(\psi)}{\partial \psi'} = \int_X \Phi(x; \theta) \frac{\partial f_j(x - \Gamma W; \phi)}{\partial \psi'} dx \tag{2.14}
\]

where \( \Phi \) stands for the Kronecker product (Magnus and Neudecker, 1988, p. 30). Finally, Assumption 5 and the DCT imply that the second derivative of \( Q(\psi) \) 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 \( n \rightarrow \infty \), \( \sqrt{n}(\hat{\psi}_n - \psi_0) \overset{\Delta}{\rightarrow} N(0, \text{B}^{-1}\text{DCD}^{-1} \text{B}^{-1}) \), where

\[
\text{B} = \text{I}_{p+q} - E \left[ \frac{\partial \rho(\psi_0)}{\partial \psi} A(W) \frac{\partial \rho(\psi_0)}{\partial \psi'} \right] \left( \text{E}WW' \otimes I_k \right)^{-1};
\]

\[
\text{C} = \begin{pmatrix} \text{C}_{11} & \text{C}_{12} \\ \text{C}_{21} & \text{C}_{22} \end{pmatrix}, \tag{2.15}
\]

\[
\text{C}_{11} = E \left[ \left( \frac{\partial \rho(\psi_0)}{\partial \psi} A(W) \frac{\partial \rho(\psi_0)}{\partial \psi'} \right) \left( \frac{\partial \rho(\psi_0)}{\partial \psi} A(W) \frac{\partial \rho(\psi_0)}{\partial \psi'} \right)' \right];
\]

\[
\text{C}_{21} = E \left[ \left( \frac{\partial \rho(\psi_0)}{\partial \psi} A(W) \frac{\partial \rho(\psi_0)}{\partial \psi'} \right) \left( \frac{\partial \rho(\psi_0)}{\partial \psi} A(W) \frac{\partial \rho(\psi_0)}{\partial \psi'} \right)' \right];
\]

and \( \text{C}_{22} = E[WW' \otimes (Z - \Gamma W)](Z - \Gamma W)' \). Furthermore,

\[
B = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^{n} \frac{\partial \hat{\psi}_n}{\partial \psi_j} A_j \frac{\partial \hat{\psi}_n}{\partial \psi'} \tag{2.16}
\]

and

\[
\text{DCD}' = \lim_{n \rightarrow \infty} \frac{1}{4n} \frac{\partial Q_n(\hat{\psi}_n)}{\partial \psi} \frac{\partial Q_n(\hat{\psi}_n)}{\partial \psi'} \tag{2.17}
\]

where

\[
\frac{\partial Q_n(\psi)}{\partial \psi} = 2 \sum_{j=1}^{n} \frac{\partial \hat{\psi}_j}{\partial \psi} A_j \frac{\partial \hat{\psi}_j}{\partial \psi'}. \tag{2.18}
\]
The asymptotic covariance of \( \hat{\psi} \) 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 \( D = (D_{i,j}, G) \), so that \( DC'D = C_1 + GC_2 + C_2'G + GC_2'G \). It is easy to see that the last three terms in \( DC'D \) are due to the least squares estimation of \( \Gamma \). To simplify discussion, assume for the moment that \( D_0 \) is known, so that these three terms do not appear in \( DC'D \).

The following discussion remains valid, when \( D_0 \) is unknown and estimated using a subset of the sample \( (Y_j, Z_j, W_j) \), \( j = 1, 2, \ldots, n \), where \( Q(\psi) \) is constructed using the rest of the sample points. Then the independence of the sample points implies that \( C_{21} = 0 \). Since \( \partial^2 \rho(\psi)/\partial \psi^2 \) depends on \( W \) only, matrix \( C_{11} \) can be written as

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

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

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

(2.15)

The equivalent of the difference of the left-hand side and right-hand (reweighted) least squares estimation, we have

\[
\hat{\psi}_{n,s} = \frac{1}{n} \sum_{j=1}^{n} \hat{\beta}_{j,s}(\psi) \text{ and } \hat{\psi}_{n,s} = \frac{1}{n} \sum_{j=1}^{n} \hat{\beta}_{j,s}(\psi) \text{ for models where } H \text{ does not depend on } W.
\]

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 \( f_0 \) is normal). In such cases, 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 \( (x_i, s = 1, 2, \ldots, 2S) \) of \( (X, \psi) \) from \( h(x) \). Then approximate \( m(W_j; \psi) \) by Monte Carlo simulations

\[
m_s(\Gamma W_j; \psi) = \frac{1}{S} \sum_{s = 1}^{S} \frac{\sum_{j=1}^{n} \psi(\psi, \theta) f_0(x_j - \Gamma W_j; \phi)}{h(x_j)}
\]

and

\[
m_{2s}(\Gamma W_j; \psi) = \frac{1}{S} \sum_{s = 1}^{S} \frac{\sum_{j=1}^{n} \psi(\psi, \theta) f_0(x_j - \Gamma W_j; \phi)}{h(x_j)}
\]

where \( \hat{x}_{\psi} = (1, x_{\psi}) \). Finally, the simulation-based estimator (SBE) for \( \psi \) is defined by \( \hat{\psi}_{n,s} = \text{argmin}_{\psi \in \phi} \rho_n(\psi) \), where

\[
\rho_n(\psi) = \sum_{j=1}^{n} \hat{\beta}_{j,s}(\psi) A_{j,s}(\psi),
\]

(3.1)

4. Article in Press


5. Article in Press
where \( \rho_2 = \rho_1(\psi_0) \) and \( \rho_3 = \sum_{j=1}^{3} x\check{g}(x; \theta_0) f_U(x_0 - \Gamma_0 W; \psi_0)/h(x_0) \) is the summand in \( \rho_3(\psi_0) = \sum_{j=1}^{3} \rho_3/jh_0/S \).

Asymptotically, the importance density \( h(x) \) has no effect on the efficiency of \( \psi_{n,s} \), 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 \( m_3(\Gamma W; \psi) \) and \( m_2(\Gamma W; \psi) \). Theoretically, the best choice of \( h(x) \) is proportional to the absolute value of the integrated \( |\check{g}(x; \theta) f_U(x - \Gamma W; \psi)| \).

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 \( \theta \) and a kernel-based nonparametric estimator for \( F_U \) using moments Eqs. (2.2) and (2.3) which become

\[
E(Y | W) = \int g(\Gamma_0 W + u; \theta_0) dF_U(u) \tag{4.1}
\]

and

\[
E(YZ | W) = \int (\Gamma_0 W + u)g(\Gamma_0 W + u; \theta_0) dF_U(u). \tag{4.2}
\]

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

**Assumption 6.** The distribution of \( W \) is absolutely continuous with respect to the 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 \( \mathcal{T} = \{ t \in \mathbb{R}^k : g(t; \theta_0) \neq 0 \} \) is dense in \( \mathbb{R}^k \), where \( \hat{g}(t; \theta_0) = \int e^{-it\cdot x} g(x; \theta) dx \) is the Fourier transform of \( g(x; \theta) \).

