Control Variables Approach to Estimate Semiparametric Models of Mismeasured Endogenous Regressors with an Application to U.K. Twin Data

Kyoo il Kim         Suyong Song*
Michigan State University University of Iowa
March 2019

Abstract

We study estimation of semiparametric models with mismeasured endogenous regressors using control variables that ensure the conditional mean independence of endogenous regressors and unobserved causes. We provide a set of sufficient conditions for identification, which control for both endogeneity and measurement error. We propose a sieve-based estimator and derive its asymptotic properties. Given the sieve approximation, our proposed estimator is easy to implement as weighted least squares. Monte Carlo simulations illustrate that our proposed estimator performs well in the finite samples. In an empirical application, we estimate the return to education on earnings using U.K. twin data, in which self-reported education is potentially measured with error and is also correlated with unobserved factors. Our approach utilizes twin’s reported education as a control variable to obtain consistent estimates. We find that a one-year increase in education leads to an 11% increase in hourly wage. The estimate is significantly higher than those from OLS and IV approaches which are potentially biased. The application underscores that our proposed estimator is useful to correct for both endogeneity and measurement error in estimating returns to education.

Keywords: Endogeneity, Control Variables, Nonclassical Measurement Error, Conditional Independence, Returns to Education, Twin Data

JEL Classifications: C14, C21, C26

*Corresponding authors: Kyoo il Kim, kyookim@msu.edu; Suyong Song, suyong-song@uiowa.edu.
1 Introduction

There have been many studies that separately control for endogeneity or measurement error in the regression models. However, estimation methods that account for both problems simultaneously have been scarce in spite of their empirical importance being well recognized. For example, Butcher and Case (1994) study the effect of women’s education on earnings. They note that education is endogenous because it is correlated with unobserved ability, and also the completed education variable may be reported with error.

To develop an approach that accounts for both problems this paper studies estimation of semiparametric models with mismeasured endogenous regressors using control variables. Our main identifying condition is that endogenous regressors and unobserved cause in the outcome equation are conditionally mean independent given the control variables.\(^1\) We then estimate the return to education on earnings using U.K. twin data, in which self-reported education is potentially measured with error and is also correlated with unobserved factors. We utilize the twin’s report as a control variable to estimate the return to education. We contribute to this literature by proposing an estimator correcting for both endogeneity and measurement error in education using only one control or instrumental variable.

We develop our approach for a model below and consider its extensions

\[
Y_1 = \pi_0(Z_1) + G(Y_2; \theta_0, h_0) + \varepsilon, \quad Y_2^* = g(Y_2, e) \tag{1.1}
\]

where \( Y \equiv (Y_1, Y_2) \) is a vector of endogenous (or dependent) variables and \( Z_1 \) is a vector of exogenous regressors. Here \( Y_2 \), that is potentially correlated with \( \varepsilon \), denotes true variable and it is only measured with error as \( Y_2^* \), which we refer to the mismeasured endogenous regressor(s) where \( e \) is the measurement error. The function \( \pi_0(\cdot) \) is nonparametric and the function \( G(\cdot) \) is a known function up to an unknown parameter vector \( (\theta_0, h_0) \), where \( \theta_0 \) and \( h_0 \) denote finite and infinite dimensional parameters, respectively. Examples include a parametric model \( G(Y_2) = Y_2^0\theta \), a partially linear model \( G(Y_2) = Y_2^1\theta + h(Y_2^2) \) such that \( Y_2 = (Y_2^1, Y_2^2) \), and a nonparametric regression model \( G(Y_2) = h(Y_2) \). We let \( (\theta_0, h_0) \) denote the true parameter. Specifically, the model allows dependence of \( h(\cdot) \) on the endogenous variables \( Y_2 \). Moreover, the measurement problem of \( Y_2 \) hinders us from using other existing semiparametric methods. We develop a new approach to tackle the problem.

We utilize the existence of control variables \( V \) such that \( Y_2 \) and \( \varepsilon \) are conditionally mean independent given \( W \equiv (Z_1, V) \).\(^2\) From this conditional mean independence we derive a cond-

\(^1\)Conditional (mean) independence has been utilized for a basis of identification in various econometric and statistical problems including estimation of treatment effects. See, among others, Heckman, Ichimura, and Todd (1998), Dehejia and Wahba (1999), Lechner (2001), Imbens and Newey (2009), and Imbens and Wooldridge (2009).

\(^2\)As a specific example, consider the problem of estimating the effect of family income on children’s health
tional moment restriction for the model (1.1) as
\[ E[(Y_2 - E[Y_2 \mid W])(Y_1 - G(Y_2; \theta_0, h_0)) \mid W] = 0. \] (1.2)

Estimation using this moment restriction, however, raises further difficulty because the endogenous regressor \( Y_2 \) is not observed but measured with error as \( Y_2^* \). We then show this moment condition can be consistently estimated using relevant density functions identified from observable data. Given the identification result we propose to estimate \((\theta_0, h_0)\) using a sieve approach.

An advantage of this approach is that given a sieve approximation of \( h \), the moment condition (1.2) yields a closed-form solution and the proposed estimator is obtained as weighted least squares. Therefore, our proposed estimator is easy to implement for this class of model.\(^3\)

To develop our estimation strategy with mismeasured endogenous regressors in a more general setting, we cast the conditional moment restriction (1.2) into a more general form
\[ m(W, \theta, h(\cdot)) \equiv E[\rho(Y, W, \theta, h(\cdot)) \mid W], \] (1.3)
\[ m(W, \theta_0, h_0(\cdot)) = 0, \]
where we let
\[ \rho(Y, W, \theta, h) = (Y_2 - E[Y_2 \mid W])(Y_1 - G(Y_2; \theta, h)) \]
for the model (1.1). For this general model, although \( Y_2 \) is not observed, we show that the conditional moment function is identified from observables \((Y_1, Y_2^*, W)\) by means of recovering relevant conditional density functions under a set of exclusion restrictions stating that (i) given the true regressors and the control variables, mismeasured regressors do not provide further information on dependent variables, (ii) given the true regressors, control variables do not provide further information on dependent variables, and (iii) given the true regressors, control variables do not provide further information on mismeasured regressors. Our approach builds on an operator-based approach for nonclassical measurement errors (e.g. Hu and Schennach (2008)).\(^4\)

Another class of model that fits into our framework is a triangular nonparametric simul-in Case, Lubotsky, and Paxson (2002), Currie and Stabile (2003), and Condliffe and Link (2008). In this example family income \((Y_2)\) is endogenous because of the dependence between household earning potential and children’s health determinants. However, given parental education as a control variable \((V)\), which acts as a proxy to parental cognitive ability, family income can be treated as being independent of children’s health determinants. Because endogeneity problem comes from the common cause, parental cognitive ability between household earning potential and children’s health determinants, a proxy to parental cognitive ability such as parental education can control for the endogeneity.

\(^3\)In our approach \((\theta_0, h_0)\) is separately estimated from \( \pi_0(Z_1) \) and it is not required to estimate \( \pi_0(Z_1) \) in (1.2). Given \((\theta_0, h_0)\), the identification of \( \pi_0(Z_1) \) is trivial as \( \pi_0(Z_1) = E[Y_1 - G(Y_2; \theta_0, h_0) \mid Z_1] \) if we further impose \( E[\varepsilon \mid Z_1] = 0 \) since \( Z_1 \) is exogenous. If \( \pi_0 \) is also a parameter of interest, once \((\theta_0, h_0)\) is estimated from the moment condition (1.2), we then go back to (1.1) and estimate \( \pi_0(Z_1) \) using a regression of \( Y_1 - G(Y_2; \hat{\theta}_n, \hat{h}_n) \) on \( Z_1 \) where \((\hat{\theta}_n, \hat{h}_n)\) denotes a consistent estimator of \((\theta_0, h_0)\).

\(^4\)For excellent reviews of other existing methods, see e.g. Carroll, Ruppert, Stefanski, and Crainiceanu (2006) and Buonaccorsi (2010).
taneous equations model (e.g. Newey, Powell, and Vella (1999), Pinkse (2000), Su and Ullah (2008), and Florens, Heckman, Meghir, and Vytlacil (2008)) but we allow for endogenous regressor being mismeasured as

\[
\begin{align*}
Y_1 &= \pi_0(Z_1) + h_0(Y_2) + \varepsilon, \\
Y_2 &= r(Z,V), \\
Y_2^* &= g(Y_2, e),
\end{align*}
\]

(1.4)

where \(e\) is the measurement error on \(Y_2\) and \(Z \equiv (Z_1, Z_2)\) with \(Z_2\) being a vector of excluded instruments. If \(Y_2\) is observable, the control variable \(V\) is obtained as the conditional cumulative distribution function (CDF) of \(Y_2\) given \(Z\), \(F_{Y_2|Z}(Y_2|Z)\), under the assumption that \(V\) is a scalar, \(r(\cdot, \cdot)\) is strictly monotonic in \(V\), and \(Z\) is independent of \(V\) (see e.g. Matzkin (2003) and Imbens and Newey (2009)). Therefore, compared to the model (1.1) that does not restrict the first stage equation of \(Y_2\), the triangular model does not allow \(Y_2\) depend on other variables than \((Z,V)\).

For this model the conditional mean independence of \(Y_2\) and \(\varepsilon\) given \(W \equiv (Z_1, V)\) implies

\[
E \left[ (Y_2 - E[Y_2|W]) (Y_1 - h_0(Y_2)) \mid W \right] = 0
\]

(1.5)

which also has the form of the moment function (1.3).

Hahn and Ridder (2017) also study a parametric version of the model (1.4) under the assumption that the instrument \(Z\) is jointly independent with the measurement \((e)\), equation error \((\varepsilon)\), and a mismeasured control \(\tilde{V}^* \equiv Y_2^* - E[Y_2^*|Z]\). Under these assumptions their approach can use this \(\tilde{V}^*\) as the control variable. As compared to their approach, we only assume \(Z\) is independent with the true control \(V\), and this control only satisfies the conditional mean independence condition \(E[Y_2 \varepsilon|W] = E[Y_2|W]E[\varepsilon|W]\).\(^5\) Indeed the three assumptions that \(V\) is a scalar, \(r(\cdot, \cdot)\) is strictly monotonic in \(V\), and \(Z\) is independent of \(V\) do not impose any restrictions on the relationship between \(Y_2\) and \(Z\) (see Hahn and Ridder (2011)). Therefore, the only material assumption for our approach is that the control variable \(V\), defined as \(V \equiv F_{Y_2|Z}(Y_2|Z)\), satisfies the conditional mean independence condition.

In our setting like (1.4) the key departure from the usual triangular equations model is that the endogenous regressor \(Y_2\) is measured with error. In this setting, besides the regressor \(Y_2\) itself being mismeasured, the measurement error further complicates the problem because, even if the distribution function \(F_{Y_2|Z}(\cdot|\cdot)\) is known, the control variable obtained by plugging in the error-laden observation \(Y_2^*\) as \(V^* \equiv F_{Y_2|Z}(Y_2^*|Z)\) is also contaminated by the measurement error. Therefore, other existing control function methods – that use the control variables as

\(^5\)Also we develop our estimation allowing for the nonparametric regression. We, however, note that although Hahn and Ridder (2017) develop their estimation for the parametric regression, their approach could potentially be generalized to the nonparametric regression as well.
additional regressors are generally not applicable in our setting. Because in the construction of
the control variable \( V \equiv F_{Y_2|Z}(Y_2|Z) \) here \( Y_2 \) is not observed, this problem can be understood
as the measurement error on the left-side variable. Nevertheless, the identification of the CDF
\( F_{Y_2|Z}(\cdot|\cdot) \) in a pre-stage will suffice for implementing our proposed estimator that utilizes the
moment condition (1.5). We propose to use two alternative approaches to tackle this problem.
One is to recover the CDF using repeated measurements of \( Y_2 \). The other is to recover the CDF
using an instrumental variable for \( Y_2 \).

