Nonseparable Triangular Simultaneous Equations Models with Unobserved Variables *

Suyong Song[†]

March 2017

Abstract

This paper studies nonparametric identification and estimation of nonseparable triangular simultaneous equations models when true endogenous variables are unobservable because of measurement errors. We provide identification of various structural parameters such as average structural functions and quantile structural effects, based on instrumental variables and control variables approaches. We show how the identification results can be extended to the case of the unobservable true endogenous variables, using repeated measurements of the true endogenous variables. Our identification strategy is constructive in the sense that it produces explicit forms of the identified parameters and suggests corresponding nonparametric estimation strategies from these closed-form expressions. Uniform convergence rates of the proposed estimators are provided. We apply the estimators to study nonseparable consumer demand system using the Panel Study of Income and Dynamics data. In many commodities, the proposed estimators obtain significantly different magnitudes and shapes of the average structural functions than methods only controlling for endogeneity in family expenditure. The estimated outcomes support that taking into account both endogeneity and measurement errors in family expenditure is substantial in analyzing the consumer demand systems.

Keywords: Nonseparable models, endogeneity, measurement errors, instrumental variables, control variables, unobserved heterogeneity, quantile effects, average derivative.

JEL Classification: C14, C31, C36

*The author would like to thank Stephane Bonhomme, Yongsung Chang, Thomas Crossley, Kirill Evdokimov, Sergio Firpo, Antonio Galvao, Marc Henry, Shakeeb Khan, Kyoo il Kim, Lorenz Kueng, Tong Li, Konrad Menzel, Alexandre Poirier, Jeffrey Racine, Christoph Rothe, Susanne M. Schennach, Ehsan Soofi, Yixiao Sun, Ben Williams, Ke-Li Xu, and seminar participants at George Washington University, Iowa State University, University of Connecticut, University of Iowa, University of Wisconsin-Madison, Midwest Econometrics Group 2014, and the 11th World Congress of the Econometric Society for their helpful comments and suggestions.

[†]Department of Economics, University of Iowa, W360 Pappajohn Business Building, 21 E Market St, Iowa City, IA 52242. Email: `suyong-song@uiowa.edu`

1 Introduction

1.1 Objectives and Outline

We consider identification and estimation of nonseparable triangular simultaneous equations:

$$Y = g(X, U) \tag{1}$$

$$X = h(Z, \eta), \tag{2}$$

where Y is dependent variable, X is a vector of endogenous regressors, U and η are unobserved heterogeneity, Z is a vector of instrumental variables satisfying $Z \perp\!\!\!\perp (U, \eta)$, and $g(\cdot)$ and $h(\cdot)$ are unknown functions. The parameters of interest are average structural functions (Blundell and Powell (2003)) and quantile structural effects (Imbens and Newey (2009)). The true endogenous regressors X are allowed to be observed with measurement errors to researchers. We obtain identification of the parameters under mild restrictions on dimensions of the unobserved heterogeneity, shapes of the unknown functions, and measurement errors on X in the nonseparable triangular simultaneous equations models.

Dealing with unobservable true endogenous variables has been of interest in many empirical and theoretical economic models. Because of complexity of economic relations, many economic variables are endogenously selected and they are even unobservable to economists because of measurement errors. In particular, when nonlinear parametric or nonparametric simultaneous equations models with measurement errors are considered, we need a great extent of treatments for the unobservable true endogenous variables due to measurement errors. This is because distribution of the unobservable true variables need to be recovered from the mismeasured observables.[1] As a result, there are presently few studies that simultaneously accommodate both the endogeneity and measurement errors in the nonlinear models, even though failure to properly accommodate these features leads to potentially serious errors in inference.[2]

Our main contribution is to provide a general approach to identification and estimation of the structural parameters in the nonseparable triangular simultaneous equations models in the presence of the true endogenous regressors observed with measurement errors. Our analysis introduces theory and methods that simultaneously accommodate both the endogeneity and measurement

[1] In linear parametric simultaneous equations models with measurement errors, Griliches (1974) and Hausman (1977) note that further presence of measurement errors on endogenous variables poses no difficulty beyond that posed by the endogeneity itself, as long as these measurement errors are uncorrelated with instrumental variables. On the other hand, in nonlinear models such as polynomial regression models, correcting for both endogeneity and measurement errors in the same variables is not as straightforward as in the linear model. In fact, Amemiya (1985) and Hsiao (1989) point out that, even in the absence of the endogeneity, the existence of valid instruments is not sufficient to control for the measurement errors.