The integrability of \( g(x; \theta) \) in Assumption 7 implies the existence of the Fourier transform \( \hat{g}(t; \theta) \). Roughly speaking, the second part of the assumption means that the zeros of \( \hat{g}(t; \theta) \) are isolated points in \( \mathbb{R}^k \). The examples given at the end of the 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

\[
m_1(v) = \int g(v + u; \theta_0) dF_U(u). \tag{4.3}
\]

Since \( \Gamma_0 \) 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 the Fourier transformation on both sides of (4.3) yields

\[
\hat{m}_1(t) = \int e^{-it\cdot v} m_1(v) dv = \int e^{-it\cdot x} g(x; \theta_0) dx \cdot \int e^{it\cdot \hat{F}_U(u)} \tag{4.4}
\]

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

\[
\hat{f}_U(t) = \hat{m}_1(t)/\hat{g}(t; \theta_0). \tag{4.5}
\]

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

\[
f_U(u; \theta_0) = \frac{1}{(2\pi)^l} \int e^{it\cdot u} \hat{m}_1(t)/\hat{g}(t; \theta_0). \tag{4.6}
\]

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

First, let \( \tilde{f} \) denote the least squares estimator in (2.6) and \( \check{f}_U = \tilde{f}_U W_j, j = 1, 2, \ldots, n \). Then the density function \( f_U(v) \) of \( V = \Gamma_0 W \) and the conditional mean function \( m_1(v) \) are estimated by

\[
\hat{f}_U(v) = \frac{1}{nm_1} \sum_{j=1}^{n} K \left( \frac{v - \check{f}_U(v)}{a_n} \right) \tag{4.7}
\]

and

\[
\hat{m}_1(v) = \frac{1}{nm_1} \sum_{j=1}^{n} K \left( \frac{v - \check{f}_U(v)}{a_n} \right) \check{f}_U(v), \tag{4.8}
\]

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

Second, let \( \hat{m}_1(t) = \int e^{-it\cdot \check{f}_U(v)} \check{f}_U(v) dv, \) where \( \check{m}_1(t) = \int e^{-it\cdot v} \check{m}_1(v) dv \), \( \check{f}_U(v) \geq b_n, 0 < b_n \to 0 \) as \( n \to \infty \), and define, for each \( \theta \in \Theta \),

\[
\hat{f}_U(t; \theta) = \frac{1}{(2\pi)^l} \int e^{it\cdot \check{f}_U(v)} \hat{m}_1(t)/\hat{g}(t; \theta) dt, \tag{4.9}
\]

where \( C_n(\theta) = \inf_{\theta \in \Theta} \theta \) and \( \theta \in C_n, 0 < c_n \to 0 \) as \( n \to \infty \).

Finally, the semiparametric estimator (SPE) for \( \theta \) is defined as

\[
\hat{\theta}(\tilde{V}) = \arg \min_{\theta \in \Theta} \check{m}_1(t)/\check{g}(t; \theta). \tag{4.10}
\]

with \( \hat{\theta}(\tilde{V}) = \hat{\theta}(\tilde{V}) = \int \check{g}(x; \theta) \check{f}_U(x - \tilde{V} W_j; \theta) dx \). The consistency of \( \hat{\theta} \) can be derived similarly as for the MME \( \hat{\theta}_n \). 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 \( Q_0(\theta) \), which have convergence rates slower 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 \( \check{f}_U(v) \), \( m_1(v) \check{f}_U(v) \) 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) \( \int K(v) dv = 1; \) \( \int v_i^d v_j^d K(v) dv = 0, \) for \( d_i \geq 0 \) and \( 1 \leq d = d_j \); and (2) \( \int v_i^d v_j^d K(v) dv < \infty, \) for \( d_i \geq 0 \) and \( 1 \leq d = d_j \); \( k \) for every \( 1 \leq j \leq k, \sup_{\|v\| \leq 1} \|K(v)\| \|v\| < \infty \), and (3) \( K(v) \in L^1(\mathbb{R}^k) \) and \( \int e^{it\cdot K(v)} dv \in L^1(\mathbb{R}^k) \).

**Assumption 10.** \( \sup_{\|v\| \leq 1} \|K(v)\| \|v\| + 1 \in L^1(\mathbb{R}^k) \) and \( \sup_{\|v\| \leq 1} \|K(v)\| \|v\| + 1 \in L^1(\mathbb{R}^k) \).

**Assumption 11.** \( g(x; \theta) \) is a measurable function of \( x \) for each \( \theta \in \Theta \) and is continuous in \( \theta \in \Theta \). Furthermore, \( \|A(W)\| \|W\| + 2 < \infty, \) \( \|A(W)\| \|W\| + 2 < \infty \) and \( \|A(W)\| \|W\| + 2 < \infty \) for all \( \mu \neq 0 \).
**Assumption 12.** \( E[\rho(\theta - \rho(\theta_0)) A(W) (\rho(\theta) - \rho(\theta_0)) ] = 0 \) if and only if \( \theta = \theta_0 \), where \( \rho(\theta) = YZ - \int f_0(x; \theta_0) f_0(x - \Gamma_0 W; \theta) \) 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 \( \hat{g}(\theta) \) exists for all \( \theta \in \Theta \) and that the density \( f_0(u; \theta) \) exists and is given by (4.6). This assumption may be weakened in the case where \( \rho(\theta) \) 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 Assumptions 2 and 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 \( n \to \infty \), \( \sqrt{n} \rho^{-1} b_{n, k} \to \infty \), \( \sqrt{n} \rho^{-1} c_{n, k} \to \infty \), and \( n^{-k} \int f_0 m_1(v) \) converge to \( 0 \), where \( B_k \) is the complement of \( B_k \) in \( \mathbb{R}^k \). Then, as \( n \to \infty \), (1) \( \hat{\theta}_n \to \theta_0 \); (2) \( \sup_{u \in \mathbb{R}^k} |f_0(u; \hat{\theta}_n) - f_0(u; \theta_0)| \to 0 \); and (3) for every \( r > 0 \), \( \sup_{u \in \mathbb{R}^k} \left| \hat{g}(u; \hat{\theta}_n) - \hat{g}(r; \theta_0) \right| \to 0 \).

The above semiparametric estimator involves three tuning parameters. In practice, these parameters can be chosen as follows. First, take the bandwidth \( a_n = n^{-1/2} \) so that \( n^{-1/2} \int f_0 m_1(v) \) is the tail behavior of \( m_1(v) \) as \( v \to \infty \). Then, choose \( b > 0 \) so that \( b \theta \) is used in the model \( g(x, \theta) \) and density \( f_0(v) \). Suppose \( a_n = n^{-1/2} \) for some \( \delta > 0 \). Then \( c_n = n^{-\delta} \theta \) where \( c_n > 0 \) satisfies \( c_n^{-1} \to 0 \). Finally, choose \( d \) so that \( d \theta \) is used in the model \( f_0(v) \).

Similar to the simulation-based estimator of Section 3, we can also construct a semiparametric version of the semiparametric estimator. Specifically, \( \hat{\theta}_n \) in (4.9) can be replaced by Monte Carlo simulations such as \( \hat{\theta}_n \) and \( \hat{\theta}_n \) in (3.1). Then a simulation-based semiparametric estimator (SBSE) \( \hat{\theta}_n \) can be defined by minimizing the semiparametric model of \( Q_n(\theta) \), i.e.,

\[
Q_n(\theta) = \sum_{j=1}^{n} \hat{\rho}_j(\theta_0) \hat{A}_j \hat{\rho}_j(\theta_0),
\]

where \( \hat{\rho}_j(\theta_0) = Y_{j} \hat{\rho}_j - \frac{1}{n} \sum_{j=1}^{n} \hat{\rho}_j g(x_j; \theta_0) f_0(x_j - \Gamma_0 W_j; \theta)/h(x_j) \), and \( \hat{\theta}_n \) is defined similarly using \( \{x_j, y_j, s = 1, 2, \ldots, 25\} \). Moreover, from Section 3 it is easy to see that \( \hat{\theta}_n \) has the same properties given in Theorem 4.1 for the SPE \( \theta_0 \). The asymptotic normality of \( \hat{\theta}_n \) can also be established in a similar way, under the following further assumptions.