We note that our approach taken to the model (1.4) is different from other control function
methods because \( V \) is not used as an additional regressor but only plays the role of a conditioning
variable. As a consequence we do not estimate a nonparametric function of \( V \) in the regression
equation. Instead, we estimate conditional density functions of relevant variables given \( V \) to
recover the conditional moment restriction (1.5). In our setting whether or not \( V \) is observable is
irrelevant as long as we recover required conditional density and distribution functions. Finally,
we note that imposing the conditional mean independence of \( Y_2 \) and \( \epsilon \) given \( W \) may serve as
an alternative to other approaches that are assuming either the sufficiency of control variables
\( E[\epsilon|V,Z] = E[\epsilon|V] \) as in Newey, Powell, and Vella (1999) or assuming the conditional moment
condition \( E[\epsilon|Z] = 0 \) as in Newey and Powell (2003) and Ai and Chen (2003). Note that for the
triangular model the sufficiency of control variable \( E[\epsilon|V,Z] = E[\epsilon|V] \) implies the conditional
mean independence because

\[
\]

where the first equality holds due to the law of iterated expectations, the second equality holds
since \( Y_2 \) is a function of \((V,Z)\) only, and the third equality is due to \( E[\epsilon|V,Z] = E[\epsilon|V] \).
From the result above we can also see that the conditional mean independence still holds even
when \( E[\epsilon|V,Z] = E[\epsilon|W] = E[\epsilon|V,Z_1] \) while it violates the sufficiency condition of the usual
control function approach. Note, however, that in general none of these modeling assumptions
including ours implies the other.\(^6\)

Given our identification results we propose a sieve-based method to estimate the parameters.
Our estimation proceeds in two stages. In the first stage we estimate the unknown densities
to recover the conditional moment function using a sieve Maximum Likelihood Estimation
(MLE), and in the second stage we estimate the structural parameters using a sieve Minimum
Distance (MD) estimation. We also provide the asymptotic properties of the estimator in the
supplementary appendix.

We run Monte Carlo simulations to illustrate the finite sample performance of our proposed

\(^6\)Song (2015) considers measurement errors in the setting of Ai and Chen (2003). Because our model is based
on the conditional mean independence, which imposes different forms of identifying restrictions, their methods
are not applicable in our setting.
estimator. We experiment with a partially linear model and an additively-separable nonparametric regression model. In particular we consider different structures of measurement errors and vary their influences. Our proposed estimator shows desirable finite sample behaviors in correcting for both endogeneity and measurement error. On the other hand a conventional instrumental variable estimator, which only corrects for endogeneity, shows considerably large biases.

As an empirical application we estimate the return to education using U.K. twin data. Common concerns in this literature are endogeneity and measurement error in self-reported education. The usual approach to control for measurement error in the twin data has been using the twin’s report as an instrument variable (IV) for the self-reported education. However, this method to correct for measurement error in the conventional IV approach may not be valid if the IV is correlated with the error term in the structural equation. This is because, given that both self-reported and twin’s reported education are measured with errors and both measurement errors are likely to be correlated, the instrument, the twin’s report, also depends on the error term containing the measurement error in the self-reported education. This violates the validity condition of the IV and hence the IV estimator becomes inconsistent. A similar argument holds when the IV approach is used to control for the endogeneity: twin’s reported education as an IV needs to be uncorrelated with other unobserved factors that could affect both education and earnings. Our approach instead utilizes the twin’s report as a control variable to control for both endogeneity and measurement error. This is particularly useful when there exists only one control or instrumental variable available to control for both confounding factors.

Estimation results show that our proposed estimator is useful to correct for both endogeneity and measurement error in education. From a conventional linear wage equation, we find that a one-year increase in education leads to an 11% increase in hourly wage. The estimate is significantly higher than those from OLS and IV approaches which are potentially biased. Since the true relationship between wage and education could be nonlinear, we have also estimated the nonlinear model of education, for which both endogeneity and measurement error problems are further exacerbated. Interestingly, OLS and IV estimates are far from reasonable values, while our estimator is more stable, predicts decreasing marginal returns of education, and is robust to different restrictions on the measurement error.

The outline of the paper is as follows. Section 2 discusses identification in the presence of mismeasured endogenous regressors. Section 3 develops a sieve estimation based on the identification result. We present Monte Carlo simulations and an empirical application in Section 4 and 5. We then conclude in Section 6. Technical results are gathered in the appendix. We also provide the online supplementary appendix for further details and the asymptotic properties of the proposed estimator.
2 Identification Using Control Variables

We introduce notations and further articulate the nature of endogeneity and measurement error. We denote the supports of the distributions of the random variables $Y_1, Y_2, Y_2^*, Z_1$, and $V$ by $\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_2^*$, $\mathcal{Z}_1$, and $\mathcal{V}$, respectively. The joint density of $Y_1$ and $(Y_2, Y_2^*, Z_1, V)$ admits a bounded density with respect to the product measure of some dominating measure $\mu$ on $\mathcal{Y}_1$ and the Lebesgue measure on $\mathcal{Y}_2 \times \mathcal{Y}_2^* \times \mathcal{Z}_1 \times \mathcal{V}$. All marginal and conditional densities are also bounded. For notational ease, let $Y \equiv (Y_1, Y_2) \in \mathcal{Y} \equiv \mathcal{Y}_1 \times \mathcal{Y}_2$, $Y^* \equiv (Y_1, Y_2^*) \in \mathcal{Y}^* \equiv \mathcal{Y}_1 \times \mathcal{Y}_2^*$, $W \equiv (Z_1, V) \in \mathcal{W} \equiv \mathcal{Z}_1 \times \mathcal{V}$. Let $X \equiv (Y, W) \in \mathcal{X} \equiv \mathcal{Y} \times \mathcal{W}$. Suppose that the true observations $\{(Y_i, W_i) : i = 1, 2, ..., n\}$ are independently drawn from the distribution of $(Y, W)$ with support $\mathcal{Y} \times \mathcal{W}$, where $\mathcal{Y}$ is a subset of $\mathcal{R}_d^{\mathcal{Y}}$ and $\mathcal{W}$ is a compact subset of $\mathcal{R}_d^{\mathcal{W}}$. Let $\alpha_0 \equiv (\theta_0, h_0) \in \mathcal{A} \equiv \Theta \times \mathcal{H}$. We assume that $\Theta \subseteq \mathcal{R}_d^{\mathcal{Y}}$ is compact with nonempty interior, and that $\mathcal{H}$ is a space of continuous functions. We use notations $f_{R_1}(r_1)$, $f_{R_1|R_2}(r_1 | r_2)$, and $F_{R_1|R_2}(r_1 | r_2)$ to denote the marginal density of variable $R_1$, the conditional density of $R_1$ conditional on $R_2$, and the cumulative distribution function of $R_1$ conditional on $R_2$, respectively.

2.1 Conditions for Identification

Our identification utilizes the notion of conditional (mean) independence or equivalent exclusion restrictions. Following Dawid (1979a), we will write $\mathcal{A} \perp \mathcal{B} | \mathcal{C}$ to denote that $\mathcal{A}$ is independent of $\mathcal{B}$ being conditioned on $\mathcal{C}$. We first state required conditions for identification using control variables when $Y_2$ is measured with error.

**Assumption 2.1** (i) $E[Y_2 \varepsilon | W] = E[Y_2 | W]E[\varepsilon | W]$; (ii) $(\theta_0, h_0)$ is the only $(\theta, h) \in \Theta \times \mathcal{H}$ satisfying the conditional moment restrictions (1.3).

**Assumption 2.2** $Y_1 \perp Y_2^* | (Y_2, W)$ for all $(Y_1, Y_2, Y_2^*, W) \in \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Y}_2^* \times \mathcal{W}$.

**Assumption 2.3** $Y_1 \perp V | (Y_2, Z_1)$ for all $(Y_1, Y_2, W) \in \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{W}$.

**Assumption 2.4** $Y_2^* \perp V | (Y_2, Z_1)$ for all $(Y_2, Y_2^*, W) \in \mathcal{Y}_2 \times \mathcal{Y}_2^* \times \mathcal{W}$.

Assumption 2.1 ensures identification of the parameters when all true $Y$ are observed and endogeneity is properly controlled. In particular the conditional moment condition (1.3) can arise from Assumption 2.1 (i), the conditional mean independence between the unobserved cause of $Y_1$ and the endogenous regressor $Y_2$ conditional on $W$. Assumption 2.1 (ii), the uniqueness of the true parameters, holds if the set $\{w \in \mathcal{W} : m(w, \theta, h) \neq m(w, \theta_0, h_0)\}$ has positive probability for any $(\theta, h) \neq (\theta_0, h_0) \in \Theta \times \mathcal{H}$ as in the standard conditional moment models. Assumptions 2.2-2.4 state additional conditional independence conditions.
which are equivalent to relevant exclusion restrictions. Several variants of these conditions have been widely used in the econometrics literature (e.g., Altonji and Matzkin (2005), Heckman and Vytlacil (2005), Su and White (2007), Imbens and Newey (2009)). In our setting these assumptions are utilized for recovering the (conditional) densities of true variables from the observed ones. Assumption 2.2 can be equivalently stated in terms of density functions as $f_{y_1|y_2^*,w}(y_1 \mid y_2, y_2^*, w) = f_{y_1|y_2,w}(y_1 \mid y_2, w)$, and Assumption 2.3 is equivalent to $f_{y_1|y_2,z_1,v}(y_1 \mid y_2, z_1, v) = f_{y_1|y_2,z_1}(y_1 \mid y_2, z_1)$. Assumption 2.2 states given the true regressors and the control variables, the mismeasured regressors do not provide further information on the dependent variables. Assumption 2.3 states the control variables do not provide any more information on the dependent variables than the true regressors do. Similarly, Assumption 2.4 can be equivalently written as $f_{y_2^*|y_2,z_1,v}(y_2^* \mid y_2, z_1, v) = f_{y_2^*|y_2,z_1}(y_2^* \mid y_2, z_1)$, which means that given the true regressors, the control variables do not have further information on the mismeasured regressors. We further discuss these identifying assumptions in the supplementary appendix.

It is interesting to note that a sufficient condition for Assumption 2.1 (i) is the conditional independence as

\textbf{Assumption 2.1'(i)} \quad Y_2 \perp \epsilon \mid (Z_1, V) \text{ for all } (Y_2, \epsilon, W) \in \mathcal{Y}_2 \times \mathcal{E} \times \mathcal{W}.

Also Assumption 2.3 is equivalent to Assumption 2.3' below by Lemma 4.2 (i) of Dawid (1979a).

\textbf{Assumption 2.3'} \quad \epsilon \perp V \mid (Z_1, Y_2) \text{ for all } (Y_2, \epsilon, W) \in \mathcal{Y}_2 \times \mathcal{E} \times \mathcal{W}.

From Assumption 2.1'(i) and 2.3' one may conclude that $\epsilon$ is jointly independent of $Y_2$ and $V$. If this were indeed true, then these assumptions would exclude endogeneity of $Y_2$ as well as dependence of $V$ on $\epsilon$. As a result, the assumptions may seem contradictory each other. However, this statement is one of common fallacious arguments. The following lemma from Dawid (1979b) clarifies the implication of Assumptions 2.1'(i) and 2.3'.

