Semiparametric Estimation of Models with Conditional Moment Restrictions in the Presence of Nonclassical Measurement Errors

SUYONG SONG*

January 27, 2011

Abstract

In this paper, we consider semiparametric estimation of models defined by conditional moment restrictions, which contain finite dimensional unknown parameters and infinite dimensional unknown functions. We extend these models to include the case where the unknown functions depend on endogenous variables which are contaminated by nonclassical measurement errors. A two-stage estimation procedure is proposed to recover the true conditional density of endogenous variables given conditioning variables masked by the nonclassical measurement errors, and to rectify the difficulty associated with endogeneity of the unknown functions. Specifically, we estimate conditional density of endogenous variables given conditioning variables in the first stage using sieve maximum likelihood estimation, and then estimate parameters of interest in the second stage using sieve minimum distance estimation. We show that the proposed estimator of the infinite dimensional unknown functions is consistent with a rate faster than $n^{-1/4}$ under a certain metric, and the proposed estimator of the finite dimensional unknown parameters obtains root-n asymptotic normality. Monte Carlo evidence and an application to semiparametric estimation of the shape-invariant IV Engel curves illustrate the usefulness of our method.

Key Words: Semiparametric conditional moment restrictions, method of sieve, endogeneity, nonclassical measurement error, instrumental variable

*Department of Economics, University of Wisconsin-Milwaukee. Email: songs@uwm.edu
1 Introduction

We consider the following models defined by conditional moment restrictions,

\[ E[\rho(Z, \theta_0, h_0(\cdot)) \mid X] = 0, \]

where \( Z \equiv (Y', X_1')' \), \( Y \equiv (Y_1', Y_2')' \) is a vector of endogenous (or dependent) variables, \( X_1 \) is a subset of conditioning variables \( X \equiv (X_1', X_2')' \), \( \rho() \) is a vector of generalized residual functions whose functional forms are known up to the unknown vector of finite dimensional parameters \( \theta_0 \) and the unknown functions \( (h_0 \equiv (h_{01}(\cdot), ..., h_{0q}(\cdot))) \), where the arguments of each function \( h_{0\ell}(\cdot), \ell = 1, ..., q \), may depend on different arguments, and, in particular, may depend on \( Y \). \( E[\rho(Z, \theta_0, h_0) \mid X] \) is the conditional expectation of \( \rho(Z, \theta_0, h_0) \) given \( X \). Classical model of conditional moment restrictions without the unknown functions \( h_0 \) has been exploited considerably in the literature on nonlinear parametric models (see, for instance, Hansen (1982), Chamberlain (1987), Newey (1990, 1993)). There has also been a lot of work on more general frameworks including the unknown function \( h_0 \) in the literature on nonparametric and semiparametric models (see, for instance, Robinson (1988), Powell, Stock, and Stoker (1989), Chamberlain (1992), Ichimura (1993)). In their seminal papers, Newey and Powell (2003), and Ai and Chen (2003) study method of sieves when the unknown functions \( h_0 \) depend on the endogenous variables. To be specific, they approximate the unknown functions \( h_0 \) by sieves, and apply the method of minimum distance to estimate parameters of interest. Ai and Chen find that an estimator of \( h_0 \) is consistent with a rate faster than \( n^{-1/4} \), and that an estimator of the parametric components \( \theta_0 \) is \( \sqrt{n} \) consistent, asymptotically normally distributed, and efficient, while Newey and Powell characterize sufficient identification conditions and propose a consistent estimator for the parameters of interest.

The main contribution of our setup to the literature is that the model (1) encompasses the case where the true \( Y_2 \), causes of interest, are unobserved due to nonclassical measurement errors on the true \( Y_2 \). There have been few works which simultaneously resolve both endogeneity and measurement errors imposed on the same variable of interest in nonparametric and semiparametric models, despite there being a number of empirical observations where endogenous variables are
also measurement error-laden. In the returns-to-education literature, for instance, education, the cause of interest, is endogenous in that it is correlated with unobserved ability which is an unobservable driver of income, dependent variable. Moreover, there is often erroneous reporting due to the nature of survey data. In the linear parametric models, the use of valid instruments could resolve issues of identification and estimation associated with measurement errors. However, the existence of valid instruments is not sufficient for the identification and estimation of parameters in nonlinear models, as demonstrated by Amemiya (1985) and Hsiao (1989). As a result, accounting for both endogeneity and measurement errors in nonparametric and semiparametric models is not straightforward.

In this paper, we propose a two-step estimation addressing the aforementioned issues. In the first step, a consistent estimate of the true conditional density of endogenous variables given conditioning variables, which are masked by the nonclassical measurement errors, is obtained. In the second step, a consistent estimate of parameters of interest, \( \alpha_0 \equiv (\theta_0, h_0) \), is obtained. For the first-step estimation, we make use of a method proposed by Hu and Schennach (2008), which relies on the eigenvalue-eigenfunction decomposition of an integral operator associated with joint densities of observables, and extend their method to allow for the presence of a vector of additional observable regressors. We also propose a sieve maximum likelihood estimator of conditional densities associated with the unobserved regressors of interest. We then propose a sieve minimum distance estimator of parameters, \( \alpha_0 \). Interestingly, we find that one instrument is sufficient to identify and estimate parameters of interest, even when one regressor of interest is endogenous and measurement error-laden. We also show that the sieve estimator of the infinite dimensional unknown functions is consistent with a rate faster than \( n^{-1/4} \) under certain metrics, and the sieve estimator of the finite dimensional unknown parameters is \( \sqrt{n} \) consistent and asymptotically normally distributed.

The rest of the paper is organized as follows. We describe the proposed two-step estimation in section 2. Issues of identification and estimation of distributions in presence of nonclassical measurement errors are discussed in section 3. In section 4, we prove consistency and \( n^{-1/4} \) convergence rates of the parameters from both steps. Asymptotic normality of finite-dimensional parameters of both steps is analyzed in section 5. In section 6, the finite-sample properties of the estimator are investigated via Monte Carlo studies, and the empirical application to IV Engel
curves using the British Family Expenditure Survey is presented. Section 7 briefly concludes. All technical proofs are included in the Mathematical Appendix.

2 Two-Stage Estimation

Let \( Y_1, Y_2, Y_2^*, X_1, \) and \( X_2 \) denote the support of the distribution of the random variables \( Y_1, Y_2, Y_2^*, X_1, \) and \( X_2, \) respectively. Let \( Y \equiv (Y_1, Y_2) \in \mathcal{Y} \equiv \mathcal{Y}_1 \times \mathcal{Y}_2, Y^* \equiv (Y_1, Y_2^*) \in \mathcal{Y}^* \equiv \mathcal{Y}_1 \times \mathcal{Y}_2^*, X \equiv (X_1, X_2) \in \mathcal{X} \equiv \mathcal{X}_1 \times \mathcal{X}_2. \) Suppose that the true observations \( \{(Y_i, X_i) : i = 1, 2, ..., n\} \) are drawn independently from the distribution of \( (Y, X) \) with support \( \mathcal{Y} \times \mathcal{X}, \) where \( \mathcal{Y} \) is a subset of \( \mathbb{R}^d_y \) and \( \mathcal{X} \) is a compact subset of \( \mathbb{R}^d_x. \) Suppose that the unknown distribution of \( (Y, X) \) satisfies the conditional moment restriction given by Eqn. (1), where \( \rho : \mathcal{Z} \times \mathcal{A} \rightarrow \mathbb{R}^{d_y} \) is a known mapping, up to an unknown vector of parameters, \( \alpha_0 \equiv (\theta_0, h_0) \in \mathcal{A} \equiv \Theta \times \mathcal{H}. \) We assume that \( \Theta \subseteq \mathbb{R}^{d_\theta} \) is compact with a nonempty interior, and that \( \mathcal{H} \equiv \mathcal{H}_1 \times \cdots \times \mathcal{H}_q \) is a space of continuous functions. We further assume that \( Z \equiv (Y', X_1') \in \mathcal{Z} \equiv \mathcal{Y} \times \mathcal{X}_1 \) and \( \mathcal{X}_1 \subseteq \mathcal{X}. \) We use the notation \( f_{R_1}(r_1), f_{R_1|R_2}(r_1 \mid r_2), \) and \( F_{R_1|R_2}(r_1 \mid r_2) \) to denote the marginal density of variable \( R_1, \) the conditional density of \( R_1 \) conditional on \( R_2, \) and the cumulative density of \( R_1 \) conditional on \( R_2, \) respectively.

Let \( m(x, \alpha) \equiv \int \rho(y, x_1, \theta, h(\cdot))dF_{Y|X}(y \mid x) \) denote the conditional mean function of the residuals, \( \rho(Y, X_1, \theta, h(\cdot)), \) given \( X. \) Under the assumption that model (1) identifies \( \alpha_0, \) one can solve for \( \alpha_0 \) as follows:

\[
\alpha_0 = \arg \inf_{\alpha=(\theta, h) \in \Theta \times \mathcal{H}} E \left[ m(X, \alpha) \left[ \Sigma(X) \right]^{-1} m(X, \alpha) \right]
\]

(2)

where \( \Sigma(X) \) is a positive-definite matrix for any given \( X. \) Because the conditional distribution \( F_{Y|X}(y \mid x) \) and conditional mean function \( m(x, \alpha) \) are not specified, Newey and Powell (2003) and Ai and Chen (2003) propose a sieve minimum distance (hereafter SMD) estimator that replaces \( m(X, \alpha) \) with a consistent nonparametric estimator \( \hat{m}(X, \alpha) \) and the function space \( \mathcal{H} \) with a sieve space \( \mathcal{H}_n \equiv \mathcal{H}_1 \times \cdots \times \mathcal{H}_n \) (Grenander, 1981). However, the method is infeasible in our setup because elements of the true \( Y \) (i.e., \( Y_2 \)) are unobserved so that the empirical distribution of \( (Y_i, X_i) \) cannot be used to estimate \( m(X, \alpha). \) Instead, we base an estimate of \( m(X, \alpha) \) on a sieve maximum likelihood (hereafter SML) estimator of \( F_{Y|X}(y \mid x). \) For this, we adapt a method proposed by
Hu and Schennach (2008). Let $F_{Y|X}(y \mid x)$ be absolutely continuous with respect to Lebesque measure. To be specific, the conditional mean function can be rewritten as follows: for true values $(\phi_0, \eta_0) \in \Phi \times M$

$$m(x, \alpha) \equiv \int_Y \rho(y, x_1, \theta, h(\cdot))dF_{Y|X}(y \mid x; \phi_0, \eta_0)$$
$$= \int_{Y_2} \left[ \int_{Y_1} \rho(y, x_1, \theta, h(\cdot))dF_{Y_1|Y_2X}(y_1 \mid y_2, x; \phi_0, \eta_0) \right] dF_{Y_2|X}(y_2 \mid x) \quad (3)$$
$$= \int_{Y_2} \left[ \int_{Y_1} \rho(y, x_1, \theta, h(\cdot))f_{Y_1|Y_2X}(y_1 \mid y_2, x; \phi_0, \eta_0)dy_1 \right] f_{Y_2|X}(y_2 \mid x)dy_2$$
$$= \int_{Y_2} \left[ \int_{Y_1} \rho(y, x_1, \theta, h(\cdot))f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_1; \phi_0, \eta_0)dy_1 \right] f_{Y_2|X_2X_1}(y_2 \mid x_2, x_1)dy_2,$$

where the last equality holds by the exclusion restriction specified in assumption 3.2 in the next section. Note that $\alpha_0 \equiv (\theta_0, h_0) \in \mathcal{A} \equiv \Theta \times \mathcal{H}$ are the second stage parameters, and $\beta_0 \equiv (\psi_0, f_1, f_2) \in \mathcal{B} \equiv \Psi \times \mathcal{F}_1 \times \mathcal{F}_2$ are the first stage parameters where $\psi_0 \equiv (\phi_0, \eta_0) \in \Psi \equiv \Phi \times M$ is a vector of parameters of $f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_1; \phi_0, \eta_0)$, $f_1 \equiv f_{Y_2|Y_2X_1}(y_2 \mid y_2, x_1)$, and $f_2 \equiv f_{Y_2|X_2X_1}(y_2 \mid x_2, x_1)$. In the first step, we use a SML estimation to estimate $f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_1; \phi_0, \eta_0)$ and $f_{Y_2|X_2X_1}(y_2 \mid x_2, x_1)$ needed for Eqn. (3). Then in the second step, the SMD estimator of $\alpha_0 \equiv (\theta_0, h_0)$ minimizes the sample analog of a nonparametric version of (2) with a sieve space $\mathcal{H}_n \equiv \mathcal{H}^1_n \times \cdots \times \mathcal{H}^m_n$ in place of $\mathcal{H}$:

$$\hat{\alpha}_n = \arg \min_{\alpha=(\theta,h) \in \Theta \times \mathcal{H}_n} \frac{1}{n} \sum_{i=1}^n \hat{m}(X_i, \alpha)[\hat{\Sigma}(X_i)]^{-1}\hat{m}(X_i, \alpha), \quad (4)$$

where $\mathcal{H}_n$ is some finite-dimensional approximation space that becomes dense in $\mathcal{H}$ as sample size $n \to \infty$ (e.g., Fourier series, power series, splines, wavelets, etc.), $\hat{\Sigma}(X)$ is a consistent estimator of $\Sigma(X)$, and $\hat{m}(X, \alpha)$ is the plug-in SML estimator of $m(X, \alpha)$ for any fixed $\alpha = (\theta, h_n)$:

$$\hat{m}(X, \alpha)$$
$$\equiv \int_{Y_2} \left[ \int_{Y_1} \rho(y, x_1, \theta, h_n(\cdot))\hat{f}_{Y_1|Y_2X_1}(y_1 \mid y_2, x_1; \hat{\phi}_n, \hat{\eta}_n)dy_1 \right] \hat{f}_{Y_2|X_2X_1}(y_2 \mid x_2, x_1)dy_2. \quad (5)$$

We now introduce useful spaces of smooth functions to analyze how well a sieve can approxi-
mate either $H$ or $M$. Let $\xi \in V \subset \mathbb{R}^{d\xi}$, $\| \cdot \|_E$ denote the Euclidean norm, and

$$\nabla^a g(\xi) \equiv \frac{\partial^{a_1+a_2+\cdots+a_{d\xi}} g(\xi)}{\partial \xi_1^{a_1} \cdots \partial \xi_{d\xi}^{a_{d\xi}}}$$

denote the $\sum_{i=1}^{d\xi} a_i$-th derivative where $a = (a_1, a_2, \ldots, a_{d\xi})$' is a vector of nonnegative integers. Let $\gamma$ denote the largest integer satisfying $\gamma < \gamma'$. The Hölder space $\Lambda^{\gamma}(V)$ of order $\gamma > 0$ is a space of functions $g : V \rightarrow \mathbb{R}$ such that the first $\gamma$ derivative is bounded and the $\gamma$-th derivatives are Hölder continuous with the exponent $\gamma - \gamma' \in (0, 1]$, i.e., for all $\xi, \xi' \in V$ and some constant $c$

$$\max_{\sum_{i=1}^{d\xi} a_i = \gamma} |\nabla^a g(\xi) - \nabla^a g(\xi')| \leq c(\|\xi - \xi'\|_E)^{\gamma - \gamma'}.$$

The space $\Lambda^{\gamma}(V)$ becomes a Banach space under the Hölder norm:

$$\|g\|_{\Lambda^{\gamma}} = \sup_{\xi} |g(\xi)| + \max_{\sum_{i=1}^{d\xi} a_i = \gamma} \sup_{\xi \neq \xi'} \frac{|\nabla^a g(\xi) - \nabla^a g(\xi')|}{\|\xi - \xi'\|_E^{\gamma - \gamma'}} < \infty.$$

A Hölder ball (of radius $c$) is defined as $\Lambda^{\gamma}_c(V) \equiv \{ g \in \Lambda^{\gamma}(V) : \|g\|_{\Lambda^{\gamma}} \leq c < \infty \}$. Let $\omega(\cdot)$ be a positive continuous weight function on $V$ where $\omega(\xi) = (1 + \|\xi\|_E^2)^{-\varsigma/2}, \varsigma > \gamma > 0$. Denote $\Lambda^{\gamma,\omega}(V)$ as the weighted Hölder space with a weighted Hölder norm $\|g\|_{\Lambda^{\gamma,\omega}} \equiv \|\tilde{g}\|_{\Lambda^{\gamma}}$ for $\tilde{g}(\xi) \equiv g(\xi)\omega(\xi)$. Also define a weighted Hölder ball $\Lambda^{\gamma,\omega}_c(V) \equiv \{ g \in \Lambda^{\gamma,\omega}(V) : \|g\|_{\Lambda^{\gamma,\omega}} \leq c < \infty \}$.

3 Identification and Estimation of Distributions

3.1 Identification of Distributions

In this section, we consider the identification of two densities, $f_{Y_1|Y_2,X_1}(y_1 \mid y_2, x_1; \psi)$ and $f_{Y_2|X_2,X_1}(y_2 \mid x_2, x_1)$. Hu and Schennach (2008) show that the joint distribution of $y_1$ and $y_2$ is identified from knowledge of the distribution of all observed variables. For our case, we straightforwardly extend the treatment in Hu and Schennach (2008) to allow for the presence of a vector $X_1$ of additional observable regressors. We consider $Y_2, Y_2^*$, and $X_2$ to be jointly continuously distributed, while $Y_1$ and $X_1$ can be either continuous or discrete. We first state a useful note that a function of three
variables can be associated with an integral operator.