**Assumption 13.** \( \theta \) contains an open neighborhood \( \theta_0 \) of \( \theta_0 \) such that (1) \( g(x; \theta) \) is twice continuously differentiable w.r.t. \( \theta \) in \( \theta_0 \); (2) the first two derivatives of \( m_1(v; \theta) \) exist w.r.t. \( \theta \) and its first derivative w.r.t. \( \gamma \) are uniformly bounded in \( \theta_0 \) and an open and bounded neighborhood of \( \gamma_0 \) by \( \eta(\gamma) \) satisfies \( \int_{\gamma_0} \eta(\gamma) d\gamma < \infty \); (3) \( g(x; \theta_0) f_0(x - \Gamma_0 W; \theta_0) \) has the same properties given in Assumptions 3 and 4 for \( g(x; \theta) f_0(x - \Gamma_0 W; \phi) \). Therefore, the asymptotic covariance matrix of \( \hat{\theta}_n \) consists of the approximation errors of \( \hat{\Gamma}_n \), \( \hat{\xi}_n \), \( \hat{\xi}_n \), \( f_0 \), \( \xi_n \), \( \xi_n \), \( \sigma_n \), and \( \xi_n \), the sampling error \( \xi_n \). It is easy to see that, if \( \Delta_0 = 0 \) and if \( f_0 = 0 \), then the asymptotic covariance matrix of our simulation estimator only depends on the sampling error \( \xi_n \) and sampling error \( \xi_n \). Since \( \xi_n \), the impact of the sampling error can be reduced by increasing the simulation size \( S \).
5. Identifiability

Identifiability is a long-standing and difficult problem in non-linear 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 are 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; Carroll et al., 1995). Hausman et al. (1991) showed that the polynomial model is identifiable using instrumental variables. Also, using the IV approach Wang and Hsiao (1995) obtained identifiability for models with integrable \( g(\theta_0) \). 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).

First, \( I_{\theta_0} \) is clearly identifiable by (2.1) and the least squares method. In the previous section, we have demonstrated that \( T_{\theta_0} \) is uniquely determined by \( \theta_0 \) and \( I_{\theta_0} \) through (4.5). In the following, we study the identifiability of \( \theta_0 \) using (4.1) and (4.2), given that \( I_{\theta_0} \) is identified. Analogous to (4.3), for every \( v \in \mathbb{R}^k \), let

\[
m_2(v) = \int (v + u)g(u; \theta_0)dF_0(u).
\]  

(5.1)

Then \( m_2(I_{\theta_0}W) = E(Y | W) \) 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 the Fubini Theorem, we obtain

\[
\int m_1(v)dv = \int g(x; \theta_0)dx : = g_1(\theta_0),
\]

(5.2)

\[
\int m_2(v)dv = \int xyg(x; \theta_0)dx : = g_2(\theta_0).
\]

(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 a unique solution \( \theta_0 \) in its neighborhood is that the Jacobian matrix

\[
J(\theta_0) = \left( \frac{\partial g_1(\theta_0)}{\partial \theta} - \frac{\partial g_2(\theta_0)}{\partial \theta} \right)
\]

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

**Theorem 5.1.** Under Assumptions 6 and 7, a sufficient condition for \( \theta_0 \) and \( I_{\theta_0} \) to be identifiable is rank \( J(\theta_0) = p \).

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

**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 \( \hat{g}_j(t; \theta_0) = \int e^{-itx}g(x; \theta_0)dx \) is dense in \( \mathbb{R}^k \), where \( x \) is the \( j \)-th coordinate of \( x = (x_1, x_2, \ldots, x_k) \), then \( \hat{g}_j(t) \) can still be identified by using the \( j \)-th equation in (5.1). This is easy to see by taking a Fourier transformation on both sides of (5.1), which yields \( \hat{m}_j(t) = \hat{g}_j(t) \int e^{-itx}g(x; \theta_0)dx \). Moreover, if \( \hat{g}_j(t) \) is analytic, then the second condition in Assumption 7 can be further weakened to the assumption that \( \hat{g}_j(t; \theta_0) \neq 0 \) from some \( t \in \mathbb{R}^k \).

This follows from the facts that the continuity of \( \hat{g}_j(t; \theta_0) \) implies that \( \hat{g}_j(t; \theta_0) \neq 0 \) 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, p. 197–198). 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.** \( E[g(x; \theta_0) ||X||] < \infty \).

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

\[
m_{1,n}(v) = \int g_n(v + u; \theta_0)dF_0(u),
\]

(5.4)

\[
m_{2,n}(v) = \int (v + u)g_n(v + u; \theta_0)dF_0(u).
\]

Then \( m_{1,n}(I_{\theta_0}W) \) and \( m_{2,n}(I_{\theta_0}W) \) approximate \( E(Y | W) \) and \( E(Y | W) \) respectively in the following sense.

**Theorem 5.2.** Under Assumption 17, it holds that

\[
\lim_{n \to \infty} E[m_{1,n}(I_{\theta_0}W)] - E(Y | W) = 0
\]

(5.5)

and

\[
\lim_{n \to \infty} E[m_{2,n}(I_{\theta_0}W)] - E(Y | W) = 0.
\]

(5.6)

Since \( E(Y | W) \) and \( E(Y | W) \) can be consistently estimated by nonparametric methods, together with Assumptions 6, Theorem 5.2 implies that \( m_{1,n}(I_{\theta_0}W) \) and \( m_{2,n}(I_{\theta_0}W) \) 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 Jacobi 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 \( I_{\theta_0} = 1 \). In this case, we need only to verify Assumptions 7 or 17.

**Example 5.1.** Exponential model \( g(x; \theta) = e^{-\theta x} \), \( \theta > 0 \). Clearly \( g(x; \theta) \) is integrable. Further, the second part of Assumption 7 is satisfied because \( \hat{g}(t; \theta) = \sqrt{\pi/\theta}e^{-\pi t^2/4\theta} \). To check the rank condition, we calculate \( g_1(\theta) = \int e^{-\theta x^2}dx = \sqrt{\pi/\theta} \) and \( g_2(\theta) = \int xe^{-\theta x^2}dx = 0 \). It follows that \( J(\theta) = (-\sqrt{\pi}/(2\theta \sqrt{\theta}) \) 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 \( T > 0 \),

\[
\hat{g}(t; \theta) = \int_T^{-T} e^{-itx}x/dx = 2\theta \left( T \cos(\theta T) - \sin(\theta T) \right) \left( T \cos(\theta T) - \sin(\theta T) \right)
\]

(5.7)

the second condition in Assumption 7 is satisfied. To check the rank condition, we calculate \( g_1(\theta) = \theta \int_T^{-T} xdx = 0 \) and \( g_2(\theta) = \theta \int_T^{-T} x^2dx = 2\theta T^2/3 \). Therefore \( J(\theta) = (0, 2\theta T^2/3 \) 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 \( E||X||^2 < \infty \). Further, because for any \( T > 0 \),

\[
\hat{g}(t; \theta) = 2\theta T \cos(t \theta) \left( \frac{\theta_1 T^2}{t^2} \right) + 2 \sin(t \theta) \left( \frac{\theta_2 T^2}{t^3} \right)
\]

(5.8)
the second condition in Assumption 7 is clearly satisfied. Again, it is straightforward to calculate $g_1(\theta) = 2\theta T^3/3$ and $g_2(\theta) = 2\theta T^3/3$. Hence

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

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 7 becomes $\int e^{\theta x} f(x) < \infty$ and $\mathbb{E} |x|$.