[2] For instance, because of complications embedded and lack of comprehensive methods in the nonlinear parametric or nonparametric models, most of previous studies of returns to schooling and of the Engel curve, where measurement errors on the endogenous variables are pervasive, address at most one of these issues.

errors under weaker conditions than what have been required in the literature. We also show that how restrictions on the dimensions of the unobserved heterogeneity and the shapes of the unknown functions (g and h) provide a great deal of informational contents for the identification.

In identifying the structural parameters such as the average structural functions and the quantile structural effects, we build upon the nonlinear measurement error literature to deal with the unobservable true endogenous variables. The innovation in this paper over previous studies is to provide a unified framework covering instrumental variables and control variables approaches in the nonseparable model where the reduced-form equation (2) has one-dimensional unobserved heterogeneity but the structural equation (1) has either multi-dimensional heterogeneity or one-dimensional heterogeneity. We show how the identification results of Matzkin (2003), Imbens and Newey (2009), and Kasy (2014) can be extended to our model with the measurement errors. We use a convenient property of the Fourier transform which relates unobservables to observables. We provide identification of various structural parameters based on quantities of a general form which are identified from observables. It is shown that two observed measurements of the unobservable true endogenous variables are sufficient to recover the structural parameters such as the average structural function and quantile structural effects. In particular, we relax full independence conditions and show that less restrictive conditions such as mean independence can still deliver identification results in the general nonseparable models.[3]

The presence of the unobservable true endogenous regressors in the nonseparable triangular simultaneous equations is non-standard issue for many reasons. In the reduced-form equation (2), dependent variable is the unobservable true endogenous regressors X. Since the instrumental variables Z and the unobserved individual heterogeneity η are nonseparable, the latter cannot merely capture the measurement error on the dependent variables X. As a result, the unobserved heterogeneity in the reduced-form equation (2) is not one-dimensional anymore due to the measurement errors. Allowing for such possibility has its own merit in many economic examples. In particular, identification strategies in many economic models such as nonadditive hedonic models (Heckman, Matzkin, and Nesheim (2010)), production function models (Olley and Pakes (1996)), and demand models (Berry, Levinsohn, and Pakes (1995)) rest on shape restrictions such as monotonicity or invertibility of functions of interest in a particular variable or unobserved individual heterogeneity. To make them operational, one-dimensional unobserved heterogeneity is crucial. When true endogenous regressors are unobservable but are replaced with their observed proxies, these approaches fail to achieve point identification due to multi-dimensionality of the unobserved heterogeneity.

Even after the true endogenous regressors in the structural equation (1) are "exogenized" by the

[3]In fact, even though we consider both the endogeneity and measurement errors in the nonseparable models (1)–(2), conditions on the measurement errors in this paper are weaker than those in Hausman, Newey, Ichimura, and Powell (1991), Schennach (2004) and Song, Schennach, and White (2015), in that we allow for the true X to depend on the measurement errors on the repeated measurements. Contrast to our condition, these papers exclude the case that the first measurement error is correlated with the true X and the second measurement error, when two repeated measurements on the true X are observed.

instrumental variables or control variables, they are still unobservable in identifying the structural parameters. Thus, an additional treatment on the measurement errors is necessary, but it is in a highly nonlinear way because of the nonseparability of the structural equation. Lastly, when distributional effects such as the quantile structural effects are particularly of interest, nonsmoothness of the problem-solving is inevitable. So it is not clear that an identification strategy for the average structural function in the presence of the measurement errors is straightforwardly conveyed to the distributional effects.

To achieve point identification of the quantile structural function and average structural function in both instrumental variables and control variables approaches, a nonparametric rank condition is necessary. When instrumental variables vary insufficiently in an empirical application so that the rank condition is too strong to be satisfied, we show that one can still bound the structural effects. Namely, partial identification results for the quantile structural function and average structural function are established when the support of Z is not large enough.