**Definition 3.1** Let \( R_1, R_2, R_3 \) and \( R_4 \) denote four random variables with respective supports \( \mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \) and \( \mathcal{R}_4, \) distributed according to the joint density \( f_{R_1 R_2 R_3 R_4}(r_1, r_2, r_3, r_4). \) Given four corresponding spaces \( \mathcal{G}(\mathcal{R}_1), \mathcal{G}(\mathcal{R}_2), \mathcal{G}(\mathcal{R}_3), \) and \( \mathcal{G}(\mathcal{R}_4) \) of functions with domains \( \mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \) and \( \mathcal{R}_4, \) respectively, let \((i)\) \( L_{R_1 | R_2 R_3} \) denote an integral operator mapping \( g \in \mathcal{G}(\mathcal{R}_2) \) to \( L_{R_1 | R_2 R_3} g \in \mathcal{G}(\mathcal{R}_1) \) for a given \( r_3 \) defined by

\[
[L_{R_1 | R_2 R_3} g](r_1) \equiv \int_{\mathcal{R}_2} f_{R_1 | R_2 R_3}(r_1 \mid r_2, r_3) g(r_2) \, dr_2; \quad (6)
\]

\((ii)\) \( L_{r_1 R_2 | R_3} \) denote an integral operator mapping \( g \in \mathcal{G}(\mathcal{R}_3) \) to \( L_{r_1 R_2 | R_3} g \in \mathcal{G}(\mathcal{R}_2) \) for a given \((r_1, r_4)\) defined by

\[
[L_{r_1 R_2 | R_3} g](r_2) \equiv \int_{\mathcal{R}_3} f_{r_1 R_2 | R_3}(r_1, r_2 \mid r_3, r_4) g(r_3) \, dr_3; \quad (7)
\]

\((iii)\) \( \Delta_{r_1 | R_2 r_3} \) denote a diagonal operator mapping \( g \in \mathcal{G}(\mathcal{R}_2) \) to \( \Delta_{r_1 | R_2 r_3} g \in \mathcal{G}(\mathcal{R}_2) \) for a given \((r_1, r_3)\) defined by

\[
\Delta_{r_1 | R_2 r_3} g \equiv f_{r_1 R_2 R_3}(r_1 \mid r_2, r_3) g(r_2). \quad (8)
\]

For the identification of distributions, we assume following hypotheses. Note that the absence of correctly measured regressors, \( X_1, \) draws on similar assumptions to those in Hu and Schennach (2008).

**Assumption 3.1** (i) The joint density of \( Y_1 \) and \( Y_2, Y_2^*, X_1, X_2 \) admits a bounded density with respect to the product measure of some dominating measure \( \mu \) (defined on \( \mathcal{Y}_1 \)) and the Lebesgue measure on \( \mathcal{Y}_2 \times \mathcal{Y}_2^* \times X_1 \times X_2. \) (ii) All marginal and conditional densities are also bounded.

**Assumption 3.2** (i) \( f_{Y_1 | Y_2 Y_2^* X_1 X_2}(y_1 \mid y_2, y_2^*, x_1, x_2) = f_{Y_1 | Y_2 X_1}(y_1 \mid y_2, x_1, x_2) \) for all \((Y_1, Y_2, Y_2^*, X_1, X_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Y}_2^* \times X_1 \times X_2 \) and (ii) \( f_{Y_2^* | Y_2 X_1 X_2}(y_2^* \mid y_2, x_1, x_2) = f_{Y_2^* | Y_2 X_1}(y_2^* \mid y_2, x_1) \) for all \((Y_2, Y_2^*, X_1, X_2) \in \mathcal{Y}_2 \times \mathcal{Y}_2^* \times X_1 \times X_2. \)

**Assumption 3.3** The operators \( L_{Y_2 | Y_2 x_1} \) and \( L_{Y_2^* | X_2 x_1} \) are one-to-one (for either \( \mathcal{G} = \mathcal{L}^1 \) or
\( \mathcal{G} = \mathcal{L}^1_{\text{bnd}} \) where \( \mathcal{L}^1(A) \) is the set of all absolutely integrable functions with domain \( A \) endowed with the norm \( ||g||_1 = \int_A |g(a)| \, da \) and where \( \mathcal{G} = \mathcal{L}^1_{\text{bnd}} \) is the set of functions in \( \mathcal{L}^1(A) \) that are also bounded such that \( \sup_{a \in A} |g(a)| < \infty \).

**Assumption 3.4** For any \( x_1 \in X_1 \) and any \( \tilde{y}_2, \bar{y}_2 \in Y_2 \), the set \( \{ y_1 : f_{Y_1|X_2}(y_1 \mid \tilde{y}_2, x_1) \neq f_{Y_1|X_2}(y_1 \mid \bar{y}_2, x_1) \} \) has positive probability (under the marginal of \( Y_1 \)) whenever \( \tilde{y}_2 \neq \bar{y}_2 \).

**Assumption 3.5** For any given \( x_1 \in X_1 \), there exists a known functional \( M \) such that \( M[f_{Y^*_2|X_1}(\cdot \mid y_2, x_1)] = y_2 \) for all \( y_2 \in Y_2 \).

A few remarks are in order. Assumption 3.1 restricts all densities to regular bounded densities. Assumption 3.2 states conditional independence restrictions which have been imposed by Altonji and Matzkin (2005), White and Chalak (2006), Chalak and White (2007), and Hoderlein and Mammen (2007), among others. To be specific, Assumption 3.2(i) states that \( Y^*_2, X_2 \) do not provide further information on \( Y_1 \), given \( Y_2, X_1 \). Similarly, Assumption 3.2(ii) indicates that \( X_2 \) does not provide further information on \( Y^*_2 \), given \( Y_2, X_1 \). Assumption 3.3 is associated with restrictions on the relationships between \( Y_2, Y^*_2, X_2, \) and \( X_1 \), which have been phrased as singular value decompositions with nonzero singular values (Darolles, Florens, and Renault (2002)), nonsingularity (Hall and Horowitz (2005), Horowitz (2006)), and completeness (or bounded completeness) (Newey and Powell (2003), Blundell, Chen, and Kristensen (2007)). Assumption 3.4 states a fairly weak condition which is only violated if the distribution of \( Y_1 \) conditional on \( Y_2, X_1 \) is identical at different values of \( Y_2 \). Assumption 3.5 places restrictions on some measure of the location of a density, denoted by \( M \). The assumption is essential in that it enables the model to include nonclassical measurement errors as well as classical measurement errors\(^1\). It is invoked by the observation that, even though the measurement error may have a nonzero mean conditional on the true value of the variable, other measures of location, such as the median, mode, or quantile, could be zero. The next theorem provides identification results of unknown distributions.

**Theorem 3.1** Under Assumptions 3.1 – 3.5, given the true observed density \( f_{Y^*|X}(y^* \mid x) \equiv \)

\(^1\)For instance, \( M \) could be the mean, mode, and \( \tau \) quantile: \( M[f] = \int_{Y^*_2} y_2 f_{Y^*_2}(y_2) \, dy_2, M[f] = \arg \max_{y_2 \in Y^*_2} f_{Y^*_2}(y_2), \) and \( M[f] = \inf \{ y_2 \in Y_2 : \int 1\{y_2 \leq y_2\} f_{Y^*_2}(y_2) \, dy_2 \geq \tau \} \), respectively.
\[ f_{Y_1 Y_2 | X_1 X_2}(y_1, y_2' \mid x_1, x_2), \text{ the equation} \]

\[ f_{Y^* | X}(y^* \mid x) = \int_{\mathcal{Y}_2} f_{Y_1 | Y_2 X_1}(y_1 \mid y_2, x_1) f_{Y_2^* | Y_2 X_1}(y_2^* \mid y_2, x_1) f_{Y_2 | X_2 X_1}(y_2 \mid x_2, x_1) dy_2 \]  \hspace{1cm} (9)

admits a unique solution \((f_{Y_1 | Y_2 X_1}, f_{Y_2^* | Y_2 X_1}, f_{Y_2 | X_2 X_1})\) for all \(y_1 \in \mathcal{Y}_1, y_2^* \in \mathcal{Y}_2^*, x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2.\)

The result is parallel to Eqn.(5) in Theorem 1 of Hu and Schennach (2008). The integral equation relates the joint densities of the observables to the product of the joint densities of the unobservables. The identification of unobserved densities enables us to propose the first-stage estimation procedure, and, in turn, to estimate the parameters of interest in the second-stage via estimates of \(f_{Y_1 | Y_2 X_1}\) and \(f_{Y_2 | X_2 X_1}.\)

### 3.2 Estimation Using Sieve Maximum Likelihood

Theorem 3.1 implies that the true value \(\beta_0\) is obtained by the maximization problem:

\[
\beta_0 = (\psi_0, f_{Y_2^* | Y_2 X_1}, f_{Y_2 | X_2 X_1})' \hspace{1cm} (10)
\]

\[
= \arg \max_{\beta=(\psi, f_1, f_2) \in B} E \left( \ln \int_{\mathcal{Y}_2} f_{Y_1 | Y_2 X_1}(y_1 \mid y_2, x_1; \psi) f_1(y_2^* \mid y_2, x_1) f_2(y_2 \mid x_2, x_1) dy_2 \right),
\]

where \(B \equiv \Psi \times \mathcal{F}_1 \times \mathcal{F}_2\) with \(\Psi \equiv \Phi \times \mathcal{M}.\) We impose some restrictions on the sets \(\mathcal{M}, \mathcal{F}_1,\) and \(\mathcal{F}_2\) to which the functions \(\eta, f_{Y_2^* | Y_2 X_1},\) and \(f_{Y_2 | X_2 X_1}\) belong, respectively, in the following assumptions.

**Assumption 3.6** \(\eta \in \Lambda_{c_{\gamma_1, \omega}}(U)\) where \(\gamma_1 > 1.\)

**Assumption 3.7** \(f_1 \in \Lambda_{c_{\gamma_1, \omega}}(\mathcal{Y}_2^* \times \mathcal{Y}_2 \times \mathcal{X}_1)\) where \(\gamma_1 > 1\) and \(\int_{\mathcal{Y}_2^*} f_1(y_2^* \mid y_2, x_1) dy_2^* = 1\) for all \(y_2 \in \mathcal{Y}_2, x_1 \in \mathcal{X}_1.\)

**Assumption 3.8** \(f_2 \in \Lambda_{c_{\gamma_1, \omega}}(\mathcal{Y}_2 \times \mathcal{X}_2 \times \mathcal{X}_1)\) where \(\gamma_1 > 1\) and \(\int_{\mathcal{Y}_2} f_2(y_2 \mid x_2, x_1) dy_2 = 1\) for all \(x_2 \in \mathcal{X}_2, x_1 \in \mathcal{X}_1.\)
Then we define three sets as follows:

\[ M = \{ \eta(\cdot, \cdot, \cdot) : \text{Assumption 3.6 holds} \}, \]
\[ F_1 = \{ f_1(\cdot | \cdot, \cdot) : \text{Assumption 3.3, 3.5, and 3.7 hold} \}, \]
\[ F_2 = \{ f_2(\cdot | \cdot, \cdot) : \text{Assumption 3.3 and 3.8 hold} \}. \]

As in Eqn. (2), the optimization method provides an inconsistent estimator for \( \beta_0 \) or a consistent estimator which converges slowly when the function spaces \( M, F_1, \) and \( F_2 \) are large. Thus, we replace \( M, F_1, \) and \( F_2 \) with finite-dimensional compact parameter spaces \( M_n, F_{1n}, \) and \( F_{2n}, \) respectively, where

\[ M_n = \{ \eta(\xi_1, \xi_2, \xi_3) = p_{kn}(\xi_1, \xi_2, \xi_3) \delta \text{ for all } \delta \text{ s.t. Assumption 3.6 holds} \}, \]
\[ F_{1n} = \{ f(y_2 | y_2, x_1) = p_{kn}(y_2, y_2, x_1) \rho \text{ for all } \rho \text{ s.t. Assumption 3.3, 3.5, and 3.7 hold} \}, \]
\[ F_{2n} = \{ f(y_2 | x_2, x_1) = p_{kn}(y_2, x_2, x_1) \pi \text{ for all } \pi \text{ s.t. Assumption 3.3 and 3.8 hold} \}. \]

Let the projection of the true parameter \( \beta_0 \) onto the space \( B_n \) where \( B_n = \Psi_n \times F_{1n} \times F_{2n} \) with \( \Psi_n = \Phi \times M_n; \)

\[ \Pi_n \beta \equiv \beta_n = \arg \max_{\beta_n = (\psi_n, f_{1n}, f_{2n})} E \left( \ln \int_{Y_2} f_{Y_1|Y_2,X_1}(y_1 | y_2, x_1; \psi) f_1(y_2 | y_2, x_1) f_2(y_2 | x_2, x_1) dy_2 \right). \]

Then a corresponding measurement-error robust sieve maximum likelihood estimator of \( \beta_0 \) maximizes the sample analog of Eqn. (10) with \( \Psi \times F_1 \times F_2 \) restricted to the sieve space \( \Psi_n \times F_{1n} \times F_{2n}: \)

\[ \hat{\beta}_n = (\hat{\psi}_n, \hat{f}_{1n}, \hat{f}_{2n})' \]
\[ = \arg \max_{(\psi, f_1, f_2) \in B_n} \frac{1}{n} \sum_{i=1}^{n} \ln \int_{Y_2} f_{Y_1|Y_2,X_1}(y_{1i} | y_2, x_{1i}; \psi) f_1(y_{2i} | y_2, x_{1i}) f_2(y_2 | x_{2i}, x_{1i}) dy_2. \]
4 Consistency and Convergence Rates

4.1 Consistency

In this section, we first obtain consistency of the SML estimator \( \hat{\beta} \) for \( \beta_0 \equiv (\psi_0, f_{Y^*|X}(y^*_2 \mid y_2, x_1), f_{Y^*|X}(y^*_2 \mid x_2, x_1)) \) under a strong metric \( \| \cdot \|_{s,\beta} \) and the SMD estimator \( \hat{\alpha} \) for \( \alpha_0 \equiv (\theta_0, h_0) \) under a strong metric \( \| \cdot \|_{s,\alpha} \) by applying the results in Newey and Powell (2003). Following Ai and Chen (2003), we then establish that \( \hat{\beta} \) and \( \hat{\alpha} \) converge to \( \beta_0 \) and \( \alpha_0 \) at a rate faster than \( n^{-1/4} \) under suitably constructed weaker metrics \( \| \cdot \|_{\beta} \) and \( \| \cdot \|_{\alpha} \), respectively. Let \( (Y^*, X')' \) be a vector of observed variables for \( Y^* \in \mathcal{Y}^*, X \in \mathcal{X} \). Define \( \| \beta \|_{s,\beta} \equiv \| \phi \|_E + \| \eta \|_{\infty,\omega} + \| f_1 \|_{\infty,\omega} + \| f_2 \|_{\infty,\omega} \) where \( \| g \|_{\infty,\omega} \equiv \sup_{\xi} |g(\xi)\omega(\xi)| \) with weight function \( \omega(\xi) = (1 + \| \xi \|_E^2)^{-\gamma_1/2}, \gamma_1 > 0 \). Note that the meaning of \( \xi \) depend on the domain of \( g \) (e.g., when \( g = f_2, \xi = (y_2, x_2, x_1) \)).

Assumption 4.1  
(i) The data \( \{ (Y_i^*, X_i) \}_{i=1}^n \) are i.i.d. (ii) The density of \( (Y^*, X')' \), \( f_{Y^*X} \), satisfies \( \int \omega(Y^*, X)^{-2} f_{Y^*X}(y^*, x) d(y^*, x) < \infty \).

Assumption 4.2  \( \alpha_0 \in \mathcal{A} \) is the only \( \alpha \in \mathcal{A} \) satisfying \( m(X, \alpha) = 0 \).

Assumption 4.3  
(i) \( \hat{\Sigma}(X) = \Sigma(X) + o_p(1) \) uniformly over \( X \in \mathcal{X} \). (ii) \( \Sigma(X) \) is finite positive-definite uniformly over \( X \in \mathcal{X} \).

Assumption 4.4  
(i) There is a metric \( \| \cdot \|_{s,\alpha} \) such that \( \mathcal{A} \equiv \Theta \times \mathcal{H} \) is compact under \( \| \cdot \|_{s,\alpha} \). (ii) For any \( \alpha \in \mathcal{A} \), there exists \( \Pi_n \alpha \in \mathcal{A}_n \equiv \Theta \times \mathcal{H}_n \) such that \( \| \Pi_n \alpha - \alpha \|_{s,\alpha} = o(1) \).

Assumption 4.5  
(i) There is a metric \( \| \cdot \|_{s,\beta} \) such that \( \mathcal{B} \equiv \Psi \times \mathcal{F}_1 \times \mathcal{F}_2 \) is compact under \( \| \cdot \|_{s,\beta} \). (ii) For any \( \beta \in \mathcal{B} \), there exists \( \Pi_n \beta \in \mathcal{B}_n \equiv \Psi_n \times \mathcal{F}_{1n} \times \mathcal{F}_{2n} \) with \( \psi_n \equiv \Phi \times \mathcal{M}_n \) such that \( \| \Pi_n \beta - \beta \|_{s,\beta} = o(1) \).

Assumption 4.6  
(i) \( E[|\rho(Z, \alpha_0)|^2 \mid X] \) is bounded. (ii) \( \rho(Z, \alpha) \) is Hölder continuous in \( \alpha \in \mathcal{A} \).

Assumption 4.7  
(i) \( E[|\ln f_{Y^*X}(y^* \mid x)|^2] \) is bounded. (ii) There exists a measurable func-
tion $h_1(y^*, x)$ with $E[|h_1(y^*, x)|^2] < \infty$ such that for any $\beta = (\bar{\psi}, \bar{f}_1, \bar{f}_2) \in B$,  
\[ \frac{|f_{Y^*|X}^{[1]}(y^* | x; \bar{\beta}, \bar{\omega})|}{f_{Y^*|X}^{[1]}(y^* | x; \beta)} \leq h_1(y^*, x), \]
where $f_{Y^*|X}^{[1]}(y^* | x; \bar{\beta}, \bar{\omega})$ is defined as $(\frac{d}{dt} f_{Y^*|X}(y^* | x; \bar{\beta} + t\bar{\omega}))_{|t=0}$ with each linear term, that is, $\frac{d}{d\psi} f_{Y^*|X}^{[1]}(y^* | x, \bar{f}_1, \bar{f}_2, replaced by its absolute value, and $\bar{\omega}(\xi, y_2, y_2, x, x_1) = [1, \omega^{-1}(\xi), \omega^{-1}((y_2, y_2, x, x_1'))', \omega^{-1}((y_2, x, x_1'))']$ with $\xi \in \mathcal{U}$. (The explicit expression of $f_{Y^*|X}^{[1]}(y^* | x; \bar{\beta}, \bar{\omega})$ can be found in the proof of Theorem 4.1(i).)