The second condition in Assumption 7 is satisfied. Moreover, since for any $\theta$, $g_1(\theta) = (e^{\theta x} - e^{-\theta x})/\theta$, and

$$g_2(\theta) = (T e^{\theta x} - e^{-\theta x}) - e^{\theta x} - e^{-\theta x}.$$

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

**Example 5.5.** Consider $g(x; \theta) = \theta_1 + \theta_2 x^p$, 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, $\theta_1$ and $\theta_2$ would be identifiable if $x$ were observable. However, $\theta_1$ and $\theta_2$ 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 Eq. (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 $f_U(u; \phi_0)$. 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), $\hat{p}_n(\theta)$ and hence $Q_n(\psi)$ are continuously differentiable with respect to $\psi = \psi(\theta)$. For sufficiently large $n$, therefore $Q_n(\psi)$ has the first-order Taylor expansion about $\theta_0 = \psi(\theta)$:

$$Q_n(\psi) = \sum_{j=1}^n \rho_j^2(\psi) A_j \psi_j(\theta) + 2 \sum_{j=1}^n \rho_j^2(\psi, \theta) \frac{\partial \rho_j(\psi, \theta)}{\partial \psi_j} (\hat{\theta} - \theta_0) + \frac{1}{2} \sum_{j=1}^n \rho_j^2(\psi, \theta) \frac{\partial^2 \rho_j(\psi, \theta)}{\partial \psi_j^2} (\hat{\theta} - \theta_0)^2 + \cdots, \quad (7.1)$$

where $\rho_j(\psi) = \mathbb{E} \hat{Z}_j - \mathbb{E} \hat{Z}_j - m \hat{W}_j | \psi_j$, $\rho_j(\psi, \theta) = \mathbb{E} \hat{Z}_j - \mathbb{E} \hat{Z}_j - m \hat{W}_j | \psi_j$.

Further, since $g(x; \theta)$ and $f_{U|W}(u; \phi)$ are continuous in $\theta$ and $\phi$ respectively, by (2.8) and the DCT, $\rho_j(\psi)$ is continuous in $\psi$ and, moreover,

$$E \sup_{\psi} \left| \rho_j(\psi) \right| A_j \psi_j(\theta) \leq E \sup_{\psi} \left| \rho_j(\psi) \right| \leq 2 E \left\| A \right\| \left( \mathbb{E} \sup_{\psi} \left| \rho_j(\psi) \right| \right)^2 \
+ 2 E \left\| A \right\| \left( \mathbb{E} \sup_{\psi} \left| g(x; \theta) f_U(x - \Gamma(W) + U) \right| \right) \left( \mathbb{E} \left| x \right| + 1 \right) dx \leq \infty,$$

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

$$\sup_{\psi} \left| \rho_j(\psi) A_j \psi_j(\theta) - Q(\psi) \right| \xrightarrow{a.s.} 0, \quad (7.2)$$

where $Q(\psi) = E \rho_j(\psi) A_j \psi_j(\theta)$. Similarly, since by the Cauchy–Schwarz inequality and (2.9),

$$E \left\| A \right\| \left( \mathbb{E} \sup_{\psi} \left| \rho_j(\psi) \right| \right)^2 \leq E \left\| A \right\| \left( \mathbb{E} \sup_{\psi} \left| \rho_j(\psi) \right| \right)^2 \frac{\partial \rho_j(\psi, \theta)}{\partial \psi_j} \left( \mathbb{E} \sup_{\psi} \left| \rho_j(\psi) \right| \right)^2 \leq k E \left\| A \right\| \left( \mathbb{E} \sup_{\psi} \left| \rho_j(\psi) \right| \right)^2 E \left\| A \right\| \left| \mathbb{E} \left| x \right| + 1 \right| dx \leq \infty,$$

and again by the ULLN we have

$$\sup_{\psi, \theta} \left| \frac{1}{n} \sum_{j=1}^n \rho_j(\psi, \theta) A_j \psi_j(\theta) \right| = O(1) \quad \text{a.s.}$$
and therefore
\[
\sup_{\psi} \left| \frac{1}{n} \sum_{j=1}^{n} \rho_j(\psi, \tilde{y}) A_j \frac{\partial \rho_j(\psi, \tilde{y})}{\partial \gamma'} (\tilde{y} - y_0) \right| \leq \sup_{\gamma, \psi} \left| \frac{1}{n} \sum_{j=1}^{n} \rho_j(\psi, \gamma) A_j \frac{\partial \rho_j(\psi, \gamma)}{\partial \gamma'} \right| \| \tilde{y} - y_0 \| \overset{a.s.}{\to} 0. \tag{7.3}
\]

It follows from (7.1)–(7.3) that
\[
\sup_{\psi} \left| \frac{1}{n} Q_n(\psi) - Q(\psi) \right| \overset{a.s.}{\to} 0. \tag{7.4}
\]

Further, because \( E(\rho_1(\psi_0)|W_1) = 0 \) and \( \rho_1(\psi) - \rho_1(\psi_0) \) depends on \( W_1 \) only, we have
\[
E[\rho_j(\psi_0) A_j (\rho_1(\psi) - \rho_1(\psi_0))] = E[E(\rho_j(\psi_0)|W_1) A_j (\rho_1(\psi) - \rho_1(\psi_0))] = 0,
\]

which implies \( Q(\psi) = Q(\psi_0) + E((\rho_1(\psi) - \rho_1(\psi_0)) A_j (\rho_1(\psi) - \rho_1(\psi_0))) \). By Assumption 2, \( Q(\psi) \geq Q(\psi_0) \) and the equality holds if and only if \( \psi = \psi_0 \). If \( Q(\psi) \) attains a unique minimum at \( \psi_0 \in \Psi \) and, therefore, by Amemiya (1973, Lemma 3) that \( \tilde{\psi}_n \overset{a.s.}{\to} \psi_0 \).