\textbf{Lemma 2.1} \quad (p.251, Dawid 1979b) Assumptions 2.1'(i) and 2.3' hold if and only if $\epsilon \perp (Y_2, V) \mid T$ where $T$ is the information in common between $(Y_2, Z_1)$ and $(Z_1, V)$.

Lemma 2.1 states Assumptions 2.1'(i) and 2.3' do not imply $\epsilon$ is jointly independent of $Y_2$ and $V$ but only imply $\epsilon$ is jointly independent of $Y_2$ and $V$ conditioned on the common information between $Y_2$ and $V$.

\footnote{See Dawid (1979b, 1980) for general interpretation of \textit{common information} and also see the supplementary appendix for examples.} It also means as long as $(Y_2, Z_1)$ and $(Z_1, V)$ share common information other than $Z_1$, $\epsilon$ is allowed to depend on $Y_2$ and $V$ while Assumptions 2.1'(i) and 2.3' hold. For illustration consider a simple example without $Z_1$

$$Y_2 = U_a + U_{Y_2}, \varepsilon = U_a + U_{\varepsilon}, V = U_a + U_V,$$
We estimate so identification of the conditional mean function of the conditional density conditional independence condition of Assumption 2.3, we note that two densities error ε covariance conditions by means of density functions. First, note that by Assumption 2.1 (i), we have rearranging both sides, we obtain \[ E[(Y_2 - E[Y_2 | W])\varepsilon | W] = 0 \] or equivalently \[ E[(Y_2 - E[Y_2 | W])(Y_1 - \pi(Z_1)) - G(Y_2; \theta, h)) | W] = 0. \]

Because \[ E[(Y_2 - E[Y_2 | W])\pi(Z_1) | W] = 0 \] since \( Z_1 \) is included in \( W \), we finally obtain \[ E[(Y_2 - E[Y_2 | W])\pi(Z_1) | W] = 0. \]

Let \( \rho(X, \theta, h) \equiv \rho(Y, W, \theta, h) = (Y_2 - E[Y_2 | W])\pi(Z_1) - G(Y_2; \theta, h) \) be the residual function. Let \( \alpha = (\theta, h) \) and let \( m(W, \alpha) \equiv E[\rho(X, \theta, h) | W] \). Then we can cast the above conditional moment restriction into a general form as (1.3). We estimate \( \alpha \) through recovering conditional density functions associated with the unobserved \( Y_2 \) from the observed data. Our main idea is that the conditional mean function (1.3) can be written as an integral form

\[
m(w, \alpha) \equiv E[\rho(X, \theta, h) | W = w] = \int_y \rho(x, \theta, h)f_{Y|W}(y | w)dy,
\]

so identification of the conditional mean function \( m(w, \alpha) \) is obtained, given the identification of the conditional density \( f_{Y|W}(y | w) \) and the residual function \( \rho(X, \theta, h) \). By utilizing the conditional independence condition of Assumption 2.3, we note that two densities \( f_{Y_1|Y_2Z_1}(y_1 | y_2, z_1) \) and \( f_{Y_2|W}(y_2 | w) \) are sufficient for this identification. We further note that the function \( E[Y_2 | W] \) inside the residual can be written as

\[
E[Y_2 | W = w] = \int_{y_2} y_2 f_{Y_2|W}(y_2 | w)dy_2,
\]

8In view of Dawid (1979b), Assumption 2.1′(i) means that \( Y_2 \) does not give extra information on the conditional distribution of \( \varepsilon \) given \( V \) because \( V \) already contains the common factor \( U_a \), so the only additional information from \( Y_2 \) is \( U_{Y_2} \), which is just a random noise. Similarly, Assumption 2.3′ means that \( V \) does not give extra information on the conditional distribution of \( \varepsilon \) given \( Y_2 \) because \( Y_2 \) already contains \( U_a \), so the only additional information from \( V \) is \( U_V \), which is just a random noise.
so that \( f_{Y_2|W}(y_2 \mid w) \) is also sufficient for the identification of the residual function \( \rho(X, \theta, h(\cdot)) \).

Rewriting the conditional mean functions in terms of conditional densities makes it clear that once \( f_{Y_1|Y_2Z_1}(y_1 \mid y_2, z_1) \) and \( f_{Y_2|W}(y_2 \mid w) \) are obtained, we then can estimate \( \alpha_0 \) using the conditional moments. In Section 3, given identification of \( m(w, \alpha) \), we propose a sieve estimator of \( \alpha_0 \) that is obtained by minimizing a sample analogue of the population criterion function

\[
Q(\alpha) \equiv E \left[ m(W_i, \alpha) \right] \left[ \Sigma(W_i) \right]^{-1} m(W_i, \alpha) \]

(2.7)

where \( \Sigma(W) \) denotes a positive-definite weighting matrix.

Therefore our problem reduces to the problem of recovering the two densities \( f_{Y_1|Y_2Z_1}(y_1 \mid y_2, z_1) \) and \( f_{Y_2|W}(y_2 \mid w) \) from observables where \( Y_2 \) is only measured with error. Note that under Assumptions 2.2-2.4, we can express the conditional density of \( Y^* \) given \( W \) as an integral equation

\[
f_{Y^*|W}(y^* \mid w) = \int_{Y_2} f_{Y_1|Y_2Z_1}(y_1 \mid y_2, z_1) f_{Y_2^*|Y_2Z_1}(y_2^* \mid y_2, z_1) f_{Y_2|W}(y_2 \mid w) dy_2. \tag{2.8}
\]

Therefore, the problem of identifying the densities \( f_{Y_1|Y_2Z_1}(y_1 \mid y_2, z_1) \) and \( f_{Y_2|W}(y_2 \mid w) \) becomes the problem of finding the unique solution to the integral equation (2.8) where \( f_{Y^*|W}(y^* \mid w) \) is directly observable from data. We now provide additional conditions that ensure the solution to the integral equation (2.8) is unique. These conditions are similar to those in Hu and Schennach (2008). Let \( R_1, R_2, \) and \( R_3 \) denote random variables with supports \( R_1, R_2, \) and \( R_3, \) respectively. Let \( L_{R_1|R_2R_3} \) denote an integral operator mapping \( g \in G(R_2) \) to \( L_{R_1|R_2R_3}g \in G(R_1) \) for a given \( r_3, \) defined by \( [L_{R_1|R_2R_3}g](r_1) \equiv \int_{R_2} f_{R_1|R_2R_3}(r_1 \mid r_2, r_3) g(r_2) dr_2, \) where \( G(R_j) \) is the corresponding function space with domain \( R_j \) with \( j = 1, 2. \)

**Assumption 2.5** (i) The operators \( L_{Y_2^*|Y_2Z_1} \) and \( L_{Y_2^*|V_2Z_1} \) are one-to-one. (ii) For any \( z_1 \in Z_1 \) and any \( \tilde{y}_2, \tilde{y}_2 \in Y_2, \) the set \( \{ y_2 \in Y_1 : f_{Y_1|Y_2Z_1}(y_1 \mid \tilde{y}_2, z_1) \neq f_{Y_1|Y_2Z_1}(y_1 \mid \tilde{y}_2, z_1) \} \) has positive probability whenever \( \tilde{y}_2 \neq \tilde{y}_2. \) (iii) For any given \( z_1 \in Z_1, \) there exists a known functional \( \mathcal{M} \) such that \( \mathcal{M}[f_{Y_2^*|Y_2Z_1}(\cdot \mid y_2, z_1)] = y_2 \) for all \( y_2 \in Y_2. \)

Assumption 2.5 (i) states completeness of the family of distributions associated with the operators \( L_{Y_2^*|Y_2Z_1} \) and \( L_{Y_2^*|V_2Z_1} \) (see Newey and Powell (2003) and Blundell, Chen, and Kristensen (2007) for related discussions). This can be regarded as a nonparametric rank condition. Assumption 2.5 (ii) excludes constant distribution of \( Y_1 \) at different values of \( Y_2. \) Assumption 2.5 (iii) places restrictions on some measure of the location of a density, denoted by \( \mathcal{M}[\cdot] \) such as the mean, mode, and quantiles. For example, it reduces to a familiar form \( \mathcal{E}[Y_2^*|Y_2, Z_1] = Y_2 \) in the classical measurement error. Given these identifying assumptions the following theorem states the true parameters in the moment condition (1.3) are identified.
Theorem 2.1  Under Assumptions 2.1-2.5, the parameters $\alpha_0 \equiv (\theta_0, h_0)$ are uniquely identified from the observables $(Y_1, Y_2^*, Z_1, V)$.

2.3 Triangular Simultaneous Equations with Conditional Mean Independence

We further discuss our approach to the triangular model (1.4), which also generates the moment condition (1.3). There are several interesting features to note in our framework. First, the reduced form equation provides a source from which the control variate $V$ can be obtained. Second, even though we know what should be the control variable (e.g. the conditional distribution of $Y_2$ given $Z$, $F_{Y_2|Z}$), a plug-in method to obtain the control variable such as $V \equiv F_{Y_2|Z}(y_2 | Z)$ is not feasible because $Y_2$ is only measured with error. In the appendix we show that in the presence of measurement error in $Y_2$, the CDF $F_{Y_2|Z}$ can be still obtained from the data $(Y_2^*, Z)$ by utilizing auxiliary data such as repeated measurements of $Y_2^*$ or existence of additional instruments. Moreover, in our setting the identification of $F_{Y_2|Z}$ suffices for recovering the structural parameters although we do not recover individual observations of $V$. This is because we need only the estimates of the relevant density functions to approximate the population criterion function (2.7) in our estimation. Whether or not individual observations of the control variable are available is not important in our approach.

3 Estimation

Based on our identification results we propose a sieve-based estimator of $\alpha_0 \equiv (\theta_0, h_0)$ for the model (1.3). First we focus on the case for which the control variables $V$ are observed and then show how the setting can extend to the case for which $V$ are generated variables as in the triangular model (1.4). In our approach, from observations on $Y_1$, $Y_2^*$, $Z_1$, and $V$, we first estimate the unknown densities $f_{Y_1|Y_2Z_1}(y_1 | y_2, z_1)$ and $f_{Y_2|W}(y_2 | w)$ in the conditional moment function (2.6) using a sieve MLE, and in the second stage we estimate parameters of interest $(\theta_0, h_0)$ using a sieve MD estimation.

We introduce additional notations. Let

$$f_{V*|W}(y^* | w; \beta_0) = \int_{y_2} f_{Y_1|Y_2Z_1}(y_1 | y_2, z_1)f_{Y_2^*|Y_2Z_1}(y_2^* | y_2, z_1)f_{Y_2|W}(y_2 | w)dy_2$$

and let $\beta_0 \equiv (f_1, f_2, f_3) \in B \equiv F_1 \times F_2 \times F_3$ denote the first stage parameters such that $f_1 \equiv f_{Y_1|Y_2Z_1}(y_1 | y_2, z_1)$, $f_2 \equiv f_{Y_2^*|Y_2Z_1}(y_2^* | y_2, z_1)$, and $f_3 \equiv f_{Y_2|W}(y_2 | w) = f_{Y_2|VZ_1}(y_2 | v, z_1)$. Define a weighted Hölder ball of radius $c$ by $\Lambda^\gamma_\omega(V) \equiv \{g \in \Lambda^\gamma_\omega(V) : \|g\|_\Lambda^\gamma_\omega \leq c < \infty\}$, where $\Lambda^\gamma_\omega(V)$ is the weighted Hölder space of order $\gamma > 0$ with a weight function $\omega$ (see Ai and
Chen (2003) and Chen, Hong, and Tamer (2005) for more details and examples). We impose restrictions on the parameter spaces $\mathcal{F}_1, \mathcal{F}_2,$ and $\mathcal{F}_3$. We take $\gamma_1 > 1$ below.