**Assumption 4.8**  
(i) $k_{1n} \to +\infty$.

**Assumption 4.9**  
(i) $k_{n}/n \to 0$.

**Theorem 4.1**  
(i) Under Assumptions 3.1-3.8, 4.5 (i) and (ii), 4.7 (i) and (ii), and 4.9, we have $\|\hat{\beta}_n - \beta_0\|_{s, \beta} = o_p(1)$.

(ii) Under Assumptions 3.1-3.8, 4.1 (i), 4.2, 4.3 (i) and (ii), 4.4 (i) and (ii), 4.5 (i) and (ii), 4.6 (i) and (ii), 4.7 (i) and (ii), 4.8 (i), and 4.9 (i), we have $\|\hat{\alpha}_n - \alpha_0\|_{s, \alpha} = o_p(1)$.

Theorem 4.1 provides consistency results under the metrics $\| \cdot \|_{s, \beta}$ and $\| \cdot \|_{s, \alpha}$, which are stepping stones to establishing the asymptotic normality of $\hat{\phi}$ and $\hat{\theta}$, respectively.

### 4.2 Convergence Rates

As in Ai and Chen (2003) and Hu and Schennach (2008), we now consider $n^{-1/4}$ convergence rates of $\hat{\beta}_n$ and $\hat{\alpha}_n$ under weaker metrics, which are sufficient to establish the asymptotic normality and $\sqrt{n}$-consistency results. First, we recall the weaker metric introduced by Ai and Chen (2003).

Suppose that the parameter space $B$ is connected in the sense that for any two points $\beta_1, \beta_2 \in B$, there exists a continuous path $\{\beta(t) : t \in [0, 1]\}$ in $B$ such that $\beta(0) = \beta_1$ and $\beta(1) = \beta_2$. And suppose that $B$ is convex at the true value $\beta_0$ in the sense that, for any $\beta \in B, (1 - t)\beta_0 + t\beta \in B$ for small $t > 0$. Furthermore, suppose that for almost all $D$ and any $\beta \in B$, $\ln f_{Y^*|X}(D, (1 - t)\beta_0 + t\beta)$ is continuously differentiable at $t = 0$. Similarly, suppose that for any two points $\alpha_1, \alpha_2 \in A$, there exists a continuous path $\{\alpha(\tau) : \tau \in [0, 1]\}$ in $A$ such that $\alpha(0) = \alpha_1$ and
\( \alpha(1) = \alpha_2 \). Also, suppose that \( \mathcal{A} \) is convex at the true value \( \alpha_0 \), and suppose that for almost all \( X \), 
\( m(X, (1 - \tau)\alpha_0 + \tau \alpha) \) is continuously differentiable at \( \tau = 0 \).

Denote the first pathwise derivative of \( \ln f_{Y^* \mid X}(y^* \mid x; \beta_0) \) at the direction \( [\beta - \beta_0] \) evaluated at \( \beta_0 \) by:

\[
\frac{d \ln f_{Y^* \mid X}(y^* \mid x; \beta_0)}{d \beta} [\beta - \beta_0] \equiv \frac{d \ln f_{Y^* \mid X}(y^* \mid x; (1 - t)\beta_0 + t\beta)}{dt} \bigg|_{t=0}
\]

almost everywhere (under the probability measure of \((Y^*, X)) \) and for \( \beta_1, \beta_2 \in \mathcal{B} \) denote

\[
\frac{d \ln f_{Y^* \mid X}(y^* \mid x; \beta_0)}{d \beta} [\beta_1 - \beta_2] \equiv \frac{d \ln f_{Y^* \mid X}(y^* \mid x; \beta_0)}{d \beta} [\beta_1 - \beta_0] - \frac{d \ln f_{Y^* \mid X}(y^* \mid x; \beta_0)}{d \beta} [\beta_2 - \beta_0].
\]

Specifically, the pathwise derivative is denoted by:

\[
\frac{d \ln f_{Y^* \mid X}(y^* \mid x; \beta_0)}{d \beta} [\beta - \beta_0] = \frac{1}{f_{Y^* \mid X}(y^* \mid x; \beta_0)} \left\{ \int_{y_2} \frac{d}{d\psi} f_{Y_2 \mid X_1}(y_1 \mid y_2, x_1, \psi_0) \left[ \psi - \psi_0 \right] 
\times f_{Y^*_2 \mid Y_2 X_1}(y^*_2 \mid y_2, x_1) f_{Y_2 \mid X_2 X_1}(y_2 \mid x_2, x_1) dy_2 
+ \int_{y_2} f_{Y_1 \mid Y_2 X_1}(y_1 \mid y_2, x_1, \psi_0) \left[ f_1(y_2 \mid y_2, x_1) - f_{Y^*_2 \mid Y_2 X_1}(y^*_2 \mid y_2, x_1) \right] f_{Y_2 \mid X_2 X_1}(y_2 \mid x_2, x_1) dy_2 
+ \int_{y_2} f_{Y_1 \mid Y_2 X_1}(y_1 \mid y_2, x_1, \psi_0) f_{Y^*_2 \mid Y_2 X_1}(y^*_2 \mid y_2, x_1) f_2(y_2 \mid x_2, x_1) - f_{Y_2 \mid X_2 X_1}(y_2 \mid x_2, x_1) dy_2 \right\}.
\]

For any \( \beta_1, \beta_2 \in \mathcal{B} \), the metric is defined as

\[
\| \beta_1 - \beta_2 \|_{\mathcal{B}} \equiv \sqrt{E \left\{ \left( \frac{d \ln f_{Y^* \mid X}(y^* \mid x; \beta_0)}{d \beta} [\beta_1 - \beta_2] \right)^2 \right\}},
\]

Similarly, denote the first pathwise derivative of \( \rho(Z, \alpha_0) \) at the direction \( [\alpha - \alpha_0] \) evaluated at \( \alpha_0 \) by:

\[
\frac{d \rho(Z, \alpha_0)}{d \alpha} [\alpha - \alpha_0] \equiv \frac{d \rho(Z, (1 - \tau)\alpha_0 + \tau \alpha)}{d \tau} \bigg|_{\tau=0}.
\]
almost everywhere (under the probability measure of \( Z \)) and for any \( \alpha_1, \alpha_2 \in A \) denote

\[
\frac{d\rho(Z, \alpha)}{d\alpha} [\alpha_1 - \alpha_2] \equiv \frac{d\rho(Z, \alpha_0)}{d\alpha} [\alpha_1 - \alpha_0] - \frac{d\rho(Z, \alpha_0)}{d\alpha} [\alpha_2 - \alpha_0],
\]

\[
\frac{dm(X, \alpha)}{d\alpha} [\alpha_1 - \alpha_2] \equiv E \left\{ \frac{d\rho(Z, \alpha_0)}{d\alpha} [\alpha_1 - \alpha_2] \bigg| X \right\}.
\]

Also, for any \( \alpha_1, \alpha_2 \in A \), the metric \( \| \cdot \|_\alpha \) is defined as

\[
\| \alpha_1 - \alpha_2 \|_\alpha \\
\equiv \sqrt{E \left\{ \left( \frac{dm(X, \alpha_0)}{d\alpha} [\alpha_1 - \alpha_2] \right)^2 \right\}^{1/2}}.
\]

**Assumption 4.3** (iii) \( \hat{\Sigma}(X) = \Sigma(X) + o_p(n^{-1/4}) \) uniformly over \( X \in \mathcal{X} \).

**Assumption 4.4** (iii) There is a constant \( \mu_1 > 0 \) such that for any \( \alpha \in A \), there exists \( \Pi_n \alpha \in A_n \) satisfying \( \| \Pi_n \alpha - \alpha \|_\alpha = O(k_n^{-\mu_1}) \), and \( k_n^{-\mu_1} = o(n^{-1/4}) \).

**Assumption 4.5** (iii) There is a constant \( \gamma_1 > 1 \) such that for any \( \beta \in B \), there exists \( \Pi_n \beta \in B_n \) satisfying \( \| \Pi_n \beta - \beta \|_\beta = O(k_n^{-\gamma_1/d_1}) \), and \( k_n^{-\gamma_1/d_1} = o(n^{-1/4}) \).

**Assumption 4.6** (iii) Each element of \( \rho(Z, \alpha) \) satisfies an envelope condition in \( \alpha \in A_n \); (iv) each element of \( \rho(\cdot, \alpha_0) \in \Lambda_c(\mathcal{X}) \) with \( \gamma > d_x/2 \), for all \( \alpha \in A_n \).

**Assumption 4.7** (iii) \( \ln f_{Y*|X}(y^* \mid x; \beta) \) satisfies an envelope condition in \( \beta \in B_n \); (iv) \( \ln f_{Y*|X}(y^* \mid x; \beta) \) is the dimension of \( (Y^*, X) \) for some constant \( c > 0 \) with \( \gamma > d_{(Y, X)}/2 \), for all \( \beta \in B_n \), where \( d_{(Y, X)} \) is the dimension of \( (Y^*, X) \).

Denote \( \xi_{0n} \equiv \sup_{(\xi_1, \xi_2, \xi_3) \in \mathcal{U}(\mathcal{Y}_n \times \mathcal{Y}_n \times \mathcal{Y}_n)} \| p^{k_n}(\xi_1, \xi_2, \xi_3) \|_E^2 \), which is nondecreasing in \( k_n \). Let \( N(\varepsilon, B_n, \| \cdot \|_{s, \beta}) \) and \( N(\delta, A_n, \| \cdot \|_{s, \alpha}) \) denote the minimal number of radius \( \varepsilon \) covering balls of \( B_n \) under the \( \| \cdot \|_{s, \beta} \) metric, and the minimal number of radius \( \delta \) covering balls of \( A_n \) under the \( \| \cdot \|_{s, \alpha} \) metric, respectively.

**Assumption 4.8** (iii) \( k_1 n \times \ln n \times \xi_{0n}^2 \times n^{-1/2} = o(1) \); (iii) \( \ln [N(\delta^{1/n}, A_n, \| \cdot \|_{s, \alpha}) \leq \text{const.} \times k_1 n \times \ln (k_1 n / \delta) \).
Assumption 4.9  (ii) \( k_n \times \ln n \times \xi_0^2 \times n^{-1/2} = o(1) \); (iii) \( \ln[N(\varepsilon, \mathcal{B}_n, \| \cdot \|_{s,\beta})] \leq \text{const.} \times k_n \times \ln(k_n/\varepsilon) \).

Assumption 4.10  (i) \( A \) is convex in \( \alpha_0 \), and \( p(Z, \alpha) \) is pathwise differentiable at \( \alpha_0 \); (ii) for some \( c_1, c_2 > 0 \),

\[
c_1 E\{m(X, \alpha)'\Sigma(X)^{-1}m(X, \alpha)\} \leq \|\alpha - \alpha_0\|_\alpha^2 \leq c_2 E\{m(X, \alpha)'\Sigma(X)^{-1}m(X, \alpha)\}
\]

holds for all \( \alpha \in A_n \) with \( \|\alpha - \alpha_0\|_{s,\alpha} = o(1) \).

Assumption 4.11  (i) \( B \) is convex in \( \beta_0 \) and \( f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_1; \psi) \) is pathwise differentiable at \( \psi_0 \); (ii) for some \( c_1, c_2 > 0 \),

\[
c_1 E\left\{ \ln \frac{f_{Y^*|X}(y^* \mid x; \beta_0)}{f_{Y^*|X}(y^* \mid x; \beta)} \right\} \leq \|\beta - \beta_0\|_\beta^2 \leq c_2 E\left\{ \ln \frac{f_{Y^*|X}(y^* \mid x; \beta_0)}{f_{Y^*|X}(y^* \mid x; \beta)} \right\}
\]

holds for all \( \beta \in B_n \) with \( \|\beta - \beta_0\|_{s,\beta} = o(1) \).

Theorem 4.2  (i) Under Assumptions 3.1-3.8, 4.1, 4.5, 4.7, 4.9 and 4.11, we have \( \|\hat{\beta}_n - \beta_0\|_\beta = o_p(n^{-1/4}) \).

(ii) Under Assumptions 3.1-3.8 and 4.1-4.11, we have \( \|\hat{\alpha}_n - \alpha_0\|_\alpha = o_p(n^{-1/4}) \).

5 Asymptotic Normality and Efficiency

5.1 Asymptotic Normality and Efficiency

We consider the asymptotic normality of \( \hat{\phi}_n \) and \( \hat{\theta}_n \), and efficiency of a three-step estimation of \( \hat{\theta}_n \). We first introduce important notation aligning with that of Ai and Chen (2003) and Hu and Schennach (2008). Let \( \mathbf{V}_1 \) denote the closure of the linear span of \( \mathcal{B} - \{\beta_0\} \) under the metric \( \| \cdot \|_\beta \) (i.e., \( \mathbf{V}_1 = \mathcal{R}^{d_\phi} \times \mathcal{W}_1 \) with \( \mathcal{W}_1 \equiv \mathcal{M} \times \mathcal{F}_1 \times \mathcal{F}_2 - \{ (\eta_0, f_{Y_2^*|Y_2X_1, f_{Y_2|X_2X_1}}) \} \) and \( \langle \mathbf{V}_1, \| \cdot \|_\beta \rangle \) is a Hilbert space with the inner product:

\[
\langle v_{11}, v_{12} \rangle_\beta = E\left\{ \left( \frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\beta} \right)[v_{11}] \left( \frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\beta} \right)[v_{12}] \right\}.
\]
Similarly, let $\mathbf{V}_2$ denote the closure of the linear span of $\mathcal{A} - \{\alpha_0\}$ under the metric $\| \cdot \|_\alpha$ (i.e., $\mathbf{V}_2 = \mathbb{R}^{d_\phi} \times \bar{\mathbb{W}}_2$ with $\bar{\mathbb{W}}_2 \equiv \mathcal{H} - \{h_0\}$). Then $(\mathbf{V}_2, \| \cdot \|_\alpha)$ is a Hilbert space with the inner product:

$$\langle v_{21}, v_{22} \rangle_\alpha = E \left\{ \left( \frac{dm(X, \alpha_0)}{d\alpha} [v_{21}] \right)' \Sigma(X)^{-1} \left( \frac{dm(X, \alpha_0)}{d\alpha} [v_{22}] \right) \right\}. $$

The pathwise derivative at $\beta_0$ is defined as

$$\frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{d\beta} [\beta - \beta_0] = \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{d\phi'} (\phi - \phi_0) + \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{d\eta} [\eta - \eta_0] + \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{df_1} [f_1 - f_{Y^*_2|Y_2x_1}] + \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{df_2} [f_2 - f_{Y^*_2|Y_2x_1}].$$

For each component $\phi_j$ of $\phi$, $j = 1, 2, \ldots, d_\phi$, we define $w^*_1 \in \bar{\mathbb{W}}_1$ as

$$w^*_{1j} \equiv (\eta^*_j, f^*_{1j}, f^*_{2j})' = \arg \min_{(\eta, f_1, f_2) \in \bar{\mathbb{W}}_1} E \left\{ \left( \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{d\phi_j} [\eta_j] - \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{df_1} [f_{1j}] - \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{df_2} [f_{2j}] \right)^2 \right\}. $$

Define

$$w^*_1 = (w^*_{11}, w^*_{12}, \ldots, w^*_{1d_\phi}),$$

$$d \ln f_{Y^*|X}(y^* | x; \beta_0) [w^*_{1j}] = \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{df} [w^*_{1j}] = \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{d\eta} [\eta_j] + \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{df_1} [f_{1j}] + \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{df_2} [f_{2j}],$$

$$d \ln f_{Y^*|X}(y^* | x; \beta_0) [w^*_1] = \left( \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{df} [w^*_{11}], \ldots, \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{df} [w^*_{1d_\phi}] \right),$$

and the row vector

$$G_{w^*_1}(Y^*, X, \beta_0) \equiv \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{d\phi'} - \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{df} [w^*_1].$$
We also introduce some notation for the second stage parameters, \( \theta_0 \). As shown before, the pathwise derivative at \( \alpha_0 \) is

\[
\frac{d m(X, \alpha_0)}{d \alpha}[\alpha - \alpha_0] \equiv \frac{d m(X, \alpha_0)}{d \theta'}(\theta - \theta_0) + \frac{d m(X, \alpha_0)}{d h}[h - h_0].
\]

For each component \( \theta_j \) of \( \theta, j = 1, 2, \ldots, d_\theta \), we define \( w^*_j \in \mathbb{W}_2 \) as

\[
w^*_j \equiv \arg \min_{w_j \in \mathbb{W}_2} E \left\{ \left( \frac{d m(X, \alpha_0)}{d \theta_j} - \frac{d m(X, \alpha_0)}{d h}[w^*_j] \right)' \Sigma(X)^{-1} \left( \frac{d m(X, \alpha_0)}{d \theta_j} - \frac{d m(X, \alpha_0)}{d h}[w^*_j] \right) \right\}.
\]

Define

\[
w^*_2 = (w^*_1, w^*_2, \ldots, w^*_d),
\]

\[
\frac{d m(X, \alpha_0)}{d h}[w^*_2] = \left( \frac{d m(X, \alpha_0)}{d h}[w^*_1], \ldots, \frac{d m(X, \alpha_0)}{d h}[w^*_d] \right),
\]

and the row vector

\[
G_{w^*_2}(X, \alpha_0) \equiv \frac{d m(X, \alpha_0)}{d \theta'^2} - \frac{d m(X, \alpha_0)}{d h}[w^*_2].
\]

Define \( s_1(\beta) \equiv \lambda'_1 \phi \) for \( \lambda_1 \in \mathbb{R}^{d_\phi} \) and \( \lambda_1 \neq 0 \), and define \( s_2(\alpha) \equiv \lambda'_2 \theta \) for \( \lambda_2 \in \mathbb{R}^{d_\theta} \) and \( \lambda_2 \neq 0 \). As mentioned in Ai and Chen (2003), \( s_1(\beta) \equiv \lambda'_1 \phi \) is bounded if and only if

\[
E[G_{w^*_1}^2(Y^*, X, \beta_0)'G_{w^*_1}(Y^*, X, \beta_0)]
\]

is finite positive-definite. The Riesz representation theorem then implies that there exists a representor \( v^*_1 \) such that

\[
s_1(\beta) - s_1(\beta_0) \equiv \lambda'_1 (\phi - \phi_0) = \langle v^*_1, \beta - \beta_0 \rangle_{\beta}
\]

for all \( \beta \in \mathcal{B} \) where \( v^*_1 \equiv (v^*_\phi, v^*_\beta) \in \mathbf{V}_1, v^*_\phi = J^{-1}_1 \lambda_1, v^*_\beta = -w^*_1 \times v^*_\phi \) with \( J_1 = E[G_{w^*_1}^2(Y^*, X, \beta_0) \Sigma(X)^{-1}G_{w^*_1}(Y^*, X, \beta_0)] \). Similarly, because of the fact that \( s_2(\alpha) \equiv \lambda'_2 \theta \) is bounded if and only if

\[
E[G_{w^*_2}^2(X, \alpha_0)'\Sigma(X)^{-1}G_{w^*_2}(X, \alpha_0)]
\]

is finite positive-definite, we have

\[
s_2(\alpha) - s_2(\alpha_0) \equiv \lambda'_2 (\theta - \theta_0) = \langle v^*_2, \alpha - \alpha_0 \rangle_{\alpha}
\]
for all \( \alpha \in \mathcal{A} \) where \( v_2^* \equiv (v_\theta^*, v_h^*) \in \nabla_2, v_\theta^* = J_2^{-1} \lambda_2, v_h^* = -w_2^* \times v_\theta^* \) with \( J_2 = E[G_w^2(X, \alpha_0)'] \Sigma(X)^{-1}G_w^2(X, \alpha_0)] \).