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 \( \psi_0 \). Since \( \partial Q_n(\tilde{\psi}_n)/\partial \psi = 0 \) and \( \tilde{\psi}_n \overset{a.s.}{\to} \psi_0 \), for sufficiently large \( n \) we have
\[
\frac{\partial Q_n(\tilde{\psi}_n)}{\partial \psi} + \frac{\partial^2 Q_n(\tilde{\psi}_n)}{\partial \psi \partial \gamma'} (\tilde{\psi}_n - \psi_0) = 0, \tag{7.5}
\]

where \( \| \tilde{\psi}_n - \psi_0 \| \leq \| \tilde{\psi}_n - \psi_0 \| \). The first derivative of \( Q_n(\psi) \) in (7.5) is given by
\[
\frac{\partial Q_n(\psi)}{\partial \psi} = 2 \sum_{j=1}^{n} \frac{\partial \tilde{\rho}_j(\psi)}{\partial \psi} A_j \tilde{\rho}_j(\psi), \tag{7.6}
\]

where \( \partial \tilde{\rho}_j(\psi)/\partial \psi \) consists of
\[
\frac{\partial \tilde{\rho}_j(\psi)}{\partial \theta} = - \int \frac{\partial g(x; \theta)}{\partial \theta} f_i(x - \tilde{W}_j; \phi) dx \tag{7.7}
\]

and
\[
\frac{\partial \tilde{\rho}_j(\psi)}{\partial \phi} = - \int \frac{\partial f_i(x - \tilde{W}_j; \phi)}{\partial \phi} g(x; \theta) dx. \tag{7.8}
\]

The second derivative in (7.5) is given by
\[
\frac{\partial^2 Q_n(\tilde{\psi}_n)}{\partial \psi \partial \gamma'} = 2 \sum_{j=1}^{n} \left[ \frac{\partial \tilde{\rho}_j(\psi)}{\partial \psi} A_j \frac{\partial \tilde{\rho}_j(\psi)}{\partial \psi} + (\tilde{\rho}_j(\psi) A_j \otimes I_{p+q}) \frac{\partial \vec{\rho}(\psi)}{\partial \psi} \right], \tag{7.9}
\]

where \( \partial \vec{\rho}(\psi)/\partial \psi \) consists of
\[
\frac{\partial \vec{\rho}(\psi)}{\partial \psi} (\tilde{\psi}_n - \psi_0) \overset{a.s.}{\to} 0.
\]

It follows from the ULN and Amemiya (1973, Lemma 4) that
\[
\frac{1}{2n} \frac{\partial^2 Q_n(\tilde{\psi}_n)}{\partial \psi \partial \gamma'} \overset{a.s.}{\to} E \left[ \frac{\partial \vec{\rho}(\psi_0)}{\partial \psi} A_1 \frac{\partial \vec{\rho}(\psi_0)}{\partial \psi} \right] + \left( \rho_1(\psi_0) A_1 \otimes I_{p+q} \right) \frac{\partial \vec{\rho}(\psi_0)}{\partial \psi} \overset{a.s.}{\to} 0, \tag{7.10}
\]

where
\[
\frac{\partial \vec{\rho}(\psi)}{\partial \psi} = \frac{\partial^2 f_i(x - \tilde{W}_j; \phi)}{\partial \phi} g(x; \theta) dx, \tag{7.11}
\]

It follows from Assumption 3 that
\[
E \left[ \left( \frac{\partial \vec{\rho}(\psi, \gamma)}{\partial \psi} A_1 \frac{\partial \vec{\rho}(\psi, \gamma)}{\partial \psi} \right) + \left( \rho_1(\psi) A_1 \otimes I_{p+q} \right) \frac{\partial \vec{\rho}(\psi)}{\partial \psi} \right] \overset{a.s.}{\to} 0.
\]

Similarly, by Assumption 3 we have
\[
\left( \frac{\partial \vec{\rho}(\psi, \gamma)}{\partial \psi} + \left( \rho_1(\psi_0) A_1 \otimes I_{p+q} \right) \frac{\partial \vec{\rho}(\psi_0)}{\partial \psi} \right) \overset{a.s.}{\to} 0.
\]

Further, by Assumption 4 and the DCT, \( \partial Q_n(\psi)/\partial \psi \) is continuously differentiable with respect to \( \gamma \) and hence, for sufficiently large \( n \), has the first-order Taylor expansion about \( \gamma_0 \):
\[
\frac{\partial Q_n(\psi)}{\partial \psi} = 2 \sum_{j=1}^{n} \frac{\partial \tilde{\rho}_j(\psi)}{\partial \psi} A_j \frac{\partial \tilde{\rho}_j(\psi)}{\partial \psi} + \frac{\partial^2 \vec{\rho}(\psi)}{\partial \psi} (\tilde{\gamma} - \gamma_0), \tag{7.12}
\]

where
\[
\frac{\partial \vec{\rho}(\psi)}{\partial \psi} = \frac{\partial^2 f_i(x - \tilde{W}_j; \phi)}{\partial \phi} g(x; \theta) dx.
\]
\[\frac{\partial^2 \hat{Q}_n(\psi_0)}{\partial \psi \partial \psi'} = 2 \sum_{j=1}^{n} \left[ \frac{\partial \rho_j'(\psi_0, \hat{\gamma})}{\partial \psi} A_j \frac{\partial \rho_j'(\psi_0, \hat{\gamma})}{\partial \psi'} \right] + \left( \rho_j'(\psi_0, \hat{\gamma}) A_j \right)^{lp+q}, \]
\[
\frac{\partial \hat{Q}_n(\psi_0)}{\partial \psi} = \frac{1}{2} \frac{\partial^2 \hat{Q}_n(\psi_0)}{\partial \psi \partial \psi'} \rightarrow N(0, DCD'). \tag{7.12} \]

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

### 7.3. Proof of Theorem 3.1

We only sketch the proofs here because they are similar to the proofs of Theorems 2.1 and 2.2. Details are available in Wang and Hsiao (2008). The proof of Theorem 3.1 is based on the first-order Taylor expansion of \(Q_{\psi,\gamma}(\psi)\) about \(\gamma_0\):

\[Q_{\psi,\gamma}(\psi) = \sum_{j=1}^{n} \rho_j'(\psi_\gamma) A_j \rho_j(\psi_\gamma) \]

and \(\rho_j(\psi_\gamma) \) are given similarly. By Assumption 4, the ULLN and the conditional independence of \(\rho_1, \rho_2, \ldots, \rho_n\) we can show that

\[\sup_{\psi} \frac{1}{n} \left| Q_{\psi,\gamma}(\psi) - E \rho_1(\psi_\gamma) A_1 \rho_2(\psi_\gamma) \right| \rightarrow 0, \tag{7.14} \]

which has been shown in the proof of Theorem 2.1 to attain a unique minimum at \(\psi_0 \in \Psi\). Therefore, \(\hat{\psi}_n \rightarrow \psi_0\) follows from Amemiya (1973, Lemma 3).