**Assumption 3.6** (i) $f_1 \in \mathcal{F}_1 \subset \Lambda_{c_1}^1(\mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Z}_1)$, $\int_{\mathcal{Y}_1} f_1 dy_1 = 1$ for all $y_2 \in \mathcal{Y}_2, z_1 \in \mathcal{Z}_1$.  
(ii) $f_2 \in \mathcal{F}_2 \subset \Lambda_{c_2}^1(\mathcal{Y}_2 \times \mathcal{Y}_2 \times \mathcal{Z}_1)$, $\int_{\mathcal{Y}_2} f_2 dy_2 = 1$ for all $y_2 \in \mathcal{Y}_2, z_1 \in \mathcal{Z}_1$. (iii) $f_3 \in \mathcal{F}_3 \subset \Lambda_{c_3}^1(\mathcal{Y}_2 \times \mathcal{V} \times \mathcal{Z}_1)$, $\int_{\mathcal{Y}_2} f_3 dy_2 = 1$ for all $v \in \mathcal{V}, z_1 \in \mathcal{Z}_1$.

Let $p_j$ denote a sequence of known basis functions (such as power series, splines, etc.). A tensor-product multivariate linear sieve basis, denoted by $p^l(\cdot, \cdot) = (p_1(\cdot, \cdot), \ldots, p_l(\cdot, \cdot))'$ is used to approximate functions of three variables. Suppose we have $n$ observations of the sample $\{y_{1i}, y_{2i}^*, z_{1i}, v_i\}$. Then based on the identification results of Section 2.2, we estimate $\beta_0$ using the sieve MLE as

$$\hat{\beta}_n = (\hat{f}_1n, \hat{f}_2n, \hat{f}_3n)$$

$$= \arg \max_{(f_1, f_2, f_3) \in \mathcal{B}_n} \frac{1}{n} \sum_{i=1}^n \ln \int_{\mathcal{Y}_2} f_1(y_{1i} \mid y_2, z_{1i}) f_2(y_{2i}^* \mid y_2, z_{1i}) f_3(y_2 \mid v_i, z_{1i}) dy_2,$$

where $\mathcal{B}_n = \mathcal{F}_1n \times \mathcal{F}_2n \times \mathcal{F}_3n$ is a sieve space approximating $\mathcal{B}$ with the sample size $n$, and where $\mathcal{F}_1n, \mathcal{F}_2n$ and $\mathcal{F}_3n$ denote sieve spaces for $\mathcal{F}_1, \mathcal{F}_2$ and $\mathcal{F}_3$, respectively. Using these estimated densities we construct $\hat{m}(w_i, \alpha)$ as the plug-in estimator of $m(w_i, \alpha)$:

$$\hat{m}(w_i, \alpha) \equiv \int_{\mathcal{Y}_2} \left[ \int_{\mathcal{Y}_1} \rho(y_1, y_2, z_{1i}, v_i, \theta, h) \hat{f}_{Y_{1|Y_2}Z_{1}}(y_1 \mid y_2, z_{1i}) dy_1 \right] \hat{f}_{Y_2|VZ_1}(y_2 \mid v_i, z_{1i}) dy_2$$

where $\hat{f}_{1n} = \hat{f}_{Y_{1|Y_2}Z_{1}}(y_1 \mid y_2, z_{1i})$ and $\hat{f}_{3n} = \hat{f}_{Y_2|VZ_1}(y_2 \mid v_i, z_{1i})$ are obtained from (3.9). Let $\mathcal{H}$ be a space of smooth functions (e.g. Hölder ball) that contains the true $h_0$ and let $\mathcal{H}_n$ be some finite-dimensional approximation space that becomes dense in $\mathcal{H}$ as the sample size $n$ tends to infinity. Then in the second stage, we obtain the (penalized) sieve MD estimator of $\alpha_0 \equiv (\theta_0, h_0)$ as

$$\hat{\alpha}_n = \arg \inf_{\alpha \in \Theta \times \mathcal{H}_n} \hat{Q}_n(\alpha) = \left\{ \frac{1}{n} \sum_{i=1}^n \hat{m}(W_i, \alpha)' [\hat{\Sigma}(W_i)]^{-1} \hat{m}(W_i, \alpha) + \lambda_n \hat{P}_n(h) \right\}$$

where $\hat{\Sigma}(W)$ is a consistent estimator of $\Sigma(W)$ (a positive-definite weighting matrix), $\lambda_n \geq 0$ is a penalization tuning parameter such that $\lambda_n = o(1)$, and $\hat{P}_n(h) \geq 0$ is a possibly random penalty function as in Chen and Pouzo (2012).\(^9\)\(^10\)

---

\(^9\)It is worth noting that if the original parameter space $\mathcal{B}$ is too large, it is useful to introduce another penalty function for the first-stage parameter, $\beta$, and estimate it via a penalized sieve MLE as in Shen (1997). For the sake of concise results, we maintain the assumption that $\mathcal{B}$ is a compact space.

\(^10\)Alternatively we can consider an approximate sieve MD estimator by adding a positive real-valued sequence,
Our estimation method can accommodate the case when $V$ are generated variables. For the model (1.4) we can estimate the densities using a sieve MLE as

$$
\hat{\beta}_n = (\hat{f}_{1n}, \hat{f}_{2n}, \hat{f}_{3n})
$$

where

$$
= \arg \max_{(f_1, f_2, f_3) \in \mathcal{B}_n} \frac{1}{n} \sum_{i=1}^{n} \ln \int_{y_2} f_1(y_{1i} \mid y_2, z_{1i}) f_2(y^*_2 \mid y_2, z_{1i}) f_3(y_2 \mid \hat{v}_i(y_2), z_{1i}) dy_2,
$$

where we define $\hat{v}_i(y_2)$ as an estimator of $F_{y_2|z}(y_2|z_i)$. We provide examples of the CDF estimator in the appendix.

Also, for the triangular model we can estimate the conditional mean function as

$$
\hat{m}(\hat{w}_i, \alpha) \\
\equiv \int_{y_2} \left[ \int_{y_1} \rho(y_1, y_2, z_{1i}, \hat{v}_i(y_2), \theta, h) \hat{f}_{1|y_2|Z_1}(y_1 \mid y_2, z_{1i}) dy_1 \right] \hat{f}_{2|VZ_1}(y_2 \mid \hat{v}_i(y_2), z_{1i}) dy_2
$$

where $\hat{f}_{1n} = \hat{f}_{1|y_2|Z_1}(y_1 \mid y_2, z_{1i})$ and $\hat{f}_{3n} = \hat{f}_{2|VZ_1}(y_2 \mid \hat{v}_i(y_2), z_{1i})$ are obtained from (3.11). The sieve MD estimator is then obtained by replacing $\hat{m}(w_i, \alpha)$ with $\hat{m}(\hat{w}_i, \alpha)$ in the construction of the sample criterion function $\hat{Q}_n(\alpha)$ in (3.10).

The supplementary appendix provides large sample properties of our estimator. There we first derive the consistency, and obtain the convergence rate and the asymptotic normality of the finite-dimensional parameters estimator following Newey and Powell (2003), Hu and Schennach (2008), and Chen and Pouzo (2009, 2012). We focus on the case for which the control variable is observable and then show how the results can extend to the case for which the control variable is a generated variable.

4 Simulations

We run Monte Carlo simulations to study the finite sample performance of the proposed estimator in a few different settings. We also provide practical details to implement our estimator in this section.

which asymptotically converges to zero, to the sample objective function (3.10) as in Chen and Pouzo (2012). This extension will be straightforward with additional notation.
4.1 Partially Linear Model

We consider a data generating process from the following partially linear model:

\[
\begin{align*}
Y_1 &= \pi(Z_1) + \theta Y_2 + \varepsilon, \\
Y_2 &= \phi_1 Z_1 + \phi_2 V + \nu, \\
\varepsilon &= \delta V + \varpi,
\end{align*}
\]

where \(Y_1\) is the dependent variable, \(Z_1\) is an exogenous covariate drawn from \(N(0,0.5^2)\), and \(V\) is a control variable drawn from \(N(0,0.5^2)\), and where \(\nu\) and \(\varpi\) are mutually independent innovations drawn from \(N(0,0.25^2)\) and \(N(0,0.5^2)\), respectively. In this DGP the common information between \(Y_2\) and \(V\) is the \(V\) itself and given \(V\), \(\varepsilon\) is independent of \((Y_2, V)\) and hence by Lemma 2.1 Assumptions 2.1 (i) and 2.3 are satisfied.

The nonparametric function is specified as \(\pi(\cdot) = \exp(\cdot)\). \(Y_2\) is an endogenous and unobserved regressor and researchers observe only its mismeasured counterpart \(Y_2^* = Y_2 + \sigma_e \exp(-Y_2)\cdot e\) where \(e\) is an measurement error and \(\sigma_e\) is its standard deviation. We consider three different structures of measurement error as follows. Design A is a non-additive error with zero median such that \(e = \ln(-\ln(1-U))\) where \(U\) is a uniformly distributed random variable over \([0,1]\) support, Design B is a heteroskedastic error with zero mean as \(e = N(0,1)\), and lastly Design C is a non-additive error with zero median such that \(e = \ln(\omega + \sqrt{\omega^2 + 2})\) where \(\omega = -0.5 + \tan(\pi U - 0.5)/\exp(-Y_2)\). Coefficients in this model are set as \(\theta = 1.5\), \(\phi_1 = 1\), \(\phi_2 = 1.5\), and \(\delta = 0.5\). For each design we also vary the size of standard deviation \(\sigma_e\) by 0.5, 1, and 1.5. In all of these designs given \(Y_2, Y_2^*\) becomes independent of \(Y_1\) and \(V\) and hence Assumptions 2.2 and 2.4 are satisfied.