We now state the sufficient conditions for the \( \sqrt{n} \)-normality of \( \hat{\phi}_n \) and \( \hat{\theta}_n \).

**Assumption 5.1** (i) \( E[G_w^1(X, \alpha_0)'] \Sigma(X)^{-1}G_w^1(X, \alpha_0)] \) exists, is bounded, and is positive-definite; (ii) \( \theta_0 \in \text{int}(\Theta) \); (iii) \( \Sigma_0(X) \equiv \text{var}[\rho(Z, \alpha_0) \mid X] \) is positive-definite for all \( X \in X \).

**Assumption 5.2** (i) \( E[G_w^1(Y^*, X, \beta_0)']G_w^1(Y^*, X, \beta_0)] \) exists, is bounded, and is positive-definite; (ii) \( \phi_0 \in \text{int}(\Phi) \).

**Assumption 5.3** There is a \( v_{2n}^* = (v_\theta, -\Pi_n w_2^* \times v_\theta) \in \mathcal{A}_n - \{\Pi_n \alpha_0\} \) such that \( \|v_{2n}^*-v_2^*\|_\alpha = O(n^{-1/4}) \).

**Assumption 5.4** There is a \( v_{1n}^* = (v_\theta, -\Pi_n w_1^* \times v_\theta) \in \mathcal{B}_n - \{\Pi_n \beta_0\} \) such that \( \|v_{1n}^*-v_1^*\|_\beta = O(n^{-1/4}) \).

Define \( \mathcal{N}_{01n} \equiv \{\beta \in \mathcal{B}_n : \|\beta - \beta_0\|_{s,\beta} \leq v_{1n}, \|\beta - \beta_0\|_\beta \leq v_{1n} n^{-1/4}\} \) with \( v_{1n} = o(1) \) and define \( \mathcal{N}_{01} \) the same way with \( \mathcal{B}_n \) replaced by \( \mathcal{B} \). Define \( \mathcal{N}_{02n} \equiv \{\alpha \in \mathcal{A}_n : \|\alpha - \alpha_0\|_{s,\alpha} \leq v_{2n}, \|\alpha - \alpha_0\|_\alpha \leq v_{2n} n^{-1/4}\} \) with \( v_{2n} = o(1) \) and define \( \mathcal{N}_{02} \) the same way with \( \mathcal{A}_n \) replaced by \( \mathcal{A} \). For \( \beta \in \mathcal{N}_{01n} \), we denote a local alternative \( \beta^*(\beta, \varepsilon_n) = (1 - \varepsilon_n)\beta + \varepsilon_n(v_1^* + \beta_0) \) with \( \varepsilon_n = o(n^{-1/2}) \). Let \( \Pi_n \beta^*(\beta, \varepsilon_n) \) be the projection of \( \beta^*(\beta, \varepsilon_n) \) onto \( \mathcal{B}_n \). We denote

\[
\frac{d\rho(Z, \alpha)}{d\alpha}[v_2] \equiv \frac{d\rho(Z, \alpha + \tau v_2)}{d\tau}|_{\tau=0} \text{ a.s. } Z,
\]

and

\[
\frac{dm(X, \alpha)}{d\alpha}[v_2] \equiv \frac{dm(X, \alpha + \tau v_2)}{d\alpha}[v_2] \text{ a.s. } X,
\]

for any \( v_2 \in \nabla_2 \). Also for any \( v_1 \in \nabla_1 \), we denote

\[
\frac{d\ln f_{Y^*|X}(y^* \mid x; \beta)}{d\beta}[v_1] \equiv \frac{d\ln f_{Y^*|X}(y^* \mid x; \beta + tv_1)}{dt}|_{t=0} \text{ a.s. } (Y^*, X).
\]
Assumption 5.5 \textit{For all }\alpha \in \mathcal{N}_{02}, \text{ the pathwise first derivative } (d\rho(Z, \alpha(\tau))/d\alpha)[v_2] \text{ exists a.s. } Z \in Z. \text{ Moreover, (i) each element of } (d\rho(Z, \alpha)/d\alpha)[v_{2n}^*] \text{ satisfies an envelope condition and is Hölder continuous in } \alpha \in \mathcal{N}_{02n}; \text{ (ii) each element of } (dm(X, \alpha)/d\alpha)[v_{2n}^*] \text{ is in } \Lambda^v_\gamma(X), \gamma > d_2/2 \text{ for all } \alpha \in \mathcal{N}_{02}.

Assumption 5.6 \textit{Uniformly over }\alpha \in \mathcal{N}_{02n}, \text{ we have}

$$E \left( \left\| \frac{dm(X, \alpha)}{d\alpha}[v_{2n}^*] - \frac{dm(X, \alpha_0)}{d\alpha}[v_{2n}^*] \right\|_E^2 \right) = o(n^{-1/2}).$$

Assumption 5.7 \textit{Uniformly over }\alpha \in \mathcal{N}_{02}, \bar{\alpha} \in \mathcal{N}_{02n}, \text{ we have}

$$E \left( \left\{ \frac{dm(X, \alpha_0)}{d\alpha}[v_2^*] \right\} \Sigma(X)^{-1} \left\{ \frac{dm(X, \alpha)}{d\alpha}[\bar{\alpha} - \alpha_0] - \frac{dm(X, \alpha_0)}{d\alpha}[\bar{\alpha} - \alpha_0] \right\} \right) = o(n^{-1/2}).$$

Assumption 5.8 \textit{For all }\alpha \in \mathcal{N}_{02n}, \text{ the pathwise second derivative } d^2\rho(Z, \alpha + \tau v_{2n}^*)/d\tau^2|_{\tau=0} \text{ exists a.s. } Z \in Z, \text{ and is bounded by a measurable function } c_5(Z) \text{ with } E[c_5(Z)^2] < \infty.

Assumption 5.9 \textit{There exists a measurable function }h_2(Y^*, X) \text{ with } E[h_2(Y^*, X)^2] < \infty \text{ such that for any } \bar{\beta} = (\bar{\psi}, \bar{f}_1, \bar{f}_2) \in \mathcal{N}_{01},

$$\left| \frac{f_{y^*|X}^{[1]}(y^* | x; \bar{\beta}, \bar{\omega})}{f_{Y^*|X}(y^* | x; \bar{\beta})} \right|^2 + \left| \frac{f_{y^*|X}^{[2]}(y^* | x; \bar{\beta}, \bar{\omega})}{f_{Y^*|X}(y^* | x; \bar{\beta})} \right| \leq h_2(Y^*, X),$$

where \( f_{y^*|X}^{[2]}(y^* | x; \bar{\beta}, \bar{\omega}) \) is defined as \( \frac{d^2}{d\alpha^2} f_{Y^*|X}(y^* | x; \bar{\beta} + t\bar{\omega})|_{t=0} \) with each linear term, that is, \( \frac{d}{d\psi} f_{Y^*|X}(y^* | x; \bar{\beta} + t\bar{\omega}), \frac{d}{df_1} f_{Y^*|X}(y^* | x; \bar{\beta} + t\bar{\omega}), \bar{f}_1, \text{ and } \bar{f}_2, \) replaced by its absolute value. (The explicit expression of \( f_{y^*|X}^{[2]}(y^* | x; \bar{\beta}, \bar{\omega}) \) can be found in the proof of Theorem 5.1(i).)

Following Hu and Schennach (2008), we write the following notations for the next assumption:

$$\frac{d \ln f_{y^*|X}(y^* | x; \beta_0)}{d\beta} [p^{k_n}] = \left( \left( \frac{d \ln f_{y^*|X}(y^* | x; \beta_0)}{d\phi} \right)' \right) \left( \frac{d \ln f_{y^*|X}(y^* | x; \beta_0)}{d\eta} \right)', \left( \frac{d \ln f_{y^*|X}(y^* | x; \beta_0)}{df_1} [p^{k_n}] \right)', \left( \frac{d \ln f_{y^*|X}(y^* | x; \beta_0)}{df_2} [p^{k_n}] \right)'.$$
where for \( \tilde{f} = \eta, f_1, \) or \( f_2, \)
\[
\frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{d f} \bigg|_{p^{k_n}} = \left( \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{d f} \bigg|_{p^{k_n}}, \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{d f} \bigg|_{p^{k_n}}, \ldots, \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{d f} \bigg|_{p^{k_n}} \right),
\]
\[
\frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{d \phi} = \left( \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{d \phi_1}, \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{d \phi_2}, \ldots, \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{d \phi_{d_\phi}} \right),
\]
and
\[
\Omega_{k_n} = E \left\{ \left( \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{d \beta} \bigg|_{p^{k_n}} \right) \left( \frac{d \ln f_{Y^*|X}(y^* | x; \beta_0)}{d \beta} \bigg|_{p^{k_n}} \right) \right\}. \]

**Assumption 5.10** The smallest eigenvalue of the matrix \( \Omega_{k_n} \) is bounded away from zero, and
\( \|p_j^{k_n}\|_{\infty, \omega} < \infty \) for \( j = 1, 2, \ldots, k_n \) uniformly in \( k_n. \)

**Assumption 5.11** For all \( \beta \in N_{01n}, \) there exists a measurable function \( h_4(Y^*, X) \) with
\( E|h_4(Y^*, X)| < \infty \) such that
\[
\left| \frac{d^4}{dt^4} \ln f_{Y^*|X}(y^* | x; \beta + t(\beta - \beta_0)) \right|_{t=0} \leq h_4(Y^*, X) \|\beta - \beta_0\|^{4}_{s, \beta}.
\]

**Theorem 5.1** (i) Under Assumptions 3.1-3.8, 4.1, 4.5, 4.7, 4.11, 5.2, 5.4, 5.9-5.11, \( \sqrt{n}(\hat{\phi}_n - \phi_0) \xrightarrow{d} N(0, J_1^{-1}), \)
where \( J_1 = E[G_{w_1}(Y^*, X, \beta_0)']G_{w_1}(Y^*, X, \beta_0)]. \)

(ii) Under Assumptions 3.1-3.8, 4.1-4.11, 5.1-5.11, \( \sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, J_2^{-1}), \) where
\[
J_2 = E[G_{w_2}(X, \alpha_0)'] \Sigma(X)^{-1}G_{w_2}(X, \alpha_0)]
\times (E[G_{w_2}(X, \alpha_0)'] \Sigma(X)^{-1} \Sigma_0(X) \Sigma(X)^{-1}G_{w_2}(X, \alpha_0)])^{-1}
\times E[G_{w_2}(X, \alpha_0)'] \Sigma(X)^{-1}G_{w_2}(X, \alpha_0)].
\]
5.2 Consistent Covariance Estimator

We now establish a consistent estimator $\tilde{J}_{02}$ of the covariance matrix $J_{02}$, which is needed to perform any statistical inference using the semiparametrically efficient estimator $\tilde{\theta}_n$.

Let $\hat{G}_w(X, \tilde{\alpha})$ be a consistent estimator of $G_w(X, \alpha_0)$ as follows:

$$\hat{G}_w(X, \tilde{\alpha}) = \frac{d\hat{m}(X, \tilde{\alpha})}{d\theta} - \frac{d\hat{m}(X, \tilde{\alpha})}{dh}[w].$$

We estimate $w_{02}$ by $\tilde{w}_{02}$, which is the solution to the minimization problem:

$$\min_{w_{02}} \frac{1}{n} \sum_{i=1}^{n} \hat{G}_w(X_i, \tilde{\alpha})' \hat{\Sigma}_0(X_i)^{-1} \hat{G}_w(X_i, \tilde{\alpha}).$$

If we let $w_{02} = (w_{02,1}, \ldots, w_{02,d})$ and $\tilde{w}_2 = (\tilde{w}_{21}, \ldots, \tilde{w}_{2d})$, then $\hat{G}_{\tilde{w}_2}(X, \tilde{\alpha})$ is a consistent estimator of $G_{w_{02}}(X, \alpha_0)$. Therefore, the estimator of $J_{02}$ is $\tilde{J}_{02} = \frac{1}{n} \sum_{i=1}^{n} \hat{G}_{\tilde{w}_2}(X_i, \tilde{\alpha})' \hat{\Sigma}_0(X_i)^{-1} \hat{G}_{\tilde{w}_2}(X_i, \tilde{\alpha})$.

\textbf{Theorem 5.2} Under the conditions of Theorem 5.1 (ii), $\tilde{J}_{02} = J_{02} + o_p(1)$.

Theorem 5.2 states that the estimator $\tilde{J}_{02}$ of the covariance matrix $J_{02}$ is consistent.

6 Simulation and Empirical Illustration

6.1 A Monte Carlo Study

We assess the finite sample performance of the proposed estimator in this section. The simulation is based on a nonparametric regression

$$Y_1 = h_0(Y_2) + U$$

where $h_0(Y_2) = \exp(Y_2)/(1 + \exp(Y_2))$. We assume that $Y_2$ is generated as $Y_2 = aX_2 + R(U + c) + b\varepsilon$. $X_2$, $\varepsilon$, and $U$ are independent and distributions of those are $X_2 \sim N(1, \sigma^2)$, $\varepsilon \sim N(1, \sigma^2)$,
and $U \sim N(0, \sigma^2)$ with $(a, b, c, R, \sigma) = (0.6, 0.2, 1, 0.2, 0.7)$. The distributions of $X_2$ and $\varepsilon$ are truncated on $[0, 2]$ and the distribution of $U$ is truncated on $[-1, 1]$. Thus the support of $Y_2$ is $[0, 2]$. As in Ai and Chen (2003), we approximate the unknown $h_0(Y_2)$ by a power series of fourth order multiplied by the cumulative distribution function of a standard normal since $h_0(Y_2)$ is bounded between zero and one. So the approximate regression model is

$$Y_1 \approx \pi_0 \Phi(Y_2) + \pi_1 \Phi(Y_2)Y_2 + \pi_2 \Phi(Y_2)Y_2^2 + \pi_3 \Phi(Y_2)Y_2^3 + U$$

where $\Phi(Y_2)$ denotes the standard normal cumulative distribution function.

We also use the general form of generating processes for the measurement error which is similar to those in Hu and Schennach (2008)

$$f_{Y_2^*|Y_2}(y_2^* | y_2) = \frac{1}{\sigma(y_2)} f_{\nu} \left( \frac{y_2^* - y_2}{\sigma(y_2)} \right),$$

where $\sigma(y_2) = 1.5 \exp(-y_2)$ and $f_{\nu}$ is a density function to be specified below for three models: heteroskedastic measurement error with zero mean, nonadditive measurement error with zero mode, and nonadditive measurement error with zero median.

1. **Heteroskedastic Measurement Error with Zero Mean**: a measurement error is

$$Y_2^* = Y_2 + \sigma(y_2) \nu$$

with $Y_2 \perp \nu$. The error structure in the simulation is $F_{\nu}(\nu) = \Phi(\nu)$.

2. **Nonadditive Measurement Error with Zero Mode**: let

$$f_{Y_2^*|Y_2}(y_2^* | y_2) = \frac{g(y_2^*, y_2)}{\int_{-\infty}^{\infty} g(y_2^*, y_2) dy_2^*},$$

$$g(y_2^*, y_2) = \exp \left\{ h(y_2) \left[ \left( \frac{y_2^* - y_2}{\sigma(y_2)} \right) - \exp \left( \frac{y_2^* - y_2}{\sigma(y_2)} \right) \right] \right\}$$

with $h(y_2) = \exp(-0.1y_2)$. Then $f_{Y_2^*|Y_2}(y_2^* | y_2)$ has the unique mode at $y_2$ for any $h(y_2) > 0$.

3. **Nonadditive Measurement Error with Zero Median**: let the corresponding cumulative dis-
tribution function be

\[ F_{Y_2^* | Y_2}(y_2^* | y_2) = \frac{1}{\pi} \arctan \left\{ h(y_2) \left[ \frac{1}{2} + \frac{1}{2} \exp \left( \frac{y_2^* - y_2}{\sigma(y_2)} \right) - \left( \frac{y_2^* - y_2}{\sigma(y_2)} \right) \right] \right\} + \frac{1}{2} \]

with \( h(y_2) = \exp(-0.1y_2) \). Then \( F_{Y_2^* | Y_2}(y_2^* | y_2) = \frac{1}{2} \) for any \( h(y_2) > 0 \).