The proof of Theorem 3.2 is based on the first-order Taylor expansion of \(\partial Q_{\psi,\gamma}(\psi)/\partial \psi\) in an open neighborhood \(\Psi_0 \subset \Psi\) of \(\psi_0\):

\[\frac{\partial Q_{\psi,\gamma}(\psi_0)}{\partial \psi} + \frac{\partial^2 Q_{\psi,\gamma}(\psi_\gamma)}{\partial \psi \partial \psi'} (\hat{\psi}_n - \psi_0) = 0. \tag{7.15} \]

Moreover, since \(T_j, j = 1, 2, \ldots, n\) are i.i.d., by the Central Limit Theorem, \(\sqrt{n} \sum_{j=1}^{n} T_j \rightarrow N(0, C)^{'}\), Therefore, according to Slutsky’s Theorem, we have

\[\frac{\partial Q_{\psi,\gamma}(\psi_0)}{\partial \psi} \rightarrow N(0, DCD'). \tag{7.12} \]
Andrews (1995, Theorem 2), under Assumptions 8 and 9, the kernel estimators in (4.7) and (4.8) satisfy

\[
\sup_{v \in \mathbb{R}^k} |\hat{f}_v(v) - f_v(v)| = O_p(n^{-1/2}a_n^{-k-1} + O_p(a_n^d)) \tag{7.20}
\]

and

\[
\sup_{u \in \mathbb{R}} |\hat{m}_1(u) - m_1(u)| = O_p(n^{-1/2}a_n^{-k-1}b_n^{-2} + O_p(a_n^d b_n^2)) \tag{7.21}
\]

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

\[
(2\pi)^{k/2} |\hat{f}_u(u; \theta) - f_u(u; \theta)| \leq \int e^{it'\hat{m}_1(t; \theta)/\hat{g}(t; \theta)} dt + \int e^{it'\hat{m}_1(t; \theta)/\hat{g}(t; \theta)} dt \tag{7.22}
\]

where

\[
sup_{v \in \mathbb{R}^k} |\hat{f}_v(v) - f_v(v)| \leq \frac{2}{b_n} \int |\hat{m}_1(v) - m_1(v)| d\nu + \frac{1}{b_n} \sup_{v \in \mathbb{R}^k} |\hat{m}_1(v)| d\nu
\]

and

\[
\sup_{u \in \mathbb{R}} |\hat{m}_1(u) - m_1(u)| = O_p(e_n^{k+1}),
\]

where the last equality follows by (7.21) and condition of Theorem 4.1. It follows from (7.22) and Assumption 10, that

\[
\sup_{\theta \in \Theta} |\hat{f}_u(u; \theta) - f_u(u; \theta)| = o_p(1) \tag{7.23}
\]

Now, we write

\[
Q_n(\theta) = \sum_{i=1}^n \rho_i(\theta)\hat{A}_i(\hat{\rho}_i(\theta) - \rho_i(\theta)) + \sum_{j=1}^n \rho_j(\theta)\hat{A}_j(\hat{\rho}_j(\theta) - \rho_j(\theta)) \tag{7.24}
\]

where

\[
\rho_i(\theta) = Y_i\hat{Z}_i - \int \tilde{x}g(x; \theta) f_0(x - \Gamma_0 \cdot W_i; \theta) dx.
\]

Then, since

\[
|\tilde{f}_u(x - \Gamma_0 \cdot W_i; \theta) - f_u(x - \Gamma_0 \cdot W_i; \theta)| dx
\]

we have

\[
\sup_{\theta \in \Theta} |\hat{\rho}_i(\theta) - \rho_i(\theta)|
\]

Finally, Theorem 3.1.2 follows from (7.15), (7.16) and (7.19).
where $\alpha_1$ and $\alpha_2$ are positive constants. Further, by the Cauchy–Schwarz inequality,
\[ E \sup_{\theta} \| A_1 \| (\| W_1 \| + 1) \| \rho_1(\theta) \|^2 \leq E \sup_{\theta} \| A_1 \| (\| W_1 \| + 1)^2 E \| A_1 \| \| \rho_1(\theta) \|^2 \]
\[ \leq 2E \sup_{\theta} \| A_1 \| (\| W_1 \|^2 + 1) \left[ E \| A_1 \| (\| Y_1 \|^2 + \| Y_2 Z_1 \|^2) \right] \]
\[ + E \sup_{\theta} \left( \int \sup_{x} \| \tilde{g}(x; \theta) f_0(x - \Gamma_0 W_1 ; \theta) \| dx \right)^2 \]
\[ < \infty, \]
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
\[ \sup_{\theta} \left| \frac{1}{n} \sum_{j=1}^{n} \rho_j(\theta) A_j (\hat{\rho}_j(\theta) - \rho_j(\theta)) \right| \]
\[ \leq \frac{1}{n} \sum_{j=1}^{n} \| A_j \| \sup_{\theta} \| \rho_j(\theta) \| \| \hat{\rho}_j(\theta) - \rho_j(\theta) \| \]
\[ \leq \frac{\alpha_1}{n} \left\| \hat{\Sigma} - \Gamma_0 \right\| + \frac{\alpha_2}{n} \sup_{\theta \in \Theta} \left\| \tilde{f}_0(u; \theta) - f_0(u; \theta) \right\| \]
\[ \times \frac{1}{n} \sum_{j=1}^{n} \| A_j \| (\| W_j \| + 1) \sup_{\theta} \| \rho_j(\theta) \| \]
\[ = o_p(1), \quad (7.25) \]
and analogously,
\[ \sup_{\theta} \left| \frac{1}{n} \sum_{j=1}^{n} (\hat{\rho}_j(\theta) - \rho_j(\theta))^2 A_j (\hat{\rho}_j(\theta) - \rho_j(\theta)) \right| \]
\[ \leq \frac{1}{n} \sum_{j=1}^{n} \| A_j \| \sup_{\theta} \| (\hat{\rho}_j(\theta) - \rho_j(\theta))^2 \| \]
\[ \leq \left( 2\alpha_1^2 \right) \left\| \hat{\Sigma} - \Gamma_0 \right\|^2 + 2\alpha_2^2 \sup_{\theta \in \Theta} \left\| \tilde{f}_0(u; \theta) - f_0(u; \theta) \right\|^2 \]
\[ \times \frac{1}{n} \sum_{j=1}^{n} \| A_j \| (\| W_j \|^2 + 1) \]
\[ = o_p(1). \quad (7.26) \]
Since
\[ E \sup_{\theta} |\rho_1(\theta)| A_1 \rho_1(\theta) | \]
\[ \leq E \sup_{\theta} \| A_1 \| \sup_{\theta} \| \rho_1(\theta) \|^2 \]
\[ \leq 2E \| A_1 \| (\| Y_1 \|^2 + \| Y_2 Z_1 \|^2) \]
\[ + E \| A_1 \| \left( \int \sup_{x} \| \tilde{g}(x; \theta) f_0(x - \Gamma_0 W_1 ; \theta) \| dx \right)^2 \]
\[ \leq \infty, \]
by the ULLN we have
\[ \sup_{\theta} \left| \frac{1}{n} \sum_{j=1}^{n} \rho_j(\theta) A_j \rho_j(\theta) - Q(\theta) \right| = o_p(1), \quad (7.27) \]
where $Q(\theta) = E \rho_1(\theta) A_1 \rho_1(\theta)$. It follows from (7.24)–(7.27) that
\[ \tilde{Q}_n(\theta) \xrightarrow{p} Q(\theta) \text{ uniformly in } \theta \in \Theta. \]
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 $\tilde{Q}_n \xrightarrow{p} \tilde{Q}_0$ follows from Amemiya (1973, Lemma 3).

Furthermore, Theorem 4.1.2 follows from (7.23) and Amemiya (1973, Lemma 4).

To prove Theorem 4.1.3, note that for any $t \in \mathbb{R}^k$, since $\hat{g}(t; \theta_0) \neq 0$ and $\hat{g}(t; \theta)$ is continuous in $\theta$, there exists an open neighborhood $\Theta_0$ of $\theta_0$ in $\Theta$, such that $\inf_{\Theta_0} \| \hat{g}(t; \theta) \| \geq c > 0$. Therefore, for any $\theta \in \Theta_0$,
\[ \| \hat{m}_1(t) - \tilde{m}_1(t) \| \leq \frac{1}{c} \int_{\Theta_0} \| \hat{m}_1(u) - \tilde{m}_1(u) \| du + \frac{1}{c} \int_{\Theta_0} \| \tilde{m}_1(u) \| du, \]
which implies that $\sup_{\Theta_0} \| \hat{m}_1(t; \theta) - \tilde{m}_1(t; \theta) \| = o_p(1)$. The result then follows from Amemiya (1973, Lemma 4).