We compare the finite sample performance of the proposed estimator with two other sieve IV estimators; infeasible estimator using the true \(Y_2\) as a benchmark and inconsistent estimator using mismeasured \(Y_2^*\). These two sieve IV estimators control for the endogeneity of \(Y_2\) using a set of instruments which is a tensor product polynomial sieve of order 3: \(P_i \equiv (1, Z_{1i}, V_i, Z_{1i}^2, Z_{1i}V_i, V_i^2, \ldots, V_i^3)'\). For instance, in order to construct the infeasible estimator, the term \(R \equiv Y_2 - E[Y_2 \mid W]\) is estimated as the regression residual of \(Y_2\) on \(P\). Then the estimator of \(\theta\) is obtained by taking the weighted regression of \(Y_1\) on \(Y_2\) treating \(R\) as the weight:

\[
\hat{\theta}_{\text{infeasible}} = \left( \sum_{i=1}^{n} \Psi_i P'_i \left( \sum_{i=1}^{n} P_i P'_i \right)^{-1} \sum_{i=1}^{n} P_i \Psi_i' \right)^{-1} \sum_{i=1}^{n} \Psi_i P'_i \left( \sum_{i=1}^{n} P_i P'_i \right)^{-1} \sum_{i=1}^{n} P_i \tilde{Y}_{1i}
\]

where \(\tilde{Y}_{1i} \equiv \hat{R}_i \times Y_{1i}, \Psi_i \equiv \hat{R}_i \times Y_{2i}\), and \(\hat{R}_i \equiv Y_{2i} - \hat{E}[Y_2 \mid W = W_i]\). The inconsistent estimator is constructed by replacing the true \(Y_2\) with the mismeasured \(Y_2^*\).
To implement the proposed estimator, in the first stage, approximating sieves for functions of three variables using tensor product bases of univariate trigonometric series are employed to approximate the densities \( f_{Y_1|Y_2Z_1} \) and \( f_{Y_2^*|Y_2Z_1} \). Following Hu and Schennach (2008), the sieve approximations are given by

\[
\begin{align*}
    f_{Y_1|Y_2Z_1}(y_1 \mid y_2, z_1) & \approx \sum_{j_1=0}^{j_{1n}} \sum_{j_2=0}^{j_{2n}} \sum_{j_3=0}^{j_{3n}} \gamma_{j_1j_2j_3} u_{j_1}(y_1 - y_2) u_{j_2}(y_2) u_{j_3}(z_1), \\
    f_{Y_2^*|Y_2Z_1}(y_2^* \mid y_2, z_1) & \approx \sum_{j_1=0}^{j_{1n}} \sum_{j_2=0}^{j_{2n}} \sum_{j_3=0}^{j_{3n}} \theta_{j_1j_2j_3} u_{j_1}(y_2^* - y_2) u_{j_2}(y_2) u_{j_3}(z_1)
\end{align*}
\]

where \( u_{j_1}(\cdot) \) is a sine or cosine function, and \( u_{j_2}(\cdot) \) and \( u_{j_3}(\cdot) \) are cosine functions. By utilizing properties of the trigonometric series, the identification restriction on \( f_{Y_2^*|Y_2Z_1} \) in Assumption 2.5 can be easily imposed. In addition, it is easy to impose integral of each density over its support being equal to one. The density \( f_{Y_2|VZ_1} \) is specified as a normal density to ease high dimensionality of the nonparametric specification. In the first stage, we estimate \( f_{Y_1|Y_2Z_1} \) and \( f_{Y_2|VZ_1} \) using the sieve MLE as in (3.9). In the second stage, the estimate of the weight \( R \equiv Y_2 - E[Y_2 \mid W] \) is constructed by \( \tilde{R}(Y_2, V, Z_1) = Y_2 - \int_{Y_2} y_2 \tilde{f}_{Y_2|VZ_1}(y_2 \mid V, Z_1)dy_2 \). Then, the estimator of \( \theta \) is obtained by taking the weighted least squares regression of \( Y_1 \) on \( Y_2 \) from the MD estimation (3.10) such that

\[
    \hat{\theta}_{\text{proposed}} = (\sum_{i=1}^{n} \tilde{Y}_{2i} \tilde{\gamma}'_{2i})^{-1} \sum_{i=1}^{n} \tilde{Y}_{2i} \tilde{Y}_{1i}
\]

(4.12)

where we let \( \tilde{Y}_{1i} \equiv \int_{Y_2} \int_{Z_1} \tilde{R}(y_2, V_i, Z_{1i}) y_1 \tilde{f}_{Y_1|Y_2Z_1}(y_1 \mid y_2, Z_{1i}) \tilde{f}_{Y_2|VZ_1}(y_2 \mid V_i, Z_{1i})dy_1dy_2 \) and \( \tilde{Y}_{2i} \equiv \int_{Y_2} \tilde{R}(y_2, V_i, Z_{1i}) y_2 \tilde{f}_{Y_2|VZ_1}(y_2 \mid V_i, Z_{1i})dy_2 \).

In both stages, we adopt a Gauss-Hermite quadrature for numerical integrals because the supports of \( Y_1 \) and \( Y_2 \) are potentially unbounded. The proposed estimator does not require numerical optimization in the second stage since it has a closed-form solution. In addition, it is not necessary to estimate the function \( \pi(\cdot) \) of exogenous covariates \( Z_1 \) when researchers are primarily interested in estimating the effect of \( Y_2 \) on \( Y_1 \).

We investigate the finite sample performance of the three estimators by calculating the squared bias (SB), variance (VAR), and mean squared error (MSE). We vary the size of \( \sigma_e \) by 0.5, 1, and 1.5. By doing so, we can investigate how different severeness of measurement error affects behaviors of the estimators. We set the sample size, \( n = 1,000 \) and repeat each experiment 200 times.

Table 1 reports the results, indicating the proposed estimator outperforms the inconsistent IV estimator. For example, when the measurement error is non-additive with zero mode and \( \sigma_e = 0.5 \), SB of the proposed estimator is 0.0188 which is close to that of the infeasible estimator, 0.0034. However, SB from the inconsistent IV estimator is significantly larger as it becomes

11\footnote{We take \( \Sigma = \Sigma = I \) (identity matrix) and \( \lambda_n = 0 \) for our experiments.}

15
2.1440. Because the proposed estimator is a semiparametric two-step estimator that requires more flexible approximation in the first stage while the other two estimators are based on least squares, the proposed estimator tends to produce larger variances. Nevertheless, MSE of the proposed estimator, 0.4813, is much smaller than 2.1457 from the IV estimator. Similar patterns are found across larger standard deviations of the measurement error. For the proposed estimator we have experimented with different numbers of sieve terms in the first stage to minimize the MSE. We find using more sieve terms generally reduces biases while increasing variances. We also find, when we fix the number of sieve terms, biases become larger but variances tend to go down as the standard deviation of the measurement error \( \sigma_e \) increases. For conciseness, we only report the estimation results which tend to minimize MSE by balancing bias and variance. With larger \( \sigma_e \) this MSE criterion allows us to use more flexible sieve terms in the first stage to reduce biases while only slightly increasing variances. The finite sample behaviors of the three estimators are similar in other structures of measurement error.

### 4.2 Additively-separable Nonparametric Model

We use the following additively-separable nonparametric model for our experiments

\[
Y_1 = h(Y_2) + \pi(Z_1) + \varepsilon,
\]
\[
Y_2 = \phi_1 Z_1 + \phi_2 V + \nu,
\]
\[
\varepsilon = \delta V + \varpi,
\]

where \( h(\cdot) = -\frac{\exp(\cdot)}{1+\exp(\cdot)} \) and \( \pi(\cdot) = 2 \sin(\cdot) \). All other variables and parameters are the same with the design in Section 4.1. The function \( h \) is the primary parameter of interest. As a sieve basis for \( h \), we use a power series of fourth order multiplied by the standard normal CDF. We report performances of the three estimators over different structures of measurement error as in Section 4.1. The finite-sample performances are evaluated over different sample sizes, \( n \in \{500, 1000\} \). We repeat each experiment for 200 times.

Table 2 and Table 3 report the integrated squared bias (ISB), integrated variance (IVAR), and integrated mean squared error (IMSE) of the \( h(\cdot) \) estimate, which are computed using numerical integral over a grid from \(-2\) to \(2\). From the estimation results we find that in all designs the proposed estimator outperforms the inconsistent IV estimator. For example, when the measurement error is non-additive with zero mode (Design A), \( \sigma_e = 0.5 \), and \( n = 1,000 \), ISB from the proposed estimator is 0.0299 which is close to that of the infeasible estimator, 0.0102 while ISB of the inconsistent IV estimator is significantly larger as 2.93. In terms of IMSE our proposed estimator clearly outperforms the inconsistent IV estimator. For the same setting above, IMSE of the proposed estimator is 0.1202 while that of the inconsistent estimator is 47.42. For other designs we observe that the finite sample behaviors of the estimators are all
similar and the results are comparable across larger standard deviations of measurement error. Finally, the performance of the proposed estimator improves as the sample size increases from 500 to 1,000. We also report graphs of estimated functions from the three estimators along with the true function in Figure G1-G12 for all designs of measurement error.

5 Empirical application

We estimate the return to education on earnings using twin data. Measuring the return of education is one of the most studied areas in the literature.\textsuperscript{12} The large volume of this research reflects the importance of the causal effect of education on earnings as well as inherent difficulties in measuring this effect. There are two confounding factors that researchers want to control for. First, education measures are frequently measured with error, particularly if the information is collected through one-time retrospective surveys, which are susceptible to recall errors (see, e.g., Ashenfelter and Krueger (1994), Kane, Rouse, and Staiger (1999), Bound, Brown, and Mathiowetz (2001)). Second, unobserved factors such as ability are potentially related to both educational level and earnings.

A strand of this literature utilizes twin data to extract information on repeated measurements of education where one twin is asked to report both his/her own schooling and also that of the other twin (see, e.g., Ashenfelter and Krueger (1994)). A common approach to correct for measurement error using twin data is to use the twin’s report as an instrument variable (IV) for the self-reported education. However, this method to correct for the measurement error in the conventional IV approach is still subject to endogeneity problem, if there exist other unobserved factors that could affect both education and earnings. This is because, given that both self-reported and twin’s reported education are mismeasured, it is likely that the instrument, the twin’s report, also depends on the error term in the structural equation, and hence becomes an invalid IV. We illustrate how our approach can be used instead in this setting to correct for both confounding factors. Our approach treats education as a continuous variable following the literature (e.g., Garen (1984), Card (2001), Florens, Heckman, Meghir and Vytlacil (2008), and Cunha, Heckman, Schennach (2010)).

We use a data set on female monozygotic twins from the Twins Research Unit, St. Thomas’ Hospital from the United Kingdom. The data are taken from Bonjour et al. (2003). The sample consists of 428 individuals comprising 214 identical twin pairs with complete wage, age, and schooling information. In Bonjour et al. (2003), to measure schooling, they use each twin’s report for their education qualification and their twin’s qualification. These qualifications were split into 12 groups, and they assign years of education to each qualification. These estimated years could be another source of measurement error. For the wage variable Bonjour et al.

\textsuperscript{12}For example, see Card (1999) for surveys on comprehensive list of studies on this topic.
(2003) ask twins to report normal earnings before taxes and deductions, and their usual hours worked, and convert the wage data into an hourly rate.

Given the data we consider the wage equation

\[ \log w_{if} = \pi(\text{age}_{if}) + \beta S_{if} + A_{if} + \varepsilon_{if} \]  

(5.13)

where \( w_{if} \) denotes the wage of twin \( i \) in family \( f \), \( \text{age}_{if} \) is the age of twin \( i \), \( S_{if} \) is the true years of schooling, \( A_{if} \) is “ability” broadly defined, and \( \varepsilon_{if} \) is an idiosyncratic wage shock for \( i = 1, 2 \). We let \( S_{if}^* \) and \( S_{if}'' \) be the self-reported education and the other twin’s report for the education of twin \( i \), respectively, such that \( S_{if}^* = S_{if} + \nu_{if} \) and \( S_{if}'' = S_{if} + \nu'_{if} \) where \( \nu_{if} \) and \( \nu'_{if} \) denote the measurement errors. In estimating the return to schooling parameter \( \beta \), Bonjour et al. (2003) instrument the self-reported education \( S_{if}^* \) with reported level of the other twin \( S_{if}'' \) to correct for measurement error. Then, to control for the endogeneity due to the unobserved ability, they also estimate the within-pair differenced equation for identical twins

\[ \log w_{1f} - \log w_{2f} = \beta(S_{1f} - S_{2f}) + (a_{1f} - a_{2f}) + (\varepsilon_{1f} - \varepsilon_{2f}) \]

where \( a_{if} \) denotes ability net of family and genetic effects. Following Ashenfelter and Krueger (1994), they instrument the reported schooling differences \( (S_{1f}^* - S_{2f}^*) \) with differences based on reports from the other twin \( (S_{1f}'' - S_{2f}'') \). Comparing their pooled OLS/IV and within-pair OLS/IV estimates they argue that measurement error appears to bias both pooled and within-pair OLS estimates downwards while it appears ability biases the pooled OLS estimate upwards. We note that for this within-pair IV approach to be valid the instrument should be uncorrelated with the term \( a_{1f} - a_{2f} \) such that

\[ E[(S_{1f}'' - S_{2f}'')(a_{1f} - a_{2f})] = E[(S_{1f} - S_{2f})(a_{1f} - a_{2f})] + E[(\nu'_{1f} - \nu'_{2f})(a_{1f} - a_{2f})] = 0 \]

which may or may not hold depending on the nature of the omitted ability and measurement errors (see Neumark (1999)).\(^\text{13}\) If this covariance \( E[(S_{1f}'' - S_{2f}'')(a_{1f} - a_{2f})] \) is negative, then the within-pair IV estimator will be biased downwards, even when the true schooling \( S_{if} \) is positively correlated with \( A_{if} \), which would imply the usual positive ability bias of the pooled OLS and IV estimators. To compare with their approach, our approach uses the whole sample without the differencing and instead utilizes the twin’s report as a control variable to control for both endogeneity and measurement error in the self-reported education. Using the whole sample without differencing could be an advantage because, as Bonjour et al. (2003) note, the within-pair results are statistically less significant due to the fact that the differencing could