Based on the identification results, we propose nonparametric estimators which are easily implementable thanks to closed-form expressions of the identified parameters. Because denominators of the estimators are associated with a conditional characteristic function which converges to zero as frequency goes to infinity by Riemann-Lebesgue lemma (Lukacs (1970)), it is natural to have so-called ill-posed inverse problem (Fan (1991)) or irregular identification (Khan and Tamer (2010)). We rectify the issue by adopting compactly-supported Fourier transform of a kernel function and derive uniform convergence rates of the estimators. The rates depend on relative tail behaviors of distributions and functions involved. As shown in Monte Carlo experiments, the proposed estimators perform well in estimating the structural parameters in the presence of the unobservable true endogenous regressors.

We apply the proposed estimators to the consumer demand using the Panel Study of Income and Dynamics data. We estimate average structural functions for six major commodities such as food, electricity, utilities, gasoline, health, and leisure in 2009. It is well-known that family expenditure in Engel curves is endogenous and mismeasured. Family income of 2009 is used as an instrumental variable to control for the endogeneity, and two years of the family expenditure (2009 and 2011) are selected to control for the measurement errors. Estimation results show that in most commodities, the proposed estimators obtain significantly different magnitudes and shapes of the average structural functions over family expenditure than estimators which suffer from the measurement errors. The results support that it is of great importance to correct for both the endogeneity and measurement errors in family expenditure in estimating nonseparable consumer demand systems.

It is worthwhile noting that instrumental variables approach such as Hu and Schennach (2008) could be adopted to control for the measurement errors under a set of different conditions, instead of repeated measurements approach. Since, however, the former requires an optimization routine in estimating parameters, we focus on the repeated measurements approach which provides closed-form solutions.

The rest of the article is organized as follows. Section 1.2 reviews the related literature. Section 2 presents the nonseparable triangular simultaneous equations models and motivates our models with an economic example of consumer demand systems. Sections 3–4 consider identification of the structural parameters, respectively, with the control variables approach and instrumental variables approach, when the unobserved heterogeneity in the structural equation is multi-dimensional. Section 5 presents identification when the unobserved heterogeneity in the structural equation is one-dimensional. In Section 6, we apply the proposed methods to study the nonseparable consumer demand systems. Section 7 summarizes and discusses possible extensions. All proofs are included in the Appendix.[4]

The online Supplementary Material contains identification of densities of the measurement errors and sketches identification results for the case of multi-dimensional variables and the case of multiple repeated measurements. It also proposes estimators of objects of interest, establishes their asymptotic properties, and presents finite-sample behaviors of the estimators via Monte Carlo simulations and more results on the empirical application to the consumer demand systems.

1.2 Related literature

This article contributes to the extensive literature on nonseparable models with endogenous regressors and nonlinear models with measurement errors. We summarize a selected set of papers in the two literature for brevity.

Researchers have previously imposed linearity or separability on systems of structural equations because of resulting ease of interpretation and implementation. But more realistic models of the economic behavior do not necessarily exhibit these convenient features. When these simplifying assumptions fail, serious errors of inference on the models may result. To overcome such difficulties, researchers have devoted increasing attention to relaxing some or all of these assumptions. In particular, identification and estimation of models with the nonseparable structural equations in the presence of observable endogenous regressors have been extensively studied under various sets of assumptions. Imbens and Newey (2009) allow for heterogeneity in the structural equation to be multi-dimensional but heterogeneity in the reduced-form equation to be one-dimensional. Their identification strategy relies on monotonicity of the reduced-form equation in the heterogeneity. In a similar model, Florens, Heckman, Meghir, and Vytlacil (2008) impose a stochastic polynomial restriction on the structural equation, but relax a large support assumption required in Imbens and Newey (2009). Chesher (2005) considers local identification of derivatives in the triangular

[4]The identification strategy can be extended to the case that both the true endogenous regressors and instrumental variables are unobservable due to measurement errors. The identification results are available from the author up on request.