We consider three estimators: (i) the (inconsistent) SMD estimator from Ai and Chen (2003) which is obtained using error-laden data, (ii) the (infeasible) SMD estimator from Ai and Chen (2003) which is obtained using error-free data, and (iii) the proposed two-stage SML-SMD estimator. We construct sieves for functions of two variables using tensor product bases of univariate trigonometric series in our estimator. In both SMD estimators, we use a tensor product polynomial sieve to approximate the conditional mean function which is the set of instruments: \( \{1, X_2, X_2^2, \ldots, X_2^{k_n}\} \) for \( k_n \geq 3 \). The sample size is 1,000 and the procedures are repeated 100 times to obtain the root integrated mean squared error (RIMSE) according to the following discrete expression:

\[
((200)^{-1} \sum_{j=0}^{199} \text{mean}\{[h_0(0 + 0.01j) - \hat{h}(0 + 0.01j)]^2\})^{1/2},
\]

where \( \text{mean}\{\cdot\} \) denotes the average over all 100 estimators \( \hat{h} \) for each procedure.

Table 1 reports estimation results. RIMSE from our proposed estimator is smaller than that from the SMD estimator obtained using error-laden data for all cases of identification conditions for measurement error.

### 7 Instrumental Variables Estimation of Engel Curves

We apply the proposed estimator to the estimation of Engel curves (or consumer demand models) using the British Family Expenditure Survey (FES) data. Findings confirm that correcting for both endogeneity and measurement error is necessary to identify the economically meaningful structural Engel curves.
7.1 Overview

Demand models play an important role in the welfare analysis. One of the reason is that the evaluation of indirect tax policy reform needs the accurate specification of demand models which is consistent with consumer theory. Because of that, the study of the Engel curves, the relationship between expenditure (or income) and budget shares, has been an area of interest among econometricians since the early studies of Engel (1895), Working (1943), Leser (1963). Many of previous studies exploits the best model specification for the Engel curves and ‘Leser-Working’ specification of Engel curve in which budget shares are a linear function of the log of income or expenditure, has been the most popular one. However, economic theory provides almost no general guidance in specification of Engel curves and recent empirical studies show that linear specification of the Engel curves is far from an accurate feature of consumer behavior. Some empirical analysis of consumer behavior suggest that nonlinear parametric or semiparametric and nonparametric models are more favorable in the specification of the Engel curves. Along with the model specification, there have been two directions in the analysis of Engel curves: endogeneity and measurement errors.

A group of studies estimate Engel curves based on that budget shares and expenditure are endogenous to the consumer and are determined simultaneously, as pointed out by Summers (1959). Using a nonparametric method and correcting for the endogeneity of the log-total expenditure, Banks, Blundell, and Lewbel (1997) suggest that Engel curves require quadratic terms in the log-total expenditure. They also find that models failing to account for nonlinearity of the Engel curves could distort the patterns of welfare losses associated with a tax increase. Blundell, Duncan, and Pendakur (1998) allow for endogeneity of the log-total expenditure by adopting a parametric additive control function approach to the partially linear regression context and find that taking account of endogeneity has an important impact on the shape of the Engel curve relationship, while Blundell, Browning, and Crawford (2003) use a nonparametric control function technique to adjust for endogeneity. Base on a nonparametric method, Lyssiotou, Pashardes and Stengos (1999) find that controlling for endogeneity tends to be more supportive of the rank 3 hypothesis. Blundell, Chen, and Kristensen (2007) (BCK) studies a shape-invariant Engel curve with endogenous log-total expenditure by applying the sieve minimum distance estimation of conditional moment restrictions and find the importance of correcting for endogeneity. Chen and Pouzo (2008a, b) studies non-
parametric or semiparametric estimation of conditional moment models with possibly nonsmooth residuals, respectively and applied their methods to estimate quantile Engel curves with endogenous log-total expenditure.

Another issue on the estimation of the Engel curves is measurement error in total expenditure. Measurement error would be because of survey errors or a form of errors which come from the discrepancy between purchases and consumption due to storage or waste. In a linear parametric model, Liviatan (1961) applies the method of instrumental variables to the Engel curves, with income serving as the instrumental variable. Aasness, Biorn, and Skjerpen (1993) model measurement error in total expenditure to estimate Engel curves with panel data. Hausman, Newey, and Powell (1995) propose consistent estimators for nonlinear regression framework in the presence of measurement error. In their application to the Engel curves, they find that measurement error in income should be accounted for and ‘Lesser-Working’ specification should be generalized to higher-order terms in log income. Lewbel (1996) develops a consistent estimator of nonlinear Engel curves to correct for measurement errors in total expenditures on the left and right hand side since an observed budget share has expenditure in its denominator. Newey (2001) studies the estimation of nonlinear errors-in-variables models using simulated moments and a flexible disturbance distribution, and applies the models to Engel curves with expenditures measurement errors on the left and right hand side. Hasegawa and Kozumi (2001) correct for expenditure measurement errors on both the left and right sides in the ‘Lesser-Working’ specification. They propose the Bayesian estimation procedure in both models without an instrument variable and with an instrument variable. Schennach (2004) proposes a general solution to measurement error in general nonlinear models when one repeated observation is available for each mismeasured variable and applies it to the estimation of Engel curves. She finds that the impact of measurement error in total expenditure can not be neglected.

Even though there are plenty of evidences that total expenditure is endogenous as well as mismeasured, there has been no study which corrects for both endogeneity and measurement error in nonlinear parametric, nonparametric, or semiparametric models. As discussed by Amemiya (1985) and Hsiao (1989), it is because nonlinear regression models with measurement error are difficult to estimate with standard linear instrumental variables approach, due to the lack of additive separability between true regressor and measurement error. The present study employs the proposed
method in order to fill this gap. So our target is to control for both endogeneity and measurement error in the nonparametric shapes of the Engel curves.

The nonparametric specification of Engel curves we consider is

\[ E[Y_{1i,l} - h_l(Y_{2i}) \mid X_i] = 0, \quad l = 1, \cdots, 7, \] (12)

where \( Y_{1i,l} \) is the budget share of household \( i \) on good \( l \) (e.g., 1: food-out, 2: food-in, 3: alcohol, 4: fares, 5: fuel, 6: leisure goods, and 7: travel). \( Y_{2i} \) is the log-total expenditure of household \( i \) that is endogenous and unobservable, and \( X_i \) is gross earnings of the head of household, which is the instrumental variable. We consider the no kids sample that consists of 628 observations. BCK have used the same data set as well as a subset of married couples with one or two children in their study of a shape-invariant system of IV Engel curves. Table 3.4 summarizes descriptive statistics for the main variables in the data set. We see that budget shares on food-in, leisure, and travel are large, while food-out, alcohol, fares, and fuel are relatively small. Leisure goods have a large standard deviation. The mean and standard deviation for log nondurable expenditure are similar to those for log gross earnings. As shown in BCK, log-total expenditure and log earnings have a strong positive correlation, which is 0.5111. We also assume that log earnings are independent of the residual, \((Y_{1i} - h_l(Y_{2i}))\). So the log gross earnings would be a proper instrumental variable to analyze the conditional moment restriction model.

BCK assume that the log of total expenditure on nondurables and services is endogenous but measurement error-free. However, their approach is infeasible if the true log-total expenditure suffers from measurement errors so that only a mismeasured version is observed.\(^2\) As reviewed above, indeed, many empirical papers on the estimation of Engel curves show that measurement errors on the log-total expenditure is considerable. As a result, failure of controlling for measurement errors makes it difficult to estimate the economically meaningful Engel curves.

\(^2\)We thank Richard Blundell for providing the UK Family Expenditure Survey data.

\(^3\)We assume that there is no measurement error on the left-hand variable in Eqn. (12) to ease the argument. It could be possible because both expenditure on good \( l \) and total expenditure might have measurement errors but the budget share could be correctly reported one if proportion of error-laden expenditure on good \( l \) to error-laden total expenditure is the same as true budget share. For instance, assume there are multiplicative measurement errors on expenditure on good \( l \) and total expenditure such that \( Y^*_{0i,l} = Y_{0i,l}e_{0i,l} \) and \( Y^*_{2i} = Y_{2i}e_{2i} \), where \( Y^*_{0i,l} \), \( Y^*_{0i,l} \), and \( e_{0i,l} \) are measurement error-laden expenditure of household \( i \) on good \( l \), true expenditure of household \( i \) on good \( l \), and its measurement error, respectively, and where \( Y^*_{2i}, Y_{2i}, \) and \( e_{2i} \) are measurement error-laden total expenditure of household \( i \), true total expenditure of household \( i \), and its measurement error, respectively. If \( e_{0i,l} = e_{2i} \), we can get \( \frac{Y^*_{0i,l}}{Y^*_{2i}} = \frac{Y_{0i,l}e_{0i,l}}{Y_{2i}e_{2i}} = \frac{Y_{0i,l}}{Y_{2i}} \).
7.2 Two-step SML-SMD Procedure

In order to use the two-step sieve maximum likelihood and sieve minimum distance (SML-SMD) estimator, we specify the conditional mean function as follows:

\[
m(x, h) \equiv \int_y (y_1 - h(y_2))dF_{Y|X}(y \mid x; \phi_0, \eta_0) = \int_{\mathcal{Y}_2} \left[ \int_{\mathcal{Y}_1} (y_1 - h(y_2))dF_{Y_1|Y_2,X}(y_1 \mid y_2, x) \right] dF_{Y_2|X}(y_2 \mid x; \phi_0, \eta_0)
\]

(13)

\[
= \int_{\mathcal{Y}_2} \left[ \int_{\mathcal{Y}_1} (y_1 - h(y_2))f_{Y_1|Y_2}(y_1 \mid y_2, x)dy_1 \right] f_{Y_2|X}(y_2 \mid x; \phi_0, \eta_0)dy_2
\]

where \( \mathcal{Y}_1 \) and \( \mathcal{Y}_2 \) are the support of the distribution of \( Y_1 \) and \( Y_2 \), respectively. In the empirical application, \( \mathcal{Y}_1 = [0, 0.350] \) and \( \mathcal{Y}_2 = [3.609, 6.947] \). Since partially parameterizing distributions eases nonparametric estimation of densities, we allow \( f_{Y_2|X}(y_2 \mid x) \) to be parameterized. In fact, the conditional distribution of log-total expenditure given log gross earnings is close to normal (see BCK), we specify \( f_{Y_2|X}(y_2 \mid x; \phi_0, \eta_0) \) as normal distribution. This is one of useful properties of the two-step SML-SMD estimator, which the sieve minimum distance procedure can not utilize because of its nature of the estimation.

In the first step, we estimate the population conditional mean function \( m(x, h) \) semiparametrically by \( \hat{m}(x, h) \). To do this, we use a SML estimation to estimate \( f_{Y_1|Y_2}(y_1 \mid y_2) \) and \( f_{Y_2|X}(y_2 \mid x; \phi_0, \eta_0) \).

\[
\beta_0 = \left( \psi_0, f_{Y_1|Y_2}, f_{Y_2|Y_2} \right) = \arg \max_{\beta = (\psi, f_0, f_1)} E \left( \ln \int_{\mathcal{Y}_2} f_{0}(y_1 \mid y_2)f_{1}(y_2^* \mid y_2)f_{Y_2|X}(y_2 \mid x; \psi)dy_2 \right),
\]

(14)

where \( \mathcal{B} = \Psi \times F_0 \times F_1 \) with \( \Psi = \Phi \times \mathcal{M} \) and \( \psi_0 = (\phi_0, \eta_0) \).

We also approximate the unknown function \( h \in \mathcal{H} \) by \( h_n \in \mathcal{H}_n \equiv \mathcal{H}_n^1 \times \cdots \times \mathcal{H}_n^q \) where \( \mathcal{H}_n \) is some finite-dimensional approximation space that becomes dense in \( \mathcal{H} \) as sample size \( n \to \infty \). In the second step, the SMD estimator of unknown sieve coefficients of \( h_0 \) is obtained by applying
the SMD procedure

$$\hat{h}_n = \arg \min_{h_n \in H_n} \frac{1}{n} \sum_{i=1}^{n} \hat{m}(X_i, h_n)' \hat{m}(X_i, h_n),$$

(15)

where $\hat{m}(X, h)$ is the plug-in SML estimator of $m(X, \alpha)$ for any fixed $h_n = (h_{1,n}, \ldots, h_{q,n})$:

$$\hat{m}(x, h_n) \equiv \int_{\mathcal{Y}_2} \left[ \int_{\mathcal{Y}_1} (y_1 - h_n(y_2)) \hat{f}_{Y_1|Y_2}(y_1 | y_2) dy_1 \right] \hat{f}_{Y_2|X}(y_2 | x; \hat{\phi}_n, \hat{\eta}_n) dy_2.$$

(16)

For the purpose of comparison, we also estimate the Engel curves using SMD estimator from BCK, which does not control for measurement errors of log-total expenditure. Both SMD and SML-SMD estimators are constructed without smoothness constraints for simplicity. We use a power series of fourth order multiplied by the cumulative distribution function of a standard normal to approximate $h_{0}(Y_2)$ for both estimators. In the SMD estimator, a set of instruments, $\{1, X_2, X_2^2, \ldots, X_2^{k_n}\}$ for $k_n \geq 3$ is used to approximate the conditional mean function.

### 7.3 Estimation Results

Figures 1～2 show estimated Engel curves for four of the goods in the system. We plot curves over a set of log-total expenditures ranging from 4.5 to 6.5. Engel curves from our SML-SMD estimator which controls for both endogeneity and measurement errors in the log-total expenditure are plotted by real curves, while those from SMD estimator which only control for endogeneity in the log-total expenditure are plotted by dashed curves.

We note several interesting features. For households with low log-total expenditure, shares of food-in from our SML-SMD estimator are bigger than those from SMD estimator. Food-out from SMD estimator is a reverse U-shape and values are similar over different level of log-total expenditure. But Food-out from SML-SMD estimator dramatically decreases as log-total expenditure increases. As a result, for households with low log-total expenditure, the estimated shares of food from our estimator, which is sum of food-in and food-out, are much bigger than those from SMD estimator, even though food shares of households with high log-total expenditure from both esti-
mators look similar. The Engel curve for fuel from SMD estimator shows a reverse S-shape and is close to that from SML-SMD estimator. However, the estimated Engel curves for leisure from both estimator show huge gaps. For example, the estimated shares of leisure for households with high log-total from SML-SMD estimator are around 0.7 bigger than those from SMD estimator. Thus measurement errors in log-total expenditure can make it difficult to estimate the Engel curves and controlling for the measurement errors are necessary to get correct estimates of the Engel curves.

Our empirical results can be extended in several directions. First, as in BCK, we could consider shape invariant Engel curves and compare the shapes of the estimated Engel curves to theirs. Then corresponding semiparametric model is

\[
E[Y_{1i,l} - h_l(Y_{2i} - \phi(X_{1i}'\theta_1)) - X_{1i}'\theta_{2,l} | X_i] = 0, \quad l = 1, \ldots, 7,
\]

where \(\phi(X_{1i}'\theta_1)\) is a known function up to a finite set of unknown parameters \(\theta_1\) and can be interpreted as the log of a general equivalence scale for household \(i\). \(X_{1i}\) is a vector of demographic variables that represent different household types and \(\theta_2\) is the vector of corresponding equivalence scales (see, e.g., Pendakur (1998) and Blundell, Browning, and Crawford (2003) ) and \(X_i = (X_{1i}, X_{2i})\). Second, we could consider smoothness constraints in the second-step of our estimation procedure and compare the shapes of the estimated Engel curves to theirs. The penalized SMD estimation is

\[
\hat{h}_n = \arg \min_{h_n \in \mathcal{H}_n} \frac{1}{n} \sum_{i=1}^{n} \tilde{m}(X_i, h_n)' \tilde{m}(X_i, h_n) + \lambda_n \hat{P}_n(h_n),
\]

where \(\hat{P}_n(h_n)\) is the penalization function on the smoothness and \(\lambda_n\) is the smoothing parameter. Third, our empirical analysis needs to carry out the robustness check of the estimated Engel curves with respect to the selection of sieve basis functions and the smoothing parameters in the smoothness constraints. Two approximations are required to proceed our SML-SMD estimation: one to approximate \(h\) and the other to approximate unknown densities. So it would be useful to examine how the choice of sieve basis and the smoothing parameter affect the shapes of the estimated Engel curves. Fourth, a semiparametric Hausman-test on measurement errors could be developed. Let \(\hat{\theta}_{SML-SMD}\) and \(\hat{\theta}_{SMD}\) denote the semiparametric estimate of \(\theta\) under \(H_0 : Y_2\) me-
measurement error-free and $H_1 : Y_2$ measurement error-laden, respectively and let $\hat{V}_{SML-SMD}$ and $\hat{V}_{SMD}$ denote the estimates of their respective variances. It then follows that $n(\hat{\theta}'_{SML-SMD} - \hat{\theta}'_{SMD}) \hat{V}^{-1}(\hat{\theta}_{SML-SMD} - \hat{\theta}_{SMD}) \overset{asy.}{\sim} \chi^2_{q+1}$ under the null, where $\hat{V} \overset{p}{\to} V$ with $V$ the asymptotic covariance matrix of $n(\hat{\theta}_{SML-SMD} - \hat{\theta}_{SMD})$.