\(^\text{13}\) In the ideal case \( a_{1f} = a_{2f} \), so the differencing would completely remove the observed ability if the omitted ability is identical between the twins.
effectively dispense with significant variation in the variables. We estimate a linear model and a nonlinear model (quadratic in years of schooling) and compare our proposed estimator with other estimators. Our proposed estimator does not need to estimate $\pi(age)$ while for other pooled estimators we specify it as a quadratic function of $age$ in (5.13). The main identifying assumption for our estimator is that the “true” schooling is conditionally mean independent with ability, being conditioned on reported level schooling of the other twin. This allows ability to be correlated with the reported education of the other twin, which is likely to happen if the true education is correlated with ability. We implement our estimator following the steps described in our Monte Carlos as (4.12).

We report the results in Table A (a) for the linear model. We find both OLS estimates (column (1) and (3)) underestimate the return due to measurement error compared to the IV estimates while our estimator (column (5-7)) yields still higher estimates than both pooled and within-pair IV estimates (column (2) and (4)). In particular our estimates suggest that the bias is negative when compared to other estimators. Indeed our estimates are close to the return to education estimate 0.110 (in Table 5 of Bonjour et al. (2003)) using smoking at ages 16 as an instrument but smaller than the IV estimate 0.122 of Evans and Montgomery (1994). Our estimator is robust to different location normalization of measurement error (zero mean, zero median, and zero mode).

Another interesting departure is observed when we estimate the nonlinear model of education, for which both endogeneity and measurement error problem is further exacerbated.\(^{14}\) The existence of valid instruments is not sufficient for consistent estimation in nonlinear models with measurement error, even in the absence of endogeneity (see e.g. Griliches and Ringstad (1970)).

---

**Table A. Returns to Education Estimation using Twin’s data**

|          | OLS Pooled | IV Within pair | OLS Pooled | IV Within pair | Proposed
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Zero Mode</td>
</tr>
<tr>
<td>(a) Educ</td>
<td>0.077***</td>
<td>0.087***</td>
<td>0.039*</td>
<td>0.077**</td>
<td>0.1070***</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.017)</td>
<td>(0.023)</td>
<td>(0.033)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>(b) Educ</td>
<td>0.371**</td>
<td>0.021</td>
<td>0.1435*</td>
<td>0.1479*</td>
<td>0.1438*</td>
</tr>
<tr>
<td></td>
<td>(0.182)</td>
<td>(0.321)</td>
<td>(0.080)</td>
<td>(0.082)</td>
<td>(0.081)</td>
</tr>
<tr>
<td>Educ(^2)</td>
<td>-1.053</td>
<td>0.238</td>
<td>-0.310</td>
<td>-0.331</td>
<td>-0.309</td>
</tr>
<tr>
<td></td>
<td>(0.652)</td>
<td>(1.15)</td>
<td>(0.316)</td>
<td>(2.414)</td>
<td>(0.309)</td>
</tr>
<tr>
<td>obs</td>
<td>428</td>
<td>428</td>
<td>214</td>
<td>214</td>
<td>428</td>
</tr>
</tbody>
</table>

**Note:** Standard errors are in parentheses. For the proposed estimator the standard errors are calculated using bootstrap with 200 repetitions. The panel (a) is the linear model and the panel (b) is the nonlinear model. *Significant at the 10-percent level.**Significant at the 5-percent level.***Significant at the 1-percent level.
wage equation becomes

\[ \log w_{if} = \pi(\text{age}_{if}) + \beta_1 S_{if} + \beta_2 S_{if}^2 + A_{if} + \varepsilon_{if}. \]

When the self-reported education is observed as \( S_{if}' = S_{if} + \nu_{if} \), the equation can be rewritten as

\[ \log w_{if} = \pi(\text{age}_{if}) + \beta_1 S_{if}' + \beta_2 (S_{if}' - \nu_{if})^2 + A_{if} + \varepsilon_{if} - \beta_1 \nu_{if}. \]

It is worth noting that the measurement error \( \nu_{if} \) is not separable from the observed education \( S_{if}' \) due to the quadratic function. As a result, the conventional IV approach which may use \( S_{if}' \) and \( (S_{if}')^2 \) as IV's for \( S_{if}' \) and \( (S_{if}')^2 \) does not resolve the issue of measurement error. By construction the measurement error inside the quadratic function requires an additional treatment. Nevertheless, our proposed approach allows for the measurement error in the nonlinear model. Table A (b) presents the estimated coefficients, \( \beta_1 \) and \( \beta_2 \). The results show that both pooled OLS and IV estimates are a bit off and they diverge to the opposite direction compared to our estimator.\(^{15}\) In particular the pooled IV estimator even predicts accelerating effects of education (the coefficient on education squared is positive). On the other hand our estimator predicts decreasing marginal returns of education and the results are robust to different location normalization of measurement error.

6 Conclusion

We study identification and estimation of regression functions for a class of semiparametric models for which the endogenous regressors are measured with errors. For these models we utilize the existence of control variables such that endogenous regressors and unobservable causes are conditionally mean independent given the control variables. Monte Carlo simulations illustrate that our proposed estimator performs well in the finite samples. In an empirical application to U.K. twin data, the proposed estimator obtains consistent estimates of the return to education, which are robust to different identification conditions on the measurement error in education.

\(^{15}\)We do not report the within-pair OLS and IV estimators for the nonlinear model because they do not produce reliable results.
Appendix

A Proof of the Identification Theorem (Theorem 2.1)

By the definition of conditional expectation, we obtain

\[ m(w, \theta, h) = \int_y \rho(x, \theta, h)f_{Y|W}(y \mid w)dy \]

\[ = \int_y (y_2 - E[Y_2 \mid W = w])(y_1 - G(y_2; \theta, h))f_{Y|W}(y \mid w)dy. \]

From Assumption 2.1, \( \alpha_0 \equiv (\theta_0, h_0) \in \Theta \times \mathcal{H} \) is the unique solution for the equation \( m(w, \theta, h) = 0 \). Thus the identification of \( f_{Y|W}(y \mid w) \) and the identification of \( \rho(x, \theta, h) \) given \( \alpha \) are sufficient for the identification of the parameter \( \alpha_0 \) through the moment equation \( m(w, \theta, h) = 0 \). First, for the identification of \( \rho(x, \theta, h) \equiv (y_2 - E[Y_2 \mid W = w])(y_1 - G(y_2; \theta, h)) \) we need to recover \( E[Y_2 \mid W] \) from the observables. We note that this conditional mean function inside the residual can be written as

\[ E[Y_2 \mid W = w] = \int_{y_2} y_2 f_{Y_2|W}(y_2 \mid w)dy_2, \]

so that \( f_{Y_2|W}(y_2 \mid w) \) is sufficient for the identification of \( E[Y_2 \mid W] \). Second, for the identification of \( f_{Y|W}(y \mid w) \), we use the fact that \( f_{Y|W}(y \mid w) = f_{Y_1|Y_2W}(y_1 \mid y_2, w)f_{Y_2|W}(y_2 \mid w) = f_{Y_1|Y_2Z_1}(y_1 \mid y_2, z_1)f_{Y_2|W}(y_2 \mid w) \) where the first equality holds by the definition of the conditional density and the second equality holds by Assumption 2.3. Thus the identification of \( m(w, \theta, h) \) is obtained by identifying the two density functions \( f_{Y_1|Y_2Z_1}(y_1 \mid y_2, z_1) \) and \( f_{Y_2|W}(y_2 \mid w) \). We use a similar argument to Hu and Schennach (2008). By Assumptions 2.2-2.4, we have the following integral equation

\[ f_{Y^*|W}(y^* \mid w) = \int_{y_2} f_{Y^*Y_2|W}(y^*, y_2 \mid w)dy_2 \]

\[ = \int_{y_2} f_{Y_1|Y_2^*Y_2W}(y_1 \mid y_2^*, y_2, w)f_{Y_2^*|Y_2W}(y_2^* \mid y_2, w)dy_2 \]

\[ = \int_{y_2} f_{Y_1|Y_2^*Y_2W}(y_1 \mid y_2^*, y_2, w)f_{Y_2^*|Y_2W}(y_2^* \mid y_2, w)f_{Y_2|W}(y_2 \mid w)dy_2 \]

\[ = \int_{y_2} f_{Y_1|Y_2Z_1}(y_1 \mid y_2, z_1)f_{Y^*_2|Y_2Z_1}(y_2^* \mid y_2, z_1)f_{Y_2|W}(y_2 \mid w)dy_2, \]

where the last equality holds by Assumptions 2.2-2.4. Recall that \( R_1, R_2, \) and \( R_3 \) denote random variables with supports \( \mathcal{R}_1, \mathcal{R}_2, \) and \( \mathcal{R}_3, \) respectively, and \( L_{R_1|R_2R_3} \) denote an integral operator mapping \( g \in \mathcal{G}(\mathcal{R}_2) \) to \( L_{R_1|R_2R_3}g \in \mathcal{G}(\mathcal{R}_1) \) for a given \( r_3 \) defined by \( [L_{R_1|R_2R_3}g](r_1) \equiv \int_{R_2} f_{R_1|R_2R_3}(r_1 \mid r_2, r_3)g(r_2)dr_2, \) where \( \mathcal{G}(\mathcal{R}_j) \) is the corresponding function space with domain
functions, we obtain the conditional moment function served regressors moment restrictions have the unique solution as measurements (for similar methods, see Li and Vuong (1998), Schennach (2004), Delaigle, Hall, Assume we observe two repeated measurements of \( Y \) only measured with error for the triangular model (1.4). One is using repeated measurements We provide two approaches to estimate the conditional distribution of \( Y \) is identified. This completes the proof.

**B Examples of Control Variable for Triangular Models**

We provide two approaches to estimate the conditional distribution of \( Y \) given \( Z \) when \( Y \) is only measured with error for the triangular model (1.4). One is using repeated measurements of \( Y \) with errors and the other is using an additional instrument for the mismeasured \( Y \).