simultaneous equations where monotonicity of both equations hold, but imposes a restriction on dimensionality of multi-dimensional heterogeneity. Kasy (2014) and Hoderlein, Holzmann, Kasy, and Meister (2016) maintain the conditions in Imbens and Newey (2009) but additionally assume monotonicity of the reduced-form in instrumental variables. They show that a reweighting method could serve as an alternative estimation procedure to the control variables approach. Chernozhukov and Hansen (2005), Chernozhukov, Imbens, and Newey (2007), Horowitz and Lee (2007), Komunjer and Santos (2010), and Gagliardini and Scaillet (2012) invoke a high-level completeness condition instead of considering the triangular simultaneous equations. They assume monotonicity of the structural equation in one-dimensional heterogeneity. Gandhi, Kim, and Petrin (2011) propose a generalized control function approach for the nonseparable models and apply it to estimating discrete choice demand models where observed factors including price interact with an unobserved demand error. Torgovitsky (2015) assumes monotonicity of both equations in one-dimensional heterogeneity and shows binary instruments can be used for identification. Sasaki (2015) shows that a quantile partial derivative identifies a weighted average of structural partial effects without monotonicity, building on results in Hoderlein and Mammen (2007). Williams (2015) considers a nonseparable model with endogeneity where binary proxies for unobserved heterogeneity are available, and provides identification results in the limit, as the number of proxies increases.

Since economic variables of interest are often misreported or missing, approaches aiming to correct for measurement errors in nonlinear or nonparametric models have received considerable attention in the literature. Unlike the proposed methods in this paper, most existing methods are restricted to estimating average effects and assume that all regressors are exogenous. In particular, approaches closely related to this paper are techniques which utilize repeated measurements.[5] For example, Hausman, Newey, Ichimura, and Powell (1991) propose a consistent estimator in a polynomial specification in the presence of instruments or repeated observations. Horowitz and Markatou (1996) propose nonparametric estimators for linear models with independent and identically distributed errors and show that every component of the convolution can be nonparametrically identified and estimated from repeated observations of dependent variable. Linton and Whang (2002) study nonparametric estimation of density functions and regression functions based on aggregated data that are grouped into family totals. They allow for a common within-family component but assume that observations are independent across different families. Li (2002) introduces a consistent estimator in an arbitrary nonlinear specification using repeated measurements, based on a technique devised by Li and Vuong (1998) who impose a mutual independence assumption on the measurement errors. A related method devised by Schennach (2004) relaxes the mutual independence assumption on one of measurement errors to mean independence in a general nonlinear parametric specification. Delaigle, Hall, and Meister (2008) study nonparametric density and regression models using repeated measurements with errors that are generally independent. They provide a sufficient

[5] We focus on papers using repeated measurements which are the most relevant. For other approaches based on (e.g.) the simulation, auxiliary data or instrumental variables methods, we refer to excellent reviews by Carroll, Ruppert, Stefanski, and Crainiceanu (2006) and Chen, Hong, and Nekipelov (2011).

condition for first-order properties and convergence rates of estimators to be equivalent when the error distribution is known and unknown. Useful results have been derived by Bonhomme and Robin (2010) in nonparametric estimation of distributions of latent factors in earnings dynamics by assuming mutual independence of latent factors. Evdokimov (2010) presents interesting results for identification of nonparametric panel data model with nonadditive individual-specific heterogeneity, when regressors are observed and idiosyncratic disturbances are additively separable from the structural function. Hu and Sasaki (2015) propose a consistent estimator in additively-separable nonparametric model with non-classical measurement errors by imposing location-scale normalization on one of repeated measurements. Song, Schennach, and White (2015) consider identification of average marginal effects in nonseparable model with mismeasured endogenous regressors. Their model is much simpler than the one considered here, in that the control variables are assumed to be observable and only average effects are studied.

2 Model and Background

2.1 Setup

Suppose (Y, X, Z) are independent and identically distributed (i.i.d.) random variables defined on a complete probability space $(\Omega, \mathfrak{F}, \mathbb{P})$. Our first assumption is about the triangular systems and random variables.

Assumption 1 *The triangular simultaneous equations (1) and (2) satisfy the following conditions:*
(i) The instrumental variables Z are continuously distributed in $\mathbb{R}$ for all z;
(ii) We have $Z \perp\!\!\!\perp (U, \eta)$;
(iii) The function $h(Z, \eta)$ is strictly monotone in η for all Z where η is a continuously distributed scalar with strictly increasing $F_\eta(\eta)$ on the support of η.