8 Summary and Concluding Remarks

We consider semiparametric estimation of models defined by conditional moment restrictions, which contain finite dimensional unknown parameters and infinite dimensional unknown functions. We extend these models to include the case where the unknown functions depend on endogenous variables which are contaminated by nonclassical measurement errors. A two-stage estimation procedure is proposed to recover the true conditional density of endogenous variables given conditioning variables masked by the nonclassical measurement errors, and to rectify the difficulty associated with endogeneity of the unknown functions. Specifically, we estimate conditional density of endogenous variables given conditioning variables in the first stage using sieve maximum likelihood estimation, and then estimate parameters of interest in the second stage using sieve minimum distance estimation. We show that the proposed estimator of the infinite dimensional unknown functions is consistent with a rate faster than $n^{-1/4}$ under a certain metric, and the proposed estimator of the finite dimensional unknown parameters obtains root-n asymptotic normality. Monte Carlo evidence and an application to semiparametric estimation of the shape-invariant IV Engel curves illustrate the usefulness of our method.
A Mathematical Appendix

Proof of Theorem 3.1 Since $Y^* \equiv (Y_1', Y_2')'$ and $X \equiv (X_1', X_2')'$, Eqn. (9) follows by the fact that

$$f_{Y^*|X}(y | x) = f_{Y_1|X_1}(y_1, y_2 | x_2, x_1)$$

$$= \int_{y_2} f_{Y_1|Y_2^*X_2X_1}(y_1 | y_2, y_2^*, x_2, x_1) \cdot f_{Y_2|Y_2^*X_2X_1}(y_2 | y_2^*, x_2, x_1) dy_2$$

where the fourth equality and the sixth equality are obtained by Assumption 3.2 (i) and (ii), respectively. The equation above relates the joint densities of the observable variables to those of unobservable variables. We need to show the solution to the equation is unique. By the definition 3.1 and the Eqn. (9), we get an operator equivalence relationship: for an arbitrary $g \in G(X_2)$

$$\left[ L_{y_1|Y_2^*X_2X_1} g \right](y_2)$$

$$= \int_{X_2} f_{Y_1|Y_2^*X_2X_1}(y_1, y_2^* | x_2, x_1) \cdot g(x_2) dx_2$$

$$= \int_{X_2} \int_{Y_2} f_{Y_2|Y_2^*X_2X_1}(y_2^* | y_2, x_2) \cdot f_{Y_1|Y_2^*X_2X_1}(y_1 | y_2, x_2, x_1) \cdot f_{Y_2|Y_2^*X_2X_1}(y_2 | x_2, x_1) dy_2 g(x_2) dx_2$$

$$= \int_{Y_2} f_{Y_2|Y_2^*X_2X_1}(y_2^* | y_2, x_2) \cdot f_{Y_1|Y_2^*X_2X_1}(y_1 | y_2, x_2, x_1) \cdot \int_{X_2} f_{Y_2|X_2X_1}(y_2 | x_2, x_1) g(x_2) dx_2 dy_2$$

$$= \int_{Y_2} f_{Y_2|Y_2^*X_2X_1}(y_2^* | y_2, x_2) \cdot f_{Y_1|Y_2^*X_2X_1}(y_1 | y_2, x_2, x_1) \cdot \left[ L_{Y_2|X_2X_1} g \right](y_2) dy_2$$

$$= \int_{Y_2} f_{Y_2|Y_2^*X_2X_1}(y_2^* | y_2, x_2) \cdot \left[ \Delta_{y_1|Y_2X_1} L_{Y_2|X_2X_1} g \right](y_2) dy_2$$

$$= \left[ L_{Y_2^*|Y_2X_2} \Delta_{y_1|Y_2X_1} L_{Y_2|X_2X_1} g \right](y_2^*),$$
where the third equality is obtained by an interchange of the order of integration. Thus Eqn. (9) defines the operator equivalence over the domain \( g \in \mathcal{G} (\mathcal{X}_2) \):

\[
L_{y_1 Y_2^*} | X_2 x_1 = L_{y_2^*} | Y_2 x_1 \triangle y_1 | Y_2 x_1 L_{Y_2} | X_2 x_1 .
\]  

(19)

Next, we note that integration of Eqn. (13) over all \( y_1 \in \mathcal{Y}_1 \) yields

\[
L_{Y_2} | X_2 x_1 = L_{Y_2^*} | Y_2 x_1 L_{Y_2} | X_2 x_1 ,
\]

since integration of \( \triangle y_1 | Y_2 x_1 \) becomes the identity operator. Since \( L_{Y_2^*} | Y_2 x_1 \) is one-to-one from Assumption 3.3, isolating \( L_{Y_2} | X_2 x_1 \) yields

\[
L_{Y_2} | X_2 x_1 = L_{Y_2^*}^{-1} | Y_2 x_1 L_{Y_2^*} | X_2 x_1 .
\]

Substitution of the expression into Eqn. (13) yields

\[
L_{y_1 Y_2^*} | X_2 x_1 = L_{y_2^*} | Y_2 x_1 \triangle y_1 | Y_2 x_1 L_{Y_2}^{-1} | Y_2^* | Y_2 x_1 L_{Y_2} | X_2 x_1 .
\]

Since \( L_{Y_2^*} | X_2 x_1 \) is one-to-one from Assumption 3.3, by rearranging, we get the operator equivalence defined over a dense subset of \( \mathcal{G} (\mathcal{Y}_2^*) \)

\[
L_{y_1 Y_2^*} | X_2 x_1 L_{Y_2}^{-1} | Y_2^* | X_2 x_1 = L_{y_2^*} | Y_2 x_1 \triangle y_1 | Y_2 x_1 L_{Y_2}^{-1} | Y_2^* | Y_2 x_1 .
\]

Thus the known operator \( L_{y_1 Y_2^*} | X_2 x_1 L_{Y_2}^{-1} | Y_2^* | X_2 x_1 \) defined in terms of densities of the observable variables \((Y^*, X)\) admits a spectral decomposition (an eigenvalue-eigenfunction decomposition). The eigenvalues of the known operator (the diagonal elements of the \( \triangle y_1 | Y_2 x_1 \) operator, i.e., \( \{ f_{Y_1} | Y_2 x_1 (y_1 \mid y_2, x_1) \} \) for a given \( (y_1, x_1) \) and for all \( Y_2 \)) and the eigenfunctions of the known operator (the kernel of the integral operator \( L_{Y_2^*} | Y_2 x_1 \), i.e., \( \{ f_{Y_2^*} | Y_2 x_1 (\cdot \mid y_2, x_1) \} \) for a given \( x_1 \) and for all \( Y_2 \)) provide the unobserved densities of interest. For the uniqueness of the spectral decomposition, we use similar arguments in Theorem 1 of Hu and Schennach (2008).  

\footnote{To ensure uniqueness of the spectral decomposition, they show four techniques: First, Theorem XV.4.5 in Dunford and Schwartz (1971) guarantees uniqueness up to some normalizations. Second, the a priori arbitrary scale of the eigenfunctions is fixed by the requirement that densities must integrate to 1. Third, Assumption 3.4 and the fact that the eigenfunctions (which do not depend on \( Y_1 \), unlike the eigenvalues) must be consistent across different values of the dependent variable \( Y_1 \) are employed to avoid any ambiguity in the definition of the eigenfunctions when there is an eigenvalue degeneracy that involves two eigenfunctions \( f_{Y_2^*} | Y_2 x_1 (\cdot \mid y_2^*, x_1) \) and \( f_{Y_2^*} | Y_2 x_1 (\cdot \mid y_2^*, x_1) \) for some}. 

31
Let

\[
\hat{Q}_n(\alpha) = \frac{1}{n} \sum_{i=1}^{n} \hat{m}(X_i, \alpha)' \hat{\Sigma}(X_i)^{-1} \hat{m}(X_i, \alpha),
\]

\[
Q_n(\alpha) = \frac{1}{n} \sum_{i=1}^{n} m(X_i, \alpha)' \Sigma(X_i)^{-1} m(X_i, \alpha),
\]

\[
Q(\alpha) = E \left[ m(X, \alpha)' \Sigma(X)^{-1} m(X, \alpha) \right].
\]

We use the following result to prove Theorem 4.1.

**Lemma A.1** Suppose that \( A \) and \( B \) are compact subsets of a space with norm \( \| \alpha \|_{s, \alpha} \) and a space with norm \( \| \beta \|_{s, \beta} \), respectively, and \( Z_t \) \( (t = 1, 2, \cdots) \) are i.i.d. Also suppose that (i) \( \text{Var}[\rho(Z, \alpha) \mid X] \) is bounded for each \( \alpha \in A; \) (ii) \( \| \beta - \beta_0 \|_{s, \beta} = o(1) \); (iii) there is \( b(Z) \) and \( \nu > 0 \) with \( |\rho(Z, \tilde{\alpha}) - \rho(Z, \alpha)| \leq b(Z)\|\tilde{\alpha} - \alpha\|_{s, \alpha} \) and \( E[\hat{g}(x_1)^2] < \infty \) where \( \hat{g} = (\hat{g}(x_1), \cdots, \hat{g}(x_n))' = (\int b(Z) \hat{f}_Y \mid X(y \mid x_1; \hat{\psi}) dy, \cdots, \int b(Z) \hat{f}_Y \mid X(y \mid x_n; \hat{\psi}) dy)' \). Then \( \sup_{\alpha \in A} |\hat{Q}_n(\alpha) - Q(\alpha)| = o_p(1) \) and \( Q(\alpha) \) is continuous.

**Proof of Lemma A.1** The proof will proceed by verifying the hypotheses of Lemma A.2 of Newey and Powell (2003). Their compactness of a parameter space is assumed directly in our hypothesis (i). To show that hypothesis (ii) holds (pointwise convergence in \( \alpha \)), let \( \hat{g}(\alpha) = (\hat{m}(X_1, \alpha), \cdots, \hat{m}(X_n, \alpha))' \), and \( g(\alpha) = (m(X_1, \alpha), \cdots, m(X_n, \alpha))' \). We use the notation \( \lesssim \) for “smaller than up to a generic constant.” Note that for some subsequence \( \{n_j\} \) a.s.,

\[
\left| \hat{Q}_n(\alpha) - Q_n(\alpha) \right| \lesssim \left\| \hat{g}(\alpha) \right\|^2_E - \left\| g(\alpha) \right\|^2_E / n
\]

\[
\leq \left( \left\| \hat{g}(\alpha) - g(\alpha) \right\|^2_E + 2\|g(\alpha)\|_E \cdot \|\hat{g}(\alpha) - g(\alpha)\|_E \right) / n.
\]

Strictly speaking, the first inequality above holds almost surely for some subsequence \( \{n_{j_k}\} \) of an arbitrary subsequence \( \{n_j\} \) of \( \{n\} \) as a consequence of Assumption 4.3, ensuring that \( \hat{\Sigma}(X) = O_p(1) \) uniformly over \( X \in X' \). For clarity and convenience, we will continue to use the notation above without explicit reference to sub-subsequences or probability zero concepts.

Also note that \( \|g(\alpha)\|^2_E / n = O_p(1) \) by the Markov inequality from \( \text{Var}(\rho(Z, \alpha) \mid X) < \infty \). Thus, it value of \( Y_1 \). Fourth, Assumption 3.5 is used to uniquely determine the ordering and indexing of the eigenvalues and eigenfunctions.
suffices to show \( \| \hat{g}(\alpha) - g(\alpha) \|_E^2 / n = o_p(1) \) to show \( \hat{Q}_n(\alpha) - Q_n(\alpha) = o_p(1) \).

\[
\begin{align*}
E \left[ \| \hat{g}(\alpha) - g(\alpha) \|_E^2 / n \right] & = E \left[ (\hat{g}(\alpha) - g(\alpha))' (\hat{g}(\alpha) - g(\alpha)) / n \right] \\
& = E \left[ \frac{1}{n} \sum_{i=1}^{n} \left( \int \rho(Z, \alpha) \hat{f}_{Y|X} (y \mid x_i; \hat{\psi}) dy - \int \rho(Z, \alpha) f_{Y|X} (y \mid x_i; \psi_0) dy \right)' \\
& \quad \times \left( \int \rho(Z, \alpha) \hat{f}_{Y|X} (y \mid x_i; \hat{\psi}) dy - \int \rho(Z, \alpha) f_{Y|X} (y \mid x_i; \psi_0) dy \right) \right] \\
& = E \left[ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d_0} \left( \int \rho_j(Z, \alpha) \hat{f}_{Y|X} (y \mid x_i; \hat{\psi}) dy - \int \rho_j(Z, \alpha) f_{Y|X} (y \mid x_i; \psi_0) dy \right)^2 \right] \\
& = E \left[ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d_0} \left( \int \rho_j(Z, \alpha) \left( \hat{f}_{Y|X} (y \mid x_i; \hat{\psi}) - f_{Y|X} (y \mid x_i; \psi_0) \right) dy \right)^2 \right] \\
& = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d_0} \left[ \left( \int \rho_j(Z, \alpha) \left( \hat{f}_{Y|X} (y \mid x_i; \hat{\psi}) - f_{Y|X} (y \mid x_i; \psi_0) \right) dy \right)^2 \right] \\
& \leq \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d_0} E \left[ \left( \int \int |\rho_j(Z, \alpha)| \right. \right. \\
& \quad \times \left\{ |\hat{f}_{Y|X} (y_1 \mid y_2, x_{1i}; \hat{\psi}) - f_{Y|X} (y_1 \mid y_2, x_{1i}; \psi_0) | \right\} d\psi dy_1 dy_2 \right] \\
& \leq \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d_0} E \left[ \left( \int \int |\rho_j(Z, \alpha)| \right. \right. \\
& \quad \times \left\{ \left| \hat{f}_{Y|X} (y_1 \mid y_2, x_{1i}; \hat{\psi}) - f_{Y|X} (y_1 \mid y_2, x_{1i}; \psi_0) \right| \right\} d\psi dy_1 dy_2 \right] \\
& = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d_0} \left[ \left( \int \int |\rho_j(Z, \alpha)| \left| \hat{f}_{Y|X} (y_1 \mid y_2, x_{1i}; \hat{\psi}) - f_{Y|X} (y_1 \mid y_2, x_{1i}; \psi_0) \right| d\psi dy_1 dy_2 \right] \right] \\
& = o(1),
\end{align*}
\]
since $\|\hat{\beta} - \beta_0\|_{s, \beta} = o_p(1)$. Therefore, we get $\|\hat{g}(\alpha) - g(\alpha)\|_E^2 / n = o_p(1)$ by the Markov inequality. Since $Q_n(\alpha) = Q(\alpha) + o_p(1)$ by the weak law of large numbers, the triangle inequality gives hypothesis (iii) of Newey and Powell (2003). To show hypothesis (iii), let $\tilde{q} = (\tilde{q}(x_1), \ldots, \tilde{q}(x_n))$, and $\tilde{B}_n = \left[ \|\tilde{q}\|_E^2 + 2\|\tilde{q}\|_E \cdot \|\tilde{g}(\alpha_0)\|_E \right] / n$. Note that $\|\tilde{q}\|_E / n = O_p(1)$ and $\|\rho(Z, \alpha_0)\|_E^2 / n = O_p(1)$ so that $\tilde{B}_n = O_p(1)$. Since $\| \cdot \|_{s, \alpha}$ is bounded on $\mathcal{A} \times \mathcal{A}$ by the compactness of the parameter space, there is a constant $C$ such that

\[
\left| \hat{Q}_n(\hat{\alpha}) - \hat{Q}_n(\alpha) \right| \\
\leq \left( \|\hat{g}(\alpha) - \tilde{g}(\alpha)\|_E^2 + 2\|\hat{g}(\alpha)\|_E \cdot \|\tilde{g}(\alpha) - \tilde{g}(\alpha)\|_E \right) / n
\]

\[
\leq \left( \|\hat{g}(\alpha) - \hat{g}(\alpha)\|_E^2 + 2\|\hat{g}(\alpha)\|_E \cdot \|\hat{g}(\alpha) - \hat{g}(\alpha)\|_E \right) / n
\]

\[
\leq \left( \|\tilde{g}(\alpha) - \hat{g}(\alpha)\|_E^2 + 2\|\hat{g}(\alpha) - \hat{g}(\alpha)\|_E \cdot \|\hat{g}(\alpha) - \hat{g}(\alpha)\|_E \right) / n
\]

\[
= \left\{ \sum_{i=1}^n \sum_{j=1}^{d_p} \left( \int (\rho_j(Z, \hat{\alpha}) - \rho_j(Z, \alpha)) \hat{f}_{Y|X}(y \mid x_i; \hat{\psi}) dy \right)^2 \right. \]

\[
+ 2 \left[ \left( \sum_{i=1}^n \sum_{j=1}^{d_p} \left( \int (\rho_j(Z, \hat{\alpha}) - \rho_j(Z, \alpha)) \hat{f}_{Y|X}(y \mid x_i; \hat{\psi}) dy \right)^2 \right) \right. \]

\[
\left. \times \left( \sum_{i=1}^n \sum_{j=1}^{d_p} \left( \int (\rho_j(Z, \hat{\alpha}) - \rho_j(Z, \alpha)) \hat{f}_{Y|X}(y \mid x_i; \hat{\psi}) dy \right)^2 \right) \right]^{1/2}
\]

\[
+ 2 \left[ \left( \sum_{i=1}^n \sum_{j=1}^{d_p} \left( \int \rho_j(Z, \alpha_0) \hat{f}_{Y|X}(y \mid x_i; \hat{\psi}) dy \right)^2 \right) \right. \]

\[
\left. \times \left( \sum_{i=1}^n \sum_{j=1}^{d_p} \left( \int \rho_j(Z, \alpha_0) \hat{f}_{Y|X}(y \mid x_i; \hat{\psi}) dy \right)^2 \right) \right]^{1/2} / n
\]

\[
\leq \left\{ \sum_{i=1}^n \sum_{j=1}^{d_p} \left( \int b_j(Z) \hat{f}_{Y|X}(y \mid x_i; \hat{\psi}) dy \right)^2 \right. \|\hat{\alpha} - \alpha\|_{s, \alpha}^{2\nu}
\]

\[
+ 2 \left[ \left( \sum_{i=1}^n \sum_{j=1}^{d_p} \left( \int b_j(Z) \hat{f}_{Y|X}(y \mid x_i; \hat{\psi}) dy \right)^2 \right) \right. \|\alpha - \alpha_0\|_{s, \alpha}^{2\nu} \right\}
\]
\[
\times \left( \sum_{i=1}^{n} \sum_{j=1}^{d_{\rho}} \left( \int b_{j}(Z) \hat{f}_{Y|X}(y \mid x_{i}; \hat{\psi}) dy \right)^{2} \| \hat{\alpha} - \alpha \|_{s,\alpha}^{2} \right) \right]^{1/2}
\]
\[+ 2 \left[ \left( \sum_{i=1}^{n} \sum_{j=1}^{d_{\rho}} \left( \int \rho_{j}(Z, \alpha_{0}) \right)^{2} \right) \times \left( \sum_{i=1}^{n} \sum_{j=1}^{d_{\rho}} \left( \int b_{j}(Z) \hat{f}_{Y|X}(y \mid x_{i}; \hat{\psi}) dy \right)^{2} \| \hat{\alpha} - \alpha \|_{s,\alpha}^{2} \right) \right]^{1/2} \right] / n
\]
\[
= \left\{ \sum_{i=1}^{n} \sum_{j=1}^{d_{\rho}} \left( \int b_{j}(Z) \hat{f}_{Y|X}(y \mid x_{i}; \hat{\psi}) dy \right)^{2} \| \hat{\alpha} - \alpha \|_{s,\alpha}^{2} \right. \\
+ 2 \sum_{i=1}^{n} \sum_{j=1}^{d_{\rho}} \left( \int b_{j}(Z) \hat{f}_{Y|X}(y \mid x_{i}; \hat{\psi}) dy \right)^{2} \| \alpha - \alpha_{0} \|_{s,\alpha}^{2} \\
+ 2 \left[ \left( \sum_{i=1}^{n} \sum_{j=1}^{d_{\rho}} \left( \int \rho_{j}(Z, \alpha_{0}) \hat{f}_{Y|X}(y \mid x_{i}; \hat{\psi}) dy \right)^{2} \right) \times \left( \sum_{i=1}^{n} \sum_{j=1}^{d_{\rho}} \left( \int b_{j}(Z) \hat{f}_{Y|X}(y \mid x_{i}; \hat{\psi}) dy \right)^{2} \right) \right]^{1/2} \left. \| \hat{\alpha} - \alpha \|_{s,\alpha}^{2} / n \right\}
\]
\[
\leq B_{n} \| \hat{\alpha} - \alpha \|_{s,\alpha}^{2} / n,
\]

where \( B_{n} = C \hat{B}_{n} \) for some constant \( C \) and

\[
\hat{B}_{n} = \left\{ \sum_{i=1}^{n} \sum_{j=1}^{d_{\rho}} \left( \int b_{j}(Z) \hat{f}_{Y|X}(y \mid x_{i}; \hat{\psi}) dy \right)^{2} \\
+ 2 \left[ \left( \sum_{i=1}^{n} \sum_{j=1}^{d_{\rho}} \left( \int (\rho_{j}(Z, \alpha_{0})) \hat{f}_{Y|X}(y \mid x_{i}; \hat{\psi}) dy \right)^{2} \right) \times \left( \sum_{i=1}^{n} \sum_{j=1}^{d_{\rho}} \left( \int b_{j}(Z) \hat{f}_{Y|X}(y \mid x_{i}; \hat{\psi}) dy \right)^{2} \right) \right]^{1/2} / n. \right\}
\]