7.5. Proof of Theorem 4.2

By Assumption 13 the first derivative $\partial Q_{n,5}(\theta)/\partial \theta$ exists and has the first order Taylor expansion in a neighborhood of $\theta_0$. Since $\partial Q_{n,5}(\theta_0)/\partial \theta = 0$ and $\tilde{Q}_{n,5} \xrightarrow{p} \tilde{Q}_0$, for sufficiently large $n$, we have
\[ 0 = \frac{\partial Q_{n,5}(\theta_0)}{\partial \theta} + \frac{\partial^2 Q_{n,5}(\theta_0)}{\partial \theta^2} \tilde{\theta}, \quad (7.28) \]
where $\tilde{\theta} \in \Theta_0$. The first derivative in (7.28) is given by
\[ \frac{\partial Q_{n,5}(\theta_0)}{\partial \theta} = \sum_{j=1}^{n} \left[ \frac{\partial \tilde{p}_{j,5}(\theta_0)}{\partial \theta} A_j \tilde{p}_{j,5}(\theta_0) + \frac{\partial \tilde{p}_{j,25}(\theta_0)}{\partial \theta} A_j \tilde{p}_{j,25}(\theta_0) \right], \]
where
\[ \frac{\partial \tilde{p}_{j,5}(\theta_0)}{\partial \theta} = -\frac{1}{S} \sum_{j=1}^{S} \frac{\tilde{X}_j}{h(x_j)} \left( \frac{d g(x_j; \theta)}{\partial \theta} \tilde{f}_0(x_j - \hat{\Sigma} W_j; \theta) \right) \]
\[ + g(x_j; \theta) \frac{\partial \tilde{f}_0(x_j - \hat{\Sigma} W_j; \theta)}{\partial \theta}, \]
and
\[ \frac{\partial^2 Q_{n,5}(\theta_0)}{\partial \theta^2} = \frac{1}{(2\pi)^S} \int_{\Theta_0} e^{i \theta^T (\hat{\Sigma} W_j - \tilde{m}_1(t))} \tilde{g}(\theta) \tilde{g}(\theta) \frac{\partial \tilde{m}_1(t)}{\partial \theta} \frac{\partial \tilde{g}(\theta)}{\partial \theta} dt. \]
The second derivative in (7.28) is given by
\[ \frac{\partial^2 Q_{n,5}(\theta_0)}{\partial \theta^2} = \sum_{j=1}^{n} \left[ \frac{\partial \tilde{p}_{j,5}(\theta_0)}{\partial \theta} A_j \frac{\partial \tilde{p}_{j,5}(\theta_0)}{\partial \theta} \right] \]
\[ + \left( \frac{\partial \tilde{p}_{j,25}(\theta_0)}{\partial \theta} A_j \right) \frac{\partial \tilde{p}_{j,25}(\theta_0)}{\partial \theta} \]
\[ + \left( \frac{\partial \tilde{p}_{j,5}(\theta_0)}{\partial \theta} A_j \right) \frac{\partial \tilde{p}_{j,25}(\theta_0)}{\partial \theta}. \]
where $\partial \tilde{p}_{j,5}(\theta)/\partial \theta$ is given by

In the following let $f(v)$, $\tilde{m}_1(v)$ be defined similarly as $\tilde{f}(v)$, $\tilde{m}_1(\cdot)$ in (4.7) and (4.8), but with $\tilde{f}$ substituted by $\Gamma_0$. Similarly, $f_0(t; \theta)$ and $\tilde{\rho}_{i,5}(\cdot)$ are used for the same situation. Then, by Assumption 13, $\partial Q_{n,s}(\cdot)/\partial \theta$ has the first-order Taylor expansion about $\gamma_0$:

$$
\frac{\partial Q_{n,s}(\theta)}{\partial \theta} = \sum_{j=1}^{n} \left[ \frac{\partial \tilde{\rho}_{i,5}(\theta)}{\partial \theta} \hat{A}_j \frac{\partial \tilde{\rho}_{i,25}(\theta)}{\partial \theta} + \frac{\partial \tilde{\rho}_{i,5}(\theta)}{\partial \gamma} \hat{A}_j \frac{\partial \tilde{\rho}_{i,25}(\theta)}{\partial \gamma} \right] + \frac{\partial^2 Q_{n,s}(\theta)}{\partial \gamma^2} (\gamma - \gamma_0) + o_p(\sqrt{n}).
$$

(7.30)

where

$$
\frac{\partial^2 Q_{n,s}(\theta)}{\partial \gamma^2} = \sum_{j=1}^{n} \left[ \frac{\partial \tilde{\rho}_{i,5}(\theta)}{\partial \gamma} \hat{A}_j \frac{\partial \tilde{\rho}_{i,25}(\theta)}{\partial \gamma} + \frac{\partial \tilde{\rho}_{i,5}(\theta)}{\partial \gamma} \hat{A}_j \frac{\partial \tilde{\rho}_{i,25}(\theta)}{\partial \gamma} \right] + \left( \frac{\partial \tilde{\rho}_{i,25}(\theta)}{\partial \gamma} \hat{A}_j \frac{\partial \tilde{\rho}_{i,25}(\theta)}{\partial \gamma} \right).
$$

and

$$
\frac{\partial \tilde{f}_u(x, \tilde{m}_1(\cdot) \otimes \delta)}{\partial \gamma'} = -\frac{1}{S} \sum_{j=1}^{S} \left[ \frac{\partial \tilde{m}_1(t)}{\partial \gamma'} \left[ \frac{\partial \tilde{m}_1(t)}{\partial \gamma} - \tilde{m}_1(t)(W \otimes t) \right] \right] dt.
$$

Here

$$
\frac{\partial \tilde{m}_1(t)}{\partial \gamma'} = \int_{B_n} e^{-u'(v)} \frac{\partial \tilde{m}_1(v)}{\partial \gamma'} dv,
$$

$$
\frac{\partial \tilde{m}_1(v)}{\partial \gamma'} = \frac{1}{f(v)(v)} \sum_{i=1}^{n} (\tilde{m}_1(v) - v_j) \frac{\partial K_n(v - v_j)}{\partial \gamma'} (W \otimes t_k)'
$$

and

$$
K_n(v - v_j) = \frac{1}{a_n} K \left( \frac{v - v_j}{a_n} \right).
$$

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

$$
\gamma_n = \frac{1}{\gamma_n} \begin{pmatrix} \hat{A}_j \frac{\partial \tilde{\rho}_{i,5}(\theta)}{\partial \theta} \frac{\partial \tilde{\rho}_{i,5}(\theta)}{\partial \gamma} \end{pmatrix}
$$