Assumption 1 (i) and (iii) imply that X and Z are real-valued and continuously distributed.[6] It excludes binary (or discrete) treatment and instruments. While we primarily focus on the case of continuous instruments, we note that the requirement of continuous instruments can be relaxed to binary or discrete case at the expense of additional restrictions (see Section 5 for more details). Assumption 1 (ii) imposes that the instrumental variables are exogenous, which is typically imposed in the literature. Assumption 1 (iii) restricts one-dimensional heterogeneity and monotonicity of $h(Z, \eta)$ in the reduced-form model (2) (see Matzkin (2003) for usual normalization in nonseparable models). Note that the structural model (1) allows for multi-dimensional heterogeneity U. We provide further identification results for one-dimensional unobserved heterogeneity U in Section 5.

[6] It can be shown to extend the scalar case to multi-dimensional endogenous regressors or multi-dimensional endogenous and exogenous regressors. Extension to multi-dimensional case involves mainly notational complications. (see the Supplementary Material for more details).

An emerging concern in economics is that the true endogenous causes of interest, X, may be unobservable to researchers because of measurement errors. Nevertheless, common practice is to assume either the endogeneity or the measurement error does not exist. This is not because of theoretical reason but rather simplification. In the context of measurement errors, when the true X is unobservable but a proxy to X is observed, we need to deal with mismeasured regressors in the structural equation (1) and mismeasured dependent variables in the reduced-form equation (2) of the model. Because the functions g and h are unknown, the measurement errors need to be solved in a nonlinear or nonparametric way.

We address both issues of the endogeneity and unobservability of the true X simultaneously without imposing strong restrictions such as independence of the measurement errors and others or known distribution of the measurement errors. We assume that, instead of the true X, two repeated measurements or proxies to X are observable to researchers as following:

Assumption 2 *Two repeated measurements of X are observable such as $X_1 = X + \varepsilon_1$ and $X_2 = X + \varepsilon_2$ with measurement errors, ε_1 and ε_2.*

We use additively-separable form of the measurement errors with very little loss of generality, because one can generate modified variables by taking reasonable transformations, while keeping most of properties of the original variables. For instance, log transformation of $\tilde{X}_1 = \tilde{X}e^{\varepsilon_1}$ ensures the form of measurement errors considered in the assumption, where $X_1 = \ln \tilde{X}_1$ and $X = \ln \tilde{X}$. There are many examples for the repeated measurements of economic variables. In twins data, twins are asked to report their own and their twin's years of education. In firm's production function, multiple proxies to intermediate input such as material, energy, and fuel are often available. Employer-employee matched data for wages can be another example. Matched federal agency and firm-level data (e.g., bank credit, customs data, and export data) is also popularly used. Market-based measures and firm-specific earnings forecasts from securities analysts have been widely used as proxies to firm's intrinsic value in the investment equation literature. Average of each site's observed catch rates and observed trip patterns are two available data set for the expected catch rates in demand models for recreational fishing sites. In panel data, any single variable can be repeatedly measured for the same individual at a different point in time. For instance, bidder's private signals in auction models and earnings data in earnings dynamics are good examples for the repeated measurements over time.

2.2 An Example

Consider nonseparable consumer demand system (Engel curves):

$$Y = g(X, U)$$

where Y is budget share of a commodity, X is family expenditure and U is (multi-dimensional) unobserved heterogeneity such as individual risk preference. Empirical studies of the consumer behavior go back to early works in Engel (1895) and Working (1943). In the literature, it is well-known that family expenditure is endogenous since consumption and saving are simultaneously determined (e.g., Banks, Blundell, and Lewbel (1997), Blundell, Browning, and Crawford (2003), Blundell, Chen, and Kristensen (2007) and Imbens and Newey (2009)). Family income is typically used as an instrument Z since it is independent of the individual unobserved heterogeneity U:

$$X = h(Z, \eta)$$

where η is a scalar unobserved heterogeneity. As pointed out in the literature (e.g., Liviatan (1961), Lewbel (1996), Hausman, Newey, Ichimura, and Powell (1991) and Schennach (2004)), the family expenditure generally has an issue of measurement errors due to survey errors or discrepancy between purchases and consumption because of storage or waste. Since the family expenditure enters both the structural and reduced-form equations in unknown nonlinear ways, standard instrumental variables or control variables approaches cannot consistently estimate the consumer demand function. Family expenditures in two different years can be used as the repeated measurements (X_1 and X_2) to correct for the measurement errors under some conditions.