Then hypothesis (iii) follows by \( B_{n} = C \cdot O_{p}(1) \). □

**Proof of Theorem 4.1** (i) See Lemma 2 in Hu and Schennach (2008).

(ii) We prove the results by verifying the hypotheses of Lemma A.1 of Newey and Powell (2003). Hypothesis (i) follows by Theorem 4.1 of Newey and Powell (2003). Hypothesis (ii) follows by Lemma A.1. Note that hypotheses (i) and (iii) of Lemma A.1 are satisfied by Assumption 4.6 and hypothesis (ii) of Lemma A.1 is satisfied by the result in Theorem 4.1 (i). Finally, we verify hypothesis (iii) by choosing \( \Pi_{n, \alpha} \in A_{n} \) such that \( \| \Pi_{n, \alpha} - \alpha \|_{s,\alpha} = o(1) \). □
Lemma A.2 Suppose that \( \| \hat{\beta} - \beta_0 \|_{\beta} = o_p(n^{-1/4}) \). Then we have (i) under Assumptions 4.1, 4.6, 4.8 and 4.9, \( \| \hat{g}(\alpha) - g(\alpha) \|^2_E/n = o_p(n^{-1/2}) \) uniformly over \( \alpha \in \mathcal{A} \); (ii) under Assumptions 4.1-4.2 and 5.1, \( \| \hat{g}(\alpha) \|^2_E/n = O_p(\delta_{1n}) \) such that \( \delta_{1n} = o(1) \).

Proof of Lemma A.2 (i) From the proof of Lemma A.1, we have

\[
E \left[ \| \hat{g}(\alpha) - g(\alpha) \|^2_E/n \right] \\
= \frac{1}{n} \sum_{j=1}^{d_{p}} \left[ \left( \int \rho_j(Z, \alpha) \left( \hat{f}_{Y|X}(y | x_i; \hat{\psi}) - f_{Y|X}(y | x_i; \psi_0) \right) dy \right)^2 \right] \\
= \frac{1}{n} \sum_{j=1}^{d_{p}} \left[ \left( \int \int \rho_j(Z, \alpha) \left( \hat{f}_{Y|X}(y | y_2, x_{1i}; \hat{\psi}) \hat{f}_{Y|X}(y_2 | x_{2i}, x_{1i}) - f_{Y|X}(y | y_2, x_{1i}; \psi_0) f_{Y|X}(y_2 | x_{2i}, x_{1i}) \right) dy_1 dy_2 \right)^2 \right] \\
= \frac{1}{n} \sum_{j=1}^{d_{p}} \left[ \left( \int \int \rho_j(Z, \alpha) \right) \times \{ f_{Y|X}(y_2 | x_{2i}, x_{1i}) \left( \hat{f}_{Y|X}(y_1 | y_2, x_{1i}; \hat{\psi}) - f_{Y|X}(y_1 | y_2, x_{1i}; \psi_0) \right) \right. \\
\left. + \hat{f}_{Y|X}(y_1 | y_2, x_{1i}; \hat{\psi}) \left( \hat{f}_{Y|X}(y_2 | x_{2i}, x_{1i}) - f_{Y|X}(y_2 | x_{2i}, x_{1i}) \right) \} dy_1 dy_2 \right)^2 \right] \\
= \frac{1}{n} \sum_{j=1}^{d_{p}} \left[ \left( \int \int \rho_j(Z, \alpha) \right) \\
\times \left\{ \left( \frac{1}{f_{Y|X}^*(y_2 | y_{2i} | y_2, x_{1i})} \right) \frac{d}{d\psi} f_{Y|X}(y_1 | y_2, x_{1i}; \psi_0) [\hat{\psi} - \psi_0] f_{Y|X}^*(y_{2i} | y_2, x_{1i}) \right. \\
\left. \times f_{Y|X}(y_2 | x_{2i}, x_{1i}) \right. \\
\left. \times f_{Y|X}(y_1 | y_2, x_{1i}; \hat{\psi}) \right. \\
\left. + \frac{\hat{f}_{Y|X}(y_1 | y_2, x_{1i}; \hat{\psi})}{f_{Y|X}^*(y_2 | y_{2i} | y_2, x_{1i})} \right) f_{Y|X}(y_1 | y_2, x_{1i}; \psi_0) \\
\left. \times f_{Y|X}^*(y_{2i} | y_2, x_{1i}) \right\} dy_1 dy_2 \right)^2 \right] 
\]
\begin{align*}
\leq & \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d_{p}} E \left[ \int \left( \int \rho_j(Z, \alpha) \right. \right. \\
& \times \left. \left. \left\{ \frac{1}{f_{Y_i^*}(y_i^* \mid y_2, x_{1i})} \frac{d}{d\psi} f_{Y_i \mid Y_2 X_1}(y_1 \mid y_2, x_{1i}; \psi_0)[\hat{\psi} - \psi_0] f_{Y_i^*}(y_i^* \mid y_2, x_{1i}) \\
& \times f_{Y_i \mid X_2 X_1}(y_2 \mid x_{2i}, x_{1i}) \right. \right. \\
& \left. \left. + \left( \frac{\hat{f}_{Y_i \mid Y_2 X_1}(y_1 \mid y_2, x_{1i}; \hat{\psi})}{f_{Y_i \mid X_2 X_1}(y_1 \mid y_2, x_{1i}; \psi_0) f_{Y_i^*}(y_i^* \mid y_2, x_{1i})} \right) f_{Y_i \mid X_2 X_1}(y_1 \mid y_2, x_{1i}; \psi_0) \\
& \times f_{Y_i^*}(y_i^* \mid y_2, x_{1i})[f_{Y_i \mid X_2 X_1}(y_2 \mid x_{2i}, x_{1i}) - f_{Y_i \mid X_2 X_1}(y_2 \mid x_{2i}, x_{1i})] \right\} dy_2 \right) dy_1 \\
= & \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d_{p}} \int E \left[ \left( \frac{1}{f_{Y_i^*}(y_i^* \mid x_i; \beta_0)} \right) \frac{d}{d\psi} f_{Y_i \mid X}(y_i^* \mid x_i; \beta_0) \int \rho_j(Z, \alpha) \right. \\
& \times \left. \left\{ \frac{1}{f_{Y_i^*}(y_i^* \mid y_2, x_{1i})} \frac{d}{d\psi} f_{Y_i \mid Y_2 X_1}(y_1 \mid y_2, x_{1i}; \psi_0)[\hat{\psi} - \psi_0] f_{Y_i^*}(y_i^* \mid y_2, x_{1i}) \\
& \times f_{Y_i \mid X_2 X_1}(y_2 \mid x_{2i}, x_{1i}) \right. \right. \\
& \left. \left. + \left( \frac{\hat{f}_{Y_i \mid Y_2 X_1}(y_1 \mid y_2, x_{1i}; \hat{\psi})}{f_{Y_i \mid X_2 X_1}(y_1 \mid y_2, x_{1i}; \psi_0) f_{Y_i^*}(y_i^* \mid y_2, x_{1i})} \right) f_{Y_i \mid X_2 X_1}(y_1 \mid y_2, x_{1i}; \psi_0) \\
& \times f_{Y_i^*}(y_i^* \mid y_2, x_{1i})[f_{Y_i \mid X_2 X_1}(y_2 \mid x_{2i}, x_{1i}) - f_{Y_i \mid X_2 X_1}(y_2 \mid x_{2i}, x_{1i})] \right\} dy_2 \right) dy_1 \\
= & \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d_{p}} \int \left( \sup_x f_{Y_i^*}(y_i^* \mid x_i; \beta_0) \right) \left( \sup_y \rho_j(Z, \alpha) \right) E \left[ \left( \frac{1}{f_{Y_i^*}(y_i^* \mid x_i; \beta_0)} \right) \frac{d}{d\psi} f_{Y_i \mid X}(y_i^* \mid x_i; \beta_0) \int \rho_j(Z, \alpha) \right. \\
& \times \left. \left\{ \frac{1}{f_{Y_i^*}(y_i^* \mid y_2, x_{1i})} \frac{d}{d\psi} f_{Y_i \mid Y_2 X_1}(y_1 \mid y_2, x_{1i}; \psi_0)[\hat{\psi} - \psi_0] f_{Y_i^*}(y_i^* \mid y_2, x_{1i}) \\
& \times f_{Y_i \mid X_2 X_1}(y_2 \mid x_{2i}, x_{1i}) \right. \right. \\
& \left. \left. + \left( \frac{\hat{f}_{Y_i \mid Y_2 X_1}(y_1 \mid y_2, x_{1i}; \hat{\psi})}{f_{Y_i \mid X_2 X_1}(y_1 \mid y_2, x_{1i}; \psi_0) f_{Y_i^*}(y_i^* \mid y_2, x_{1i})} \right) f_{Y_i \mid X_2 X_1}(y_1 \mid y_2, x_{1i}; \psi_0) \\
& \times f_{Y_i^*}(y_i^* \mid y_2, x_{1i})[f_{Y_i \mid X_2 X_1}(y_2 \mid x_{2i}, x_{1i}) - f_{Y_i \mid X_2 X_1}(y_2 \mid x_{2i}, x_{1i})] \right\} dy_2 \right) dy_1 \\
\leq & \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d_{p}} \int \left( \sup_x f_{Y_i^*}(y_i^* \mid x_i; \beta_0) \right) \left( \sup_y \rho_j(Z, \alpha) \right) \\
& \times \left( \max_x \left( \sup_y \frac{1}{f_{Y_i^*}(y_i^* \mid y_2, x_{1i})} \frac{\hat{f}_{Y_i \mid Y_2 X_1}(y_1 \mid y_2, x_{1i}; \hat{\psi})}{f_{Y_i \mid X_2 X_1}(y_1 \mid y_2, x_{1i}; \psi_0) f_{Y_i^*}(y_i^* \mid y_2, x_{1i})} \right) \right) \text{ dy}_1 \\
\end{align*}
\[
\times E \left[ \left( \frac{1}{f_{Y*|X}(y_i^* | x_i; \beta_0)} \right) \int \frac{d}{d\psi} f_{Y_1|X_1}(y_1 | y_2, x_{1i}; \psi_0) [\hat{\psi} - \psi] f_{Y_2|X_2X_1}(y_{2i}^* | y_2, x_{1i}) \\
\times f_{Y_2|X_2X_1}(y_2 | x_{2i}, x_{1i}) \right] dy_2 \\
+ \int f_{Y_1|X_1}(y_1 | y_2, x_{1i}; \psi_0) \\
\times f_{Y_2|X_2X_1}(y_{2i}^* | y_2, x_{1i}) [\hat{f}_{Y_1|X_2X_1}(y_1 | y_2, x_{1i}) - f_{Y_2|X_2X_1}(y_2 | x_{2i}, x_{1i})] dy_2 \right)^2 \right] dy_1
\]

\[
\leq \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d_p} \int \left( \sup_x f_{Y*|X}(y_i^* | x_i; \beta_0) \right) \left( \sup_{y_2} \rho_j(Z, \alpha) \right) \\
\times \left( \max \left\{ \sup_x \left( \frac{1}{f_{Y_2*|X_2X_1}(y_{2i}^* | y_2, x_{1i})} \right), \sup_x \left( \frac{\hat{f}_{Y_1|X_2X_1}(y_1 | y_2, x_{1i}; \hat{\psi})}{f_{Y_1|X_2X_1}(y_1 | y_2, x_{1i}; \psi_0) f_{Y_2*|X_2X_1}(y_{2i}^* | y_2, x_{1i})} \right) \right\} \right) \\
\times E \left\{ \left( \frac{d \ln f_{Y*|X}(y_i^* | x_i; \beta_0)}{d\beta} \right)^2 \right\} dy_1
\]

\[
\leq \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d_p} \int \left( \sup_x f_{Y*|X}(y_i^* | x_i; \beta_0) \right) \left( \sup_{y_2} \rho_j(Z, \alpha) \right) \\
\times \left( \max \left\{ \sup_x \left( \frac{1}{f_{Y_2*|X_2X_1}(y_{2i}^* | y_2, x_{1i})} \right), \sup_x \left( \frac{\hat{f}_{Y_1|X_2X_1}(y_1 | y_2, x_{1i}; \hat{\psi})}{f_{Y_1|X_2X_1}(y_1 | y_2, x_{1i}; \psi_0) f_{Y_2*|X_2X_1}(y_{2i}^* | y_2, x_{1i})} \right) \right\} \right) dy_1 \|\hat{\beta} - \beta_0\|_\beta^2
\]

= o(n^{-1/2}),

since \(\|\hat{\beta} - \beta_0\|_\beta = o_p(n^{-1/4})\). Thus the result follows by the Markov inequality.

(iii) See Corollary A.1 (iii) in Ai and Chen (2003).  \(\square\)

**Lemma A.3**  
(i) Under Assumptions 4.1-4.2, 4.3(ii), 4.6(iii), and 4.10, we obtain uniformly over \(\alpha \in \{A_n : \|\alpha - \alpha_0\|_\alpha = o(1)\} : (1/n) \sum_{i=1}^{n} \|m(X_i, \alpha)\|_E^2 - E[\|m(X, \alpha)\|_E^2] = o_p(n^{-1/2}).\)

(ii) Suppose that \(\|\hat{\beta} - \beta_0\|_\beta = o_p(n^{-1/4})\). Under Assumptions 4.1-4.3, 4.6, 4.8, 4.10, we obtain uniformly over \(\alpha \in A, \|\alpha - \alpha_0\|_\alpha = o(\eta_n) : (1/n) \sum_{i=1}^{n} \|m(X_i, \alpha)\|_E^2 = o_p(\eta_n^2)\) and \((1/n) \sum_{i=1}^{n} \|m(X_i, \alpha)\|_E^2 = o_p(\eta_n^2),\) where \(\eta_n = n^{-\tau}\) with \(\tau \leq 1/4.\)

**Proof of Lemma A.3**  
(i) See Corollary A.2(i) of Ai and Chen (2003).

(ii) The result follows from applying Lemma A.2(i) and A.3(i), and \(E[\|m(X, \alpha)\|_E^2] = o(\eta_n^2)\) by Assumptions 4.3(ii) and 4.9.  \(\square\)
Lemma A.4 Suppose that $\|\hat{\beta} - \beta_0\|_\beta = o_p(n^{-1/4})$. Assumptions 4.1-4.3, 4.6, and 4.8-4.9 imply: (i) $\hat{Q}_n(\alpha) - Q_n(\alpha) = o_p(n^{-1/4})$ uniformly over $\alpha \in \mathcal{A}_n$; and (ii) $\hat{Q}_n(\alpha) - Q_n(\alpha_0) - \{Q_n(\alpha) - Q_n(\alpha_0)\} = o_p(\eta_n n^{-1/4})$ uniformly over $\alpha \in \mathcal{A}_n$ with $\|\alpha - \alpha_0\|_\alpha \leq o(\eta_n)$, where $\eta_n = n^{-\tau}$ with $\tau \leq 1/4$.

Proof of Lemma A.4 (i) The result follows from Lemma A.2(i) and Assumption 4.3. (ii) The result follows from Lemma A.3 and Lemma A.4(ii). □

Proof of Theorem 4.2 (i) See Theorem 2 in Hu and Schennach (2008). (ii) It follows from a similar argument of Theorem 3.1 in Ai and Chen (2003). □

Let
\[
\frac{dg(\alpha)}{d\alpha}[v_{2n}^*] = \left( \frac{dm(X,\alpha)}{d\alpha}[v_{2n}^*], \cdots, \frac{d\hat{m}(X_n,\alpha)}{d\alpha}[v_{2n}^*] \right),
\]
where
\[
\frac{dm(X,\alpha)}{d\alpha}[v_{2n}^*] = \int \frac{d\rho(Z,\alpha)}{d\alpha}[v_{2n}^*] f_{Y|x}(y|x;\hat{\psi}) dy,
\]
by the interchangability of integral and derivative. Recall the definition of neighborhoods $\mathcal{N}_{02n}$ and $\mathcal{N}_{02}$ introduced in Section 5.