In the following, because all functions of parameters are evaluated at $\theta_0$, we will omit it to further simplify notation. In addition, let $u_0 = x_0 - \Gamma_0 W_0$, $x_s = g(x_0; \theta_0)$, $h_p = h(x_0)$ and $f_0$ be similarly defined. Now we express the first term of (7.30) in terms of $\xi_j$, $\tilde{\xi}_j$, $\tilde{\xi}_j$, $\tilde{\xi}_j$. To this end, we write

$$
\rho_{i,5} - \rho_{i,5} = -\frac{1}{S} \sum_{j=1}^{S} \tilde{x}_j g_j h_p (f_0 - f_0) + o_p(\sqrt{n}),
$$

where

$$
\int f_s - f_s = \frac{1}{(2\pi)^k} \int_{C_0} g(t; \theta_0) \int_{B_n} e^{-u'(v)} (\tilde{m}_1(t) - m_1(t)) dv dt,
$$

$$
= \frac{1}{n(2\pi)^k} \sum_{j=1}^{n} \int_{C_0} g(t; \theta_0) \int_{B_n} e^{-u'(v)} (\tilde{m}_1(t) - m_1(t)) dv dt,
$$

$$
\int_{B_n} e^{-u'(v)} (Y - m_1(t)) K_n(v - V) dv dt,
$$

$$
= \frac{1}{n(2\pi)^k} \sum_{j=1}^{n} \int_{C_0} g(t; \theta_0) \int_{B_n} e^{-u'(v)} (\tilde{m}_1(t) - m_1(t)) dv dt,
$$

and $n_t$ is defined in (4.15). Since, by (7.24), $\lim_{n \to \infty} P(\inf_{0 \geq b_n/2}) = 1$ and $\lim_{n \to \infty} P(\inf_{0 \geq b_n/2}) = 1$, with...
probability approaching one, the absolute value of the second term of the last equation of (7.32) satisfies
\[
\left| \int_{c_0}^{1} \frac{e^{-i\eta_i \ell}}{\theta_i} \left( \int_{c_0}^{1} \frac{e^{-i\eta_i \ell}}{\theta_i} \right) d\theta_i \right| \leq \frac{1}{n \sqrt{2\pi}} \sum_{i=1}^{n} \left| \int g(t; \theta_i) \left( \int \frac{e^{-i\eta_i \ell}}{f(v)} \right) d\theta_i \right| d\eta_i dt = o_p(n^{-1/2}).
\]
where the last two equations follow from Assumptions 13 and 14 and theorem conditions. Furthermore, since \( \sum_{i} \eta_i = O_p(\sqrt{n}) \),
\[
\frac{1}{n} \sum_{i=1}^{n} \int_{c_0}^{1} \frac{e^{-i\eta_i \ell}}{\theta_i} \left( \int_{c_0}^{1} \frac{e^{-i\eta_i \ell}}{f(v)} \right) d\theta_i = o_p(n^{-1/2}).
\]
and, similarly,
\[
\frac{1}{n} \sum_{i=1}^{n} \int_{c_0}^{1} \frac{e^{-i\eta_i \ell}}{\theta_i} \left( \int_{c_0}^{1} \frac{e^{-i\eta_i \ell}}{f(v)} \right) d\eta_i d\theta_i = o_p(n^{-1/2}).
\]
It follows that
\[
\tilde{f}_n = f_n = n \sqrt{2\pi} \sum_{i=1}^{n} \int_{c_0}^{1} \frac{e^{-i\eta_i \ell}}{\theta_i} \left( \int_{c_0}^{1} \frac{e^{-i\eta_i \ell}}{f(v)} \right) d\theta_i + o_p(n^{-1/2}).
\]
Therefore, the second term on the right-hand side of (7.31) equals
\[
- \frac{1}{nS(2\pi)^k} \sum_{s=1}^{S} \sum_{j=1}^{n} \int_{c_0}^{1} \frac{e^{-i\eta_i \ell}}{\theta_i} \left( \int_{c_0}^{1} \frac{e^{-i\eta_i \ell}}{f(v)} \right) d\theta_i + o_p(\sqrt{n})
\]
where \( U_n \) is a so-called U-Statistic of degree two. By Theorem 1 of Section 2.1 of Lee (1990), pp. 76), \( U_n = \sum_{i=1}^{n} \xi_2 + o_p(\sqrt{n}) \).
Further, note that in \( \rho_{25} \) and \( \rho_{25} \) only \( \hat{\xi}_2 \) and \( \hat{\xi}_2 \) involve \( \delta \), completely analogously it can be shown that the third term on the right-hand side of (7.31) equals \( \sum_{i=1}^{n} \xi_2 + o_p(\sqrt{n}) \). Finally, it is obvious that the fourth term is \( o_p(\sqrt{n}) \).
It follows that
\[
\sum_{j=1}^{n} \frac{\partial Q_{n,5}(\theta_0)}{\partial \theta} A_j \rho_{25}(\theta_0) = \sum_{j=1}^{n} \xi_2 + \xi_2 + \xi_2 + \xi_2 + o_p(\sqrt{n}),
\]
where \( \xi_1, \xi_2, \xi_2, \xi_2 \) are given in (4.12)-(4.14). Then we can rewrite (7.30) as
\[
\frac{\partial Q_{n,5}(\theta_0)}{\partial \theta} = 2D_{n,5} \sum_{j=1}^{n} \xi_j + o_p(\sqrt{n}),
\]
where
\[
D_{n,5} = \left( \beta_0, \beta_0, \beta_0, \frac{1}{n} \frac{\partial Q_{n,5}(\theta_0)}{\partial \theta} \left( \int_{c_0}^{1} \frac{e^{-i\eta_i \ell}}{f(v)} \right) d\eta_i \right) \to D.
\]
Since \( E[\{W_j \cap (Z_j - \bar{\nu}_j W_j) \}] = 0, E\eta_j = 0 \) and \( E \left( \frac{\partial Q_{n,5}}{\partial \theta} A \rho_{25} \right) = 0 \), we have \( E_b = 0 \). Furthermore, by model assumption, Assumption 13, the covariance matrix \( C = E\xi_j \xi_j^T \) is finite. It follows by Lindeberg–Levy CLT and the Slutsky Theorem,
\[
\frac{1}{2\sqrt{\pi}} \frac{\partial Q_{n,5}(\theta_0)}{\partial \theta} \to N(0, \sigma_D)'(7.33)
\]
where \( D \) and \( C \) are as in Theorem 5.2. The theorem follows then from (7.28), (7.29), (7.33) and Assumption 16.

7.6. Proof of Theorem 5.2
Let \( F_n \) and \( F_n \) denote the distribution function of \( W \) and \( X \). Then by (4.1) and the Fubini Theorem,
\[
E_W[m_{1,n}(T_0 W) - E(Y | W)] = \int m_{1,n}(T_0 w) - E(Y | w) dF_n(w) \leq \int \int g(T_0 w + u; \theta_0) 1(\{D_T w + u \geq T_n\}) dF_n(w) \]
\[
\leq \int \int g(T_0 w + u; \theta_0) 1(\{D_T w + u \geq T_n\}) dF_n(w) dF_n(w) \leq \int \int g(x; \theta_0) 1(\{|x| \geq T_n\}) dF_n(x) \]
\[
\int \int g(x; \theta_0) 1(\{|x| \geq T_n\}) dF_n(x),
\]
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 \( T_n \to \infty \), which implies (5.5). The proof of (5.6) is analogous.

Acknowledgments
The authors are grateful to the anonymous referees for their helpful comments and suggestions. A preliminary version of this paper was completed in Wang and Hsiao (1995) while Linquan 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