**B.1 Using Repeated Measurements**

Assume we observe two repeated measurements of \( Y \) as \( Y_{2a} = Y_2 + e_a \) and \( Y_{2b}^* = Y_2 + e_b \) where \( e_a \) and \( e_b \) are measurement errors. Then \( F_{Y_2|Z}(y_2 \mid z) \) can be identified from the repeated measurements (for similar methods, see Li and Vuong (1998), Schennach (2004), Delaigle, Hall,
As a result, we get (other equivalent results) allows us to recover values of the control variable for any given values $Y$ of the second measurement error and the true value. Here the second measurement error has conditional mean zero given $V$ such that it follows a uniform distribution over $V$. Then $F_{Y_2|Z}(y_2 \mid z)$ is identified from the observables $(Y_{2a}^*, Y_{2b}^*, Z)$. In particular we obtain

$$F_{Y_2|Z}(y_2 \mid z) = \frac{1}{2} + \frac{1}{2\pi} \int_0^\infty \frac{E[\exp(-i\zeta Y_2) \mid Z = z] \exp(i\zeta y_2) - E[\exp(i\zeta Y_2) \mid Z = z] \exp(-i\zeta y_2)}{i\zeta} \, d\zeta$$

where $E[\exp(i\zeta Y_2) \mid Z] = E[\exp(i\zeta Y_{2a}^*)] \exp \left( \int_0^\zeta \frac{iE[\exp(i\zeta Y_{2a}^*)]}{E[\exp(i\zeta Y_{2a}^*)]} \, d\zeta \right)$.

Note that in Lemma B.2 the assumption $E[e_a \mid Y_{2b}^*] = 0$ states the first measurement error has conditional mean zero given $Y_{2b}^*$. The assumption $e_b \perp (Y_2, Z)$ implies the independence of the second measurement error and the true value. Here the second measurement error does not need to have mean zero. It allows for systematic drift term in the second measurement $Y_{2b}^*$, which may be a useful property when panel data is used for estimation. Lemma B.2 (or other equivalent results) allows us to recover values of the control variable for any given values of $(y_2, z)$ when $Y_2$ is only measured with error. In particular we can estimate $F_{Y_2|Z}(y_2 \mid z)$ as a sample analogue to (B.14) by approximating the (conditional) expectations with corresponding sample (conditional) means.

### B.2 Proof of Lemma B.2

Proof of Lemma B.2 (i): By a similar argument to Matzkin (2003), we have

$$F_{V}(v) = F_{V}(r^{-1}(z, y_2)) = P(V \leq r^{-1}(Z, y_2) \mid Z = z) = P(r(Z, V) \leq y_2 \mid Z = z) = F_{Y_2|Z}(y_2 \mid z)$$

from the monotonicity of $r(Z, V)$ in $V$ and the independence of $Z$ and $V$. Then by normalizing $V$ such that it follows a uniform distribution over $[0, 1]$ as $V = F_{V}(V)$, we obtain the result.

Proof of Lemma B.2 (ii): Denoting the support of $Z$ by $Z$, we note $E[\exp(i\zeta Y_2) \mid Z = z] = \int f_{Y_2|Z}(y_2 \mid z) \exp(i\zeta y_2) \, dy_2$ is the Fourier transform of $f_{Y_2|Z}(y_2 \mid z)$ and that $\frac{1}{2\pi} \int E[\exp(i\zeta Y_2) \mid Z = z] \exp(-i\zeta y_2) \, d\zeta$ is the inverse Fourier transform of $E[\exp(i\zeta Y_2) \mid Z = z]$ for $(y_2, z) \in Y_2 \times Z$. As a result, we get

$$f_{Y_2|Z}(y_2 \mid z) = \frac{1}{2\pi} \int E[\exp(i\zeta Y_2) \mid Z = z] \exp(-i\zeta y_2) \, d\zeta.$$
Then the inversion theorem (e.g., Gurland (1948)) provides the conditional CDF of $Y_2$ given $Z = z$ as (B.14). We now show the identification of $E[\exp(i\zeta Y_2) \mid Z = z]$. From (B.14) it is clear that the identification of $E[\exp(i\zeta Y_2) \mid Z = z]$ suffices to recover the CDF, $F_{Y_2|Z}(y_2 \mid z)$. First observe that

$$\int_0^\zeta \frac{iE[Y_{2a}^* \exp(i\zeta Y_{2b})]}{E[\exp(i\zeta Y_{2b})]} \, d\zeta = \int_0^\zeta \frac{iE[(Y_2 + e_a) \exp(i\zeta Y_{2b})]}{E[\exp(i\zeta Y_{2b})]} \, d\zeta$$

$$= \int_0^\zeta \frac{iE[Y_2 \exp(i\zeta (Y_2 + e_b))] + iE[e_a \exp(i\zeta Y_{2b})]}{E[\exp(i\zeta (Y_2 + e_b))]} \, d\zeta$$

$$= \int_0^\zeta \frac{iE[Y_2 \exp(i\zeta (Y_2 + e_b))] + iE[E(e_a \exp(i\zeta Y_{2b}) \mid Y_{2b})]}{E[\exp(i\zeta (Y_2 + e_b))]} \, d\zeta$$

$$= \int_0^\zeta \frac{iE[Y_2 \exp(i\zeta (Y_2 + e_b))]}{E[\exp(i\zeta (Y_2 + e_b))]} \, d\zeta = \int_0^\zeta \frac{iE[Y_2 \exp(i\zeta Y_2)]E[\exp(i\zeta e_b)]}{E[\exp(i\zeta Y_2)]E[\exp(i\zeta e_b)]} \, d\zeta$$

$$= \int_0^\zeta \frac{iE[Y_2 \exp(i\zeta Y_2)]}{E[\exp(i\zeta Y_2)]} \, d\zeta = \int_0^\zeta \frac{\partial}{\partial \zeta} \ln(E[\exp(i\zeta Y_2)]) \, d\zeta$$

$$= \int_0^\zeta \left( \frac{\partial}{\partial \zeta} \ln(E[\exp(i\zeta Y_2)]) - \ln 1 \right) \, d\zeta = \ln(E[\exp(i\zeta Y_2)])$$

where the law of iterated expectation is used in the third equality, $E[e_a \mid Y_{2b}] = 0$ is used in the fifth equality, $e_b \perp Y_2$ is used in the sixth equality, and $\ln 1 = 0$ is used in the ninth equality. Thus we get $E[\exp(i\zeta Y_2)] = \exp \left( \int_0^\zeta \frac{iE[Y_{2a}^* \exp(i\zeta Y_{2b})]}{E[\exp(i\zeta Y_{2b})]} \, d\zeta \right)$. Further observe that from $e_b \perp Y_2 \mid Z$ (which is implied by $e_b \perp (Y_2, Z)$)

$$E[\exp(i\zeta Y_2) \mid Z] = \frac{E[\exp(i\zeta Y_2) \mid Z]E[\exp(i\zeta Y_2)]E[\exp(i\zeta e_b)]}{E[\exp(i\zeta Y_2)]E[\exp(i\zeta e_b)]}$$

$$= \frac{E[\exp(i\zeta Y_2) \mid Z]E[\exp(i\zeta e_b) \mid Z]}{E[\exp(i\zeta Y_{2b})]} E[\exp(i\zeta Y_2)]$$

$$= \frac{E[\exp(i\zeta (Y_2 + e_b)) \mid Z]}{E[\exp(i\zeta Y_{2b})]} E[\exp(i\zeta Y_2)]$$

$$= \frac{E[\exp(i\zeta Y_{2b}) \mid Z]}{E[\exp(i\zeta Y_{2b})]} \exp \left( \int_0^\zeta \frac{iE[Y_{2a}^* \exp(i\zeta Y_{2b})]}{E[\exp(i\zeta Y_{2b})]} \, d\zeta \right)$$

where the right-hand side is a function of all observables, which implies the identification of $E[\exp(i\zeta Y_2) \mid Z]$. This completes the identification result for $F_{Y_2|Z}(y_2 \mid z)$ through (B.14).
B.3 Using Instrumental Variables

Suppose we observe instrumental variables $U$ for the unobservable $Y_2$. The variables $U$ can be thought as some observables other than $Z$ that are correlated with $Y_2$ but are independent with the measurement error given $Y_2$. Then using a similar argument in Theorem 2.1, $F_{Y_2|Z}$ can be identified from the observables $(Y_2^*, Z, U)$. To see this we rewrite the CDF as $F_{Y_2|Z}(y_2 \mid z) \equiv \int_{-\infty}^{y_2} f_{Y_2|Z}(\tilde{y}_2 \mid z)d\tilde{y}_2$ where $f_{Y_2|Z}(y_2 \mid z) = \frac{f_{Y_2Z}(y_2, z)}{f_Z(z)}$. Then since $f_Z(z)$ is identified from the data, the identification of $F_{Y_2|Z}(y_2 \mid z)$ rests on the identification of $f_{Y_2Z}(y_2, z)$. We state conditions similar to Assumptions 2.2-2.5 for the identification of $f_{Y_2Z}(y_2, z)$.

Assumption B.7 $Z \perp Y_2^* \mid (Y_2, U)$ for all $(U, Y_2, Y_2^*, Z) \in U \times \mathcal{Y}_2 \times \mathcal{Y}_2^* \times \mathcal{Z}$.

Assumption B.8 $Z \perp U \mid Y_2$ for all $(U, Y_2, Z) \in U \times \mathcal{Y}_2 \times \mathcal{Z}$.

Assumption B.9 $Y_2^* \perp U \mid Y_2$ for all $(Y_2, Y_2^*, U) \in \mathcal{Y}_2 \times \mathcal{Y}_2^* \times U$.

Assumption B.10 (i) The operators $L_{Y_2^*|U}$ and $L_{Y_2^*|Y_2}$ are one-to-one. (ii) For any $\tilde{y}_2, \bar{y}_2 \in \mathcal{Y}_2$, the set \{ $z \in \mathcal{Z} : f_{Z|Y_2}(z \mid \tilde{y}_2) \neq f_{Z|Y_2}(z \mid \bar{y}_2)$ \} has positive probability whenever $\tilde{y}_2 \neq \bar{y}_2$. (iii) There exists a known functional $\tilde{M}$ such that $\tilde{M}[f_{Y_2|Y_2^*}(\cdot \mid y_2)] = y_2$ for all $y_2 \in \mathcal{Y}_2$.

The following lemma provides an identification of $f_{Y_2Z}(y_2, z)$ which is sufficient for the identification of the CDF $F_{Y_2|Z}(y_2 \mid z)$ given that $f_Z(z)$ is observable from the data.

Lemma B.3 Under Assumptions B.7-B.10, the densities $(f_{Y_2Z}, f_{Y_2^*|Y_2}, f_{U|Y_2})$ are uniquely identified from the observables $(Y_2^*, Z, U)$.