Lemma A.5 (i) Assumptions 4.1, 4.8 and 5.1, 5.3, 5.5-5.6 imply:
\[
\sup_{\alpha \in \mathcal{N}_{02n}} \frac{1}{n} \left\| \frac{d\hat{g}(\hat{\alpha})}{d\alpha}[v_{2n}^*] - \frac{dg(\alpha_0)}{d\alpha}[v_{2n}^*] \right\|_E^2 = o_p(n^{-1/2}).
\]
(ii) In addition, if $\|\hat{\beta} - \beta_0\|_\beta = o_p(n^{-1/4})$ holds, then
\[
\sup_{\alpha \in \mathcal{N}_{02n}} \frac{1}{n} \left\| \frac{d\hat{g}(\hat{\alpha})}{d\alpha}[v_{2n}^*] - \frac{dg(\hat{\alpha})}{d\alpha}[v_{2n}^*] \right\|_E^2 = o_p(n^{-1/2}).
\]

Proof of Lemma A.5 (i) The result can be proved by the same argument of Corollary C.1 (ii) of Ai and Chen (2003).
(ii) We have

\[
E \left[ \frac{1}{n} \left\| \frac{d\hat{g}(\hat{\alpha})}{d\alpha} [v_{2n}^*] - \frac{dg(\hat{\alpha})}{d\alpha} [v_{2n}^*] \right\|_E^2 \right]
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d_p} E \left[ \left( \int \frac{d\rho_j(Z, \hat{\alpha})}{d\alpha} [v_{2n}^*] \left( \hat{f}_{Y|X}(y \mid x_i; \hat{\psi}) - f_{Y|X}(y \mid x_i; \psi_0) \right) \right)^2 \right]
\]

\[= o(n^{-1/2}), \]

uniformly over \( \hat{\alpha} \in \mathcal{N}_{02n} \) since \( \|\hat{\beta} - \beta_0\|_\beta = o_p(n^{-1/4}) \). Thus the result follows by the Markov inequality. 

\[\blacksquare\]

Let

\[
\frac{dQ_n(\alpha)}{d\alpha} [v_{2n}^*] = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d\hat{m}(X_i, \alpha)}{d\alpha} [v_{2n}^*] \right\} \left( \bar{\Sigma}(X_i) \right)^{-1} \left\{ \frac{d\hat{m}(X_i, \alpha)}{d\alpha} [v_{2n}^*] \right\},
\]

and

\[
\frac{d^2\hat{g}(\alpha)}{d\alpha d\alpha} [v_{2n}^*, v_{2n}^*] = \left( \frac{d^2\hat{m}(X, \alpha)}{d\alpha d\alpha} [v_{2n}^*, v_{2n}^*], \cdots, \frac{d^2\hat{m}(X, \alpha)}{d\alpha d\alpha} [v_{2n}^*, v_{2n}^*] \right)^\prime,
\]

where

\[
\frac{d^2\hat{m}(X, \alpha)}{d\alpha d\alpha} [v_{2n}^*, v_{2n}^*] = \int \frac{d^2\rho(Z, \alpha)}{d\alpha d\alpha} [v_{2n}^*, v_{2n}^*] \hat{f}_{Y|X}(y \mid x; \hat{\psi}) dy,
\]

\[
\frac{d^2m(X, \alpha)}{d\alpha d\alpha} [v_{2n}^*, v_{2n}^*] = \int \frac{d^2\rho(Z, \alpha)}{d\alpha d\alpha} [v_{2n}^*, v_{2n}^*] f_{Y|X}(y \mid x; \psi_0) dy.
\]

by the interchangability of integral and derivative.

**Lemma A.6** Suppose that \( \|\hat{\beta} - \beta_0\|_\beta = o_p(n^{-1/4}) \). (i) Under Assumptions 4.1, 4.3-4.4 4.6, 4.8, 4.10, and 5.8, we have

\[
\sup_{\hat{\alpha} \in \mathcal{N}_{02n}} \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d^2\hat{m}(X_i, \hat{\alpha})}{d\alpha d\alpha} [v_{2n}^*, v_{2n}^*] \right\} \left( \bar{\Sigma}(X_i) \right)^{-1} \hat{m}(X_i, \hat{\alpha}) = o_p(n^{-1/4}).
\]
(ii) Under Assumptions 4.1, 4.3, 4.8, 5.1(ii), 5.3, 5.5-5.6, we have

\[ \sup_{\tilde{\alpha} \in N_{02n}} \frac{d\hat{Q}_n(\tilde{\alpha})}{d\alpha}[v_{2n}^*] = \frac{dQ_n(\alpha_0)}{d\alpha}[v_{2n}^*] + o_p(n^{-1/4}). \]

**Proof of Lemma A.6** Proof is similar to Ai and Chen (2003). (i) For some constant \( C \), Assumption 4.3 implies

\[ \left| \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d^2\hat{m}(X_i, \tilde{\alpha})}{d\alpha^2}[v_{2n}^*, v_{2n}^*] \right\} \right| \leq C \sqrt{\frac{\|d\hat{g}(\alpha)[v_{2n}^*, v_{2n}^*]\|^2}{n}}. \]

Then the result follows from Lemma A.3(ii) because we have that uniformly over \( \tilde{\alpha} \in N_{02n} \),

\[ \frac{d^2\hat{g}(\alpha)[v_{2n}^*, v_{2n}^*]}{d\alpha^2} / n \leq c_1(Z)^2 = O_p(1) \]

by Assumption 5.8.

(ii) Uniformly over \( \tilde{\alpha} \in N_{02n} \),

\[ \frac{d\hat{Q}_n(\tilde{\alpha})}{d\alpha}[v_{2n}^*] = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d^2\hat{m}(X_i, \tilde{\alpha})}{d\alpha^2}[v_{2n}^*, v_{2n}^*] \right\} \cdot \left[ \hat{\Sigma}(X_i) \right]^{-1} \left\{ \frac{d\hat{m}(X_i, \tilde{\alpha})}{d\alpha}[v_{2n}^*] \right\} \]

\[ + \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{dm(X_i, \alpha_0)}{d\alpha}[v_{2n}^*] \right\} \cdot \left[ \hat{\Sigma}(X_i) \right]^{-1} \left\{ \frac{d\hat{m}(X_i, \tilde{\alpha})}{d\alpha}[v_{2n}^*] - \frac{dm(X_i, \alpha_0)}{d\alpha}[v_{2n}^*] \right\} \]

\[ + \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{dm(X_i, \alpha_0)}{d\alpha}[v_{2n}^*] \right\} \cdot \left[ \hat{\Sigma}(X_i) \right]^{-1} \left\{ \frac{dm(X_i, \alpha_0)}{d\alpha}[v_{2n}^*] \right\} \]

\[ + \frac{dQ_n(\alpha_0)}{d\alpha}[v_{2n}^*]. \]

The result follows from Assumption 4.3 and Lemma A.5. \( \square \)

**Lemma A.7** (i) Under Assumptions 4.1 4.3-4.4, 4.6, 4.8, 4.10, 5.1(ii), 5.3, 5.5-5.6, we have uniformly
over $\tilde{\alpha} \in \mathbb{N}_{02n}$:

$$
\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d\hat{m}(X_i, \tilde{\alpha})}{d\alpha} [v_{2n}^*] \right\}' [\tilde{\Sigma}(X_i)]^{-1} \hat{m}(X_i, \tilde{\alpha})
= \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{m(X_i, \alpha_0)}{d\alpha} [v_{2}^*] \right\}' [\Sigma(X_i)]^{-1} \hat{m}(X_i, \tilde{\alpha}) + o_p(n^{-1/2}).
$$

(ii) Under Assumptions 4.1, 5.1(ii), 5.3, 5.5-5.7, we have uniformly over $\tilde{\alpha} \in \mathbb{N}_{02n}$:

$$
\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d \hat{m}(X_i, \alpha_0)}{d\alpha} [v_{2}^*] \right\}' [\Sigma(X_i)]^{-1} \{ \hat{m}(X_i, \tilde{\alpha}) - \hat{m}(X_i, \alpha_0) \} = (v_{2}^*, \tilde{\alpha} - \alpha_0) + o_p(n^{-1/2}).
$$

(iii) Under Assumptions 4.1, 4.3(ii), 4.8, 5.1(iii), 5.3, we have

$$
\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d \hat{m}(X_i, \alpha_0)}{d\alpha} [v_{2}^*] \right\}' [\Sigma(X_i)]^{-1} \hat{m}(X_i, \alpha_0)
= \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{m(X_i, \alpha_0)}{d\alpha} [v_{2}^*] \right\}' [\Sigma(X_i)]^{-1} \tilde{\rho}(X_i, \alpha_0) + o_p(n^{-1/2}).
$$

**Proof of Lemma A.7** (i) Uniformly over $\tilde{\alpha} \in \mathbb{N}_{02n}$,

$$
\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d \hat{m}(X_i, \tilde{\alpha})}{d\alpha} [v_{2n}^*] \right\}' [\tilde{\Sigma}(X_i)]^{-1} \hat{m}(X_i, \tilde{\alpha})
- \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{m(X_i, \alpha_0)}{d\alpha} [v_{2}^*] \right\}' [\Sigma(X_i)]^{-1} \hat{m}(X_i, \tilde{\alpha})
= \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d \hat{m}(X_i, \tilde{\alpha})}{d\alpha} [v_{2n}^*] - \frac{d \hat{m}(X_i, \alpha_0)}{d\alpha} [v_{2n}^*] \right\}' [\tilde{\Sigma}(X_i)]^{-1} \hat{m}(X_i, \tilde{\alpha})
+ \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{m(X_i, \alpha_0)}{d\alpha} [v_{2n}^*] \right\}' [\tilde{\Sigma}(X_i)]^{-1} \hat{m}(X_i, \tilde{\alpha})
+ \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{m(X_i, \alpha_0)}{d\alpha} [v_{2n}^*] \right\}' [\Sigma(X_i)]^{-1} \hat{m}(X_i, \tilde{\alpha})
\equiv A_1 + A_2 + A_3.
$$

Then the result follows from the fact that $A_1 = o_p(n^{-1/2})$ by Lemma A.3(ii), A.5(ii), and Assumption 4.3(ii); $A_2 = o_p(n^{-1/2})$ by Assumption 4.3(iii) and Lemma A.3(ii); $A_3 = o_p(n^{-1/2})$ by Assumption 5.3 and Lemma A.3(ii).
(ii) Let \( \varphi(X, v^*_2) = \left( \frac{d m(X, \alpha_0)}{d \alpha} [v^*_2] \right)' \Sigma(X)^{-1} \) and let

\[
\tilde{F} = \left\{ \varphi(X, v^*)\tilde{m}(X, \alpha) : \alpha \in \mathcal{N}_{02n}, \tilde{m} \in \Lambda^2_2(\mathcal{X}) \text{ s.t.} \right. \\
\sup_{x \in \mathcal{X}, \alpha \in \mathcal{N}_{02n}} |\tilde{m}(x, \alpha) - m(x, \alpha)| = o(1) \left. \right\},
\]

\[
\mathcal{F} = \left\{ \varphi(X, v^*)m(X, \alpha) : \alpha \in \mathcal{N}_{02n} \right\}.
\]

By a similar argument to Corollary C.3(ii) of Ai and Chen (2003), \( \tilde{F} \) and \( \mathcal{F} \) are Donsker classes, and we have uniformly over \( \alpha \in \mathcal{N}_{02n} \),

\[
\frac{1}{n} \sum_{i=1}^{n} \varphi(X_i, v^*_2)\{\hat{m}(X_i, \alpha) - m(X_i, \alpha)\} - E \left[ \varphi(X_i, v^*_2)\{\hat{m}(X_i, \alpha) - m(X_i, \alpha)\} \right] = o_p(n^{-1/2}),
\]

(20)

\[
= \frac{1}{n} \sum_{i=1}^{n} \varphi(X_i, v^*_2)\{\hat{m}(X_i, \alpha_0) - m(X_i, \alpha_0)\} - E \left[ \varphi(X_i, v^*_2)\{\hat{m}(X_i, \alpha_0) - m(X_i, \alpha_0)\} \right] = o_p(n^{-1/2}),
\]

(21)

\[
= \frac{1}{n} \sum_{i=1}^{n} \varphi(X_i, v^*_2)\{\hat{m}(X_i, \alpha) - m(X_i, \alpha)\} - E \left[ \varphi(X_i, v^*)\{m(X_i, \alpha) - m(X_i, \alpha)\} \right] = o_p(n^{-1/2}).
\]

(22)

From Eqns. (14) and (15),

\[
\frac{1}{n} \sum_{i=1}^{n} \varphi(X_i, v^*_2)\{\hat{m}(X_i, \hat{\alpha}) - \hat{m}(X_i, \alpha_0)\} \]

\[
= \frac{1}{n} \sum_{i=1}^{n} \varphi(X_i, v^*_2)\{m(X_i, \hat{\alpha}) - m(X_i, \alpha_0)\} + E \left[ \varphi(X_i, v^*_2)\{\hat{m}(X_i, \hat{\alpha}) - \hat{m}(X_i, \alpha_0)\} \right] - E \left[ \varphi(X_i, v^*_2)\{m(X_i, \hat{\alpha}) - m(X_i, \alpha_0)\} \right] + o_p(n^{-1/2}).
\]

(23)

Let \( \tilde{\varphi}(X, v^*_2) = \int \int \varphi(X, v^*_2) \hat{f}_{Y_1|X_2|X_1}(y_1 | y_2, x_1; \hat{\phi}, \hat{\eta})dy_1 \int \hat{f}_{Y_2|X_2|X_1}(y_2 | x_2, x_1)dy_2 \). Then we have that

\[
E \left[ \varphi(X_i, v^*_2)\{\hat{m}(X_i, \hat{\alpha}) - \hat{m}(X_i, \alpha_0)\} \right] = E \left[ \tilde{\varphi}(X_i, v^*_2)\{m(X_i, \hat{\alpha}) - m(X_i, \alpha_0)\} \right],
\]

(24)
\[ E[\tilde{\varphi}(X_i, v_2^*)\{m(X_i, \hat{\alpha}) - m(X_i, \alpha_0)\}] - E[\varphi(X_i, v_2^*)\{m(X_i, \hat{\alpha}) - m(X_i, \alpha_0)\}] \\
= E[(\tilde{\varphi}(X_i, v_2^*) - \varphi(X_i, v_2^*)\{m(X_i, \hat{\alpha}) - m(X_i, \alpha_0)\}] \\
= o_p(n^{-1/2}). \]  

Plugging Eqns. (16), (18) and (19) into (17) gives for some \( \bar{\alpha} \in \mathcal{N}_{02} \), a convex combination of \( \hat{\alpha} \) and \( \alpha_0 \) that

\[
\frac{1}{n} \sum_{i=1}^{n} \varphi(X_i, v_2^*)\{m(X_i, \hat{\alpha}) - m(X_i, \alpha_0)\} \\
= \frac{1}{n} \sum_{i=1}^{n} \varphi(X_i, v_2^*)\{m(X_i, \hat{\alpha}) - m(X_i, \alpha_0)\} + o_p(n^{-1/2}) \\
= E[\varphi(X_i, v_2^*)\{m(X_i, \hat{\alpha}) - m(X_i, \alpha_0)\}] + o_p(n^{-1/2}) \\
= E[\varphi(X_i, v_2^*)\frac{dm(X_i, \hat{\alpha})}{d\alpha} [\hat{\alpha} - \alpha_0] + \varphi(X_i, v_2^*)\frac{dm(X_i, \alpha_0)}{d\alpha} [\hat{\alpha} - \alpha_0]] + o_p(n^{-1/2}) \\
= (v_2^*, \hat{\alpha} - \alpha_0)_{\alpha} + E[\varphi(X_i, v_2^*)\left(\frac{dm(X_i, \hat{\alpha})}{d\alpha} [\hat{\alpha} - \alpha_0] - \frac{dm(X_i, \alpha_0)}{d\alpha} [\hat{\alpha} - \alpha_0]\right)] + o_p(n^{-1/2}) \\
= (v_2^*, \hat{\alpha} - \alpha_0)_{\alpha} + o_p(n^{-1/2}),
\]

where the third, fourth and fifth equalities follow from the mean value theorem, the definition of \( (v_2^*, \hat{\alpha} - \alpha_0)\), and Assumption 5.1(ii) and 5.7, respectively.

(iii) From the definition of \( \tilde{\varphi}(X_i, v_2^*)\), we have

\[
\frac{1}{n} \sum_{i=1}^{n} \varphi(X_i, v_2^*)\{m(X_i, \alpha_0) - \frac{1}{n} \sum_{i=1}^{n} \varphi(X_i, v_2^*)\rho(Z_i, \alpha_0)\} \\
= \frac{1}{n} \sum_{i=1}^{n} (\tilde{\varphi}(X_i, v_2^*) - \varphi(X_i, v_2^*)) \rho(Z_i, \alpha_0).
\]

Then the result follows from the same argument of Corollary C.3(iii) in Ai and Chen (2003). □

**Proof of Theorem 5.1** (i) See Theorem 3 in Hu and Schennach (2008).

(ii) It follows from a similar argument of Theorem 4.1 in Ai and Chen (2003). □

**Proof of Theorem 5.2**

See Theorem 5.1 in Ai and Chen (2003). □
References


## Table 1: Monte Carlo simulation results

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Zero Mode</th>
<th>Zero Mean</th>
<th>Zero Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infeasible SMD</td>
<td>0.14334</td>
<td>0.15796</td>
<td>0.14059</td>
</tr>
<tr>
<td>SML-SMD</td>
<td>0.14683</td>
<td>0.17255</td>
<td>0.14683</td>
</tr>
<tr>
<td>Inconsistent SMD</td>
<td>0.23990</td>
<td>0.18691</td>
<td>0.15668</td>
</tr>
</tbody>
</table>
Notes: Top figure is food-in and bottom figure is food-out. Our SML-SMD is the solid curve and SMD is dashed curve.

Figure 1: Engel curves for food-in and food-out
Notes: Top figure is fuel and bottom figure is leisure. Our SML-SMD is the solid curve and SMD is dashed curve.

Figure 2: Engel curves for fuel and leisure