B.4 Proof of Lemma B.3

For the identification of $f_{Y_2Z}(y_2, z)$, we use a similar argument to Theorem 2.1. By Assumptions B.7-B.9, we have the integral equation $f_{Z|Y_2^*|U}(z, y_2^* \mid u) = \int_{y_2} f_{Z|Y_2}(z \mid y_2)f_{Y_2^*|Y_2}(y_2^* \mid y_2)f_{Y_2|U}(y_2 \mid u)dy_2$. Then from the proof of Theorem 1 of Hu and Schennach (2008), we note that the densities $(f_{Z|Y_2}, f_{Y_2^*|Y_2}, f_{Y_2|U})$ are uniquely identified from the observable joint density $f_{Z|Y_2^*|U}(z, y_2^* \mid u)$ by Assumption B.10. Then because $f_{Y_2Z}(y_2) = \int f_{Y_2|U}(y_2 \mid u)f_U(u)du$, $f_{Y_2|Z}(y_2 \mid y_2) = f_{Y_2|U}(y_2 \mid u)f_U(u)/f_{Y_2}(y_2)$, $f_{Y_2Z}(y_2, z) = f_{Z|Y_2}(z \mid y_2)f_{Y_2}(y_2)$, and $f_U(u)$ is directly observable from data, we conclude the densities $(f_{Y_2Z}, f_{Y_2^*|Y_2}, f_{U|Y_2})$ are uniquely identified from the observables $(Y_2^*, Z, U)$.
Figure G1: Design A, n=1,000 and s.d. = 0.5

Figure G2: Design B, n=1,000 and s.d. = 0.5

Figure G3: Design C, n=1,000 and s.d. = 0.5

Figure G4: Design A, n=1,000 and s.d. = 1

Figure G5: Design B, n=1,000 and s.d. = 1

Figure G6: Design C, n=1,000 and s.d. = 1
Figure G7: Design A, n=1,000 and s.d. = 1.5

Figure G8: Design B, n=1,000 and s.d. = 1.5

Figure G9: Design C, n=1,000 and s.d. = 1.5

Figure G10: Design A, n=500 and s.d. = 1

Figure G11: Design B, n=500 and s.d. = 1

Figure G12: Design C, n=500 and s.d. = 1
<table>
<thead>
<tr>
<th>Estimator</th>
<th>ME Structure</th>
<th>Design A</th>
<th>Design B</th>
<th>Design C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zero Mode</td>
<td>Zero Mean</td>
<td>Zero Median</td>
<td></td>
</tr>
<tr>
<td>Infeasible</td>
<td>Squared Bias</td>
<td>0.0034</td>
<td>0.0034</td>
<td>0.0034</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>0.0055</td>
<td>0.0055</td>
<td>0.0055</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.0089</td>
<td>0.0089</td>
<td>0.0089</td>
</tr>
<tr>
<td>Proposed</td>
<td>Squared Bias</td>
<td>0.0188</td>
<td>0.0262</td>
<td>0.0093</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>0.4625</td>
<td>0.4499</td>
<td>0.4899</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.4813</td>
<td>0.4761</td>
<td>0.4992</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>Squared Bias</td>
<td>2.1440</td>
<td>2.1920</td>
<td>2.2200</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>0.0017</td>
<td>0.0042</td>
<td>0.0022</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>2.1457</td>
<td>2.1962</td>
<td>2.2222</td>
</tr>
<tr>
<td>s.d. of ME</td>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Design A</th>
<th>Design B</th>
<th>Design C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infeasible</td>
<td>Squared Bias</td>
<td>0.0034</td>
<td>0.0034</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>0.0055</td>
<td>0.0055</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.0089</td>
<td>0.0089</td>
</tr>
<tr>
<td>Proposed</td>
<td>Squared Bias</td>
<td>0.0050</td>
<td>0.0246</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>0.4875</td>
<td>0.4588</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.4925</td>
<td>0.4834</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>Squared Bias</td>
<td>2.2050</td>
<td>2.2350</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>0.0005</td>
<td>0.0011</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>2.2055</td>
<td>2.2361</td>
</tr>
<tr>
<td>s.d. of ME</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Design A</th>
<th>Design B</th>
<th>Design C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infeasible</td>
<td>Squared Bias</td>
<td>0.0034</td>
<td>0.0034</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>0.0055</td>
<td>0.0055</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.0089</td>
<td>0.0089</td>
</tr>
<tr>
<td>Proposed</td>
<td>Squared Bias</td>
<td>0.0002</td>
<td>0.0234</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>0.5155</td>
<td>0.4878</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.5157</td>
<td>0.5112</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>Squared Bias</td>
<td>2.2220</td>
<td>2.2430</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>0.0002</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>2.2222</td>
<td>2.2435</td>
</tr>
<tr>
<td>s.d. of ME</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2

Estimation of $h_0$ in the additively-separable model ($n = 500$)

<table>
<thead>
<tr>
<th>Estimator</th>
<th>ME Structure</th>
<th>Design A</th>
<th>Design B</th>
<th>Design C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Zero Mode</td>
<td>Zero Mean</td>
<td>Zero Median</td>
</tr>
<tr>
<td>Infeasible</td>
<td>ISB</td>
<td>0.0034</td>
<td>0.0034</td>
<td>0.0034</td>
</tr>
<tr>
<td></td>
<td>IVAR</td>
<td>0.0881</td>
<td>0.0881</td>
<td>0.0881</td>
</tr>
<tr>
<td></td>
<td>IMSE</td>
<td>0.0915</td>
<td>0.0915</td>
<td>0.0915</td>
</tr>
<tr>
<td>Proposed</td>
<td>ISB</td>
<td>0.0346</td>
<td>0.0434</td>
<td>0.0346</td>
</tr>
<tr>
<td></td>
<td>IVAR</td>
<td>0.1499</td>
<td>0.1822</td>
<td>0.1354</td>
</tr>
<tr>
<td></td>
<td>IMSE</td>
<td>0.1845</td>
<td>0.2256</td>
<td>0.1700</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>ISB</td>
<td>2.0700</td>
<td>0.8595</td>
<td>0.0732</td>
</tr>
<tr>
<td></td>
<td>IVAR</td>
<td>53.1300</td>
<td>10.9300</td>
<td>6.5100</td>
</tr>
<tr>
<td></td>
<td>IMSE</td>
<td>55.2000</td>
<td>11.7900</td>
<td>6.5800</td>
</tr>
<tr>
<td>s.d. of ME</td>
<td></td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimator</th>
<th></th>
<th>Design A</th>
<th>Design B</th>
<th>Design C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infeasible</td>
<td>ISB</td>
<td>0.0034</td>
<td>0.0034</td>
<td>0.0034</td>
</tr>
<tr>
<td></td>
<td>IVAR</td>
<td>0.0881</td>
<td>0.0881</td>
<td>0.0881</td>
</tr>
<tr>
<td></td>
<td>IMSE</td>
<td>0.0915</td>
<td>0.0915</td>
<td>0.0915</td>
</tr>
<tr>
<td>Proposed</td>
<td>ISB</td>
<td>0.0326</td>
<td>0.0355</td>
<td>0.0340</td>
</tr>
<tr>
<td></td>
<td>IVAR</td>
<td>0.1690</td>
<td>0.1552</td>
<td>0.1372</td>
</tr>
<tr>
<td></td>
<td>IMSE</td>
<td>0.2016</td>
<td>0.1907</td>
<td>0.1712</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>ISB</td>
<td>1.7600</td>
<td>0.3273</td>
<td>0.2002</td>
</tr>
<tr>
<td></td>
<td>IVAR</td>
<td>112.3000</td>
<td>12.5000</td>
<td>9.5900</td>
</tr>
<tr>
<td></td>
<td>IMSE</td>
<td>114.0600</td>
<td>12.8300</td>
<td>9.7900</td>
</tr>
<tr>
<td>s.d. of ME</td>
<td></td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimator</th>
<th></th>
<th>Design A</th>
<th>Design B</th>
<th>Design C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infeasible</td>
<td>ISB</td>
<td>0.0034</td>
<td>0.0034</td>
<td>0.0034</td>
</tr>
<tr>
<td></td>
<td>IVAR</td>
<td>0.0881</td>
<td>0.0881</td>
<td>0.0881</td>
</tr>
<tr>
<td></td>
<td>IMSE</td>
<td>0.0915</td>
<td>0.0915</td>
<td>0.0915</td>
</tr>
<tr>
<td>Proposed</td>
<td>ISB</td>
<td>0.0356</td>
<td>0.0369</td>
<td>0.0342</td>
</tr>
<tr>
<td></td>
<td>IVAR</td>
<td>0.2160</td>
<td>0.1491</td>
<td>0.1244</td>
</tr>
<tr>
<td></td>
<td>IMSE</td>
<td>0.2516</td>
<td>0.1860</td>
<td>0.1587</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>ISB</td>
<td>4.0800</td>
<td>0.7860</td>
<td>0.1974</td>
</tr>
<tr>
<td></td>
<td>IVAR</td>
<td>146.3000</td>
<td>22.8500</td>
<td>13.8500</td>
</tr>
<tr>
<td></td>
<td>IMSE</td>
<td>150.3800</td>
<td>23.6400</td>
<td>14.0500</td>
</tr>
<tr>
<td>s.d. of ME</td>
<td></td>
<td>1.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3

Estimation of $h_0$ in the additively-separable model ($n = 1,000$)

<table>
<thead>
<tr>
<th>Estimator</th>
<th>ME Structure</th>
<th>Design A</th>
<th>Design B</th>
<th>Design C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Zero Mode</td>
<td>Zero Mean</td>
<td>Zero Median</td>
</tr>
<tr>
<td>Infeasible</td>
<td>ISB</td>
<td>0.0102</td>
<td>0.0102</td>
<td>0.0102</td>
</tr>
<tr>
<td></td>
<td>IVAR</td>
<td>0.0459</td>
<td>0.0459</td>
<td>0.0459</td>
</tr>
<tr>
<td></td>
<td>IMSE</td>
<td>0.0561</td>
<td>0.0561</td>
<td>0.0561</td>
</tr>
<tr>
<td>Proposed</td>
<td>ISB</td>
<td>0.0299</td>
<td>0.0300</td>
<td>0.0300</td>
</tr>
<tr>
<td></td>
<td>IVAR</td>
<td>0.0903</td>
<td>0.0909</td>
<td>0.0859</td>
</tr>
<tr>
<td></td>
<td>IMSE</td>
<td>0.1202</td>
<td>0.1209</td>
<td>0.1159</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>ISB</td>
<td>2.9300</td>
<td>0.4569</td>
<td>0.0441</td>
</tr>
<tr>
<td></td>
<td>IVAR</td>
<td>44.5000</td>
<td>6.6400</td>
<td>2.7300</td>
</tr>
<tr>
<td></td>
<td>IMSE</td>
<td>47.4300</td>
<td>7.1000</td>
<td>2.7700</td>
</tr>
<tr>
<td>s.d. of ME</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Design A  | Design B  | Design C  |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Infeasible</td>
<td>ISB</td>
<td>0.0102</td>
</tr>
<tr>
<td></td>
<td>IVAR</td>
<td>0.0459</td>
</tr>
<tr>
<td></td>
<td>IMSE</td>
<td>0.0561</td>
</tr>
<tr>
<td>Proposed</td>
<td>ISB</td>
<td>0.0304</td>
</tr>
<tr>
<td></td>
<td>IVAR</td>
<td>0.0890</td>
</tr>
<tr>
<td></td>
<td>IMSE</td>
<td>0.1194</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>ISB</td>
<td>0.7639</td>
</tr>
<tr>
<td></td>
<td>IVAR</td>
<td>28.9700</td>
</tr>
<tr>
<td></td>
<td>IMSE</td>
<td>29.7300</td>
</tr>
<tr>
<td>s.d. of ME</td>
<td>1.0</td>
<td></td>
</tr>
</tbody>
</table>

Design A   | Design B   | Design C   |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Infeasible</td>
<td>ISB</td>
<td>0.0102</td>
</tr>
<tr>
<td></td>
<td>IVAR</td>
<td>0.0459</td>
</tr>
<tr>
<td></td>
<td>IMSE</td>
<td>0.0561</td>
</tr>
<tr>
<td>Proposed</td>
<td>ISB</td>
<td>0.0299</td>
</tr>
<tr>
<td></td>
<td>IVAR</td>
<td>0.0856</td>
</tr>
<tr>
<td></td>
<td>IMSE</td>
<td>0.1155</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>ISB</td>
<td>0.9545</td>
</tr>
<tr>
<td></td>
<td>IVAR</td>
<td>26.8600</td>
</tr>
<tr>
<td></td>
<td>IMSE</td>
<td>27.8100</td>
</tr>
<tr>
<td>s.d. of ME</td>
<td>1.5</td>
<td></td>
</tr>
</tbody>
</table>

30
References