3 Identification with Control Variables

In this section, we establish identification using control variables in the models where the heterogeneity U in equation (1) is multi-dimensional while the heterogeneity η in equation (2) is one-dimensional.

3.1 General Identification

We show various structural effects can be identified, based on a control variable V. Let $\mathcal{X}$, $\mathcal{Z}$, and $\mathcal{V}$ denote the support of X, Z, and V, respectively. We impose the following condition:

Assumption 3 *A control variable $V \in \mathcal{V}$ satisfies $X \perp\!\!\!\perp U \mid V$.*

This is a conditional exogeneity assumption which has been used in many studies, e.g., Altonji and Matzkin (2005), Firpo (2007), Rothe (2012), Chiappori, Komunjer, and Kristensen (2015) and the references therein. In particular, Imbens and Newey (2009) consider a control variable which can be either observable or estimable when X is observable. Song, Schennach, and White (2015) assumes the conditional independence when the true X is unobservable due to measurement errors, but a control variable is assumed to be observable. The methodology we develop comprehensively extends Imbens and Newey (2009) and Song, Schennach, and White (2015) into the model where both X and V are unobservable.

As shown in the proof of Theorem 1 in Imbens and Newey (2009), the monotonicity of $h(Z, \eta)$ in η and the exogenous instruments condition $Z \perp\!\!\!\perp (U, \eta)$ from equations (1) and (2) imply that a control variable is the uniformly distributed $V = F_{X|Z}(X \mid Z) = F_\eta(\eta)$. This is because of the fact that

$$F_{X|Z}(x \mid z) = \mathbb{P}(h(Z, \eta) \leq x \mid Z = z) = \mathbb{P}(\eta \leq h^{-1}(Z, x) \mid Z = z) = F_\eta(h^{-1}(z, x)),$$

à la Matzkin (2003). Thus, when X is observable, Imbens and Newey (2009) show that V can be identified from the reduced form and therefore structural effects can be identified from the conditional distribution of Y given X and V.

One of central object of interest is the average structural function (Blundell and Powell (2003)), denoted by

$$\bar{g}(x) \equiv \mathbb{E}[g(x, U)] = \int g(x, u) F_U(\mathrm{d}u) = \int \mathbb{E}[Y \mid X = x, V = v] F_V(\mathrm{d}v)$$

where $F_U(u)$ and $F_V(v)$ are the cumulative distribution of U and V, respectively. This is identified from the observables through the identification of $\mathbb{E}[Y \mid X = x, V = v]$ and $F_V(v)$.

Another useful object is the quantile structural function which is defined as the τth quantile of the function $g(x, U)$ (Imbens and Newey (2009)), denoted by $\vartheta^\tau(x)$.[7] As shown below, the object is also identified from the observables when $F_{Y|X,V}(Y \mid X = x, V = v)$ and $F_V(v)$ are identified.

3.2 Failure of Point Identification

In many economic models, the true endogenous cause X is unobservable to researchers because of the measurement errors. Therefore, structurally identified objects in the previous literature might not be identified in this case, in that they cannot be defined solely in terms of the joint distribution of observables. A primary contribution of this paper is to provide new identification results for various structural effects when the true variables are observed with measurement errors.

When measurement errors occur, point identification using the conventional approaches is not feasible. Suppose that the true X is observed with a measurement error such as $X^* = X + \varepsilon$, where

[7] Distributional effects based on other forms of measures have been extensively studied in the literature. For instance, Firpo, Fortin, and Lemieux (2009) analyze an unconditional quantile regression building on the recentered influence function. Carneiro, Heckman, and Vytlacil (2010) introduce the marginal policy relevant treatment effects and connect it to the average marginal treatment effects. Rothe (2010, 2012) proposes a method for partial distributional policy effects in a nonseparable model.

ε represents the measurement error, the reduced-form equation (2) can be rewritten as

$$X^* = h(Z, \eta) + \varepsilon = \tilde{h}(Z, \tilde{\eta})$$

where $\tilde{\eta} \equiv (\epsilon, \eta)^\top$ is a vector of unobservables. Thus, identification of the control variables $V = F_{X|Z}(X \mid Z)$ is not feasible since X is unobservable. Moreover, the function $\tilde{h}$ violates the condition of one-dimensional unobserved heterogeneity, which is essential to the control variables approach in Imbens and Newey (2009). Finally, in the structural equation (1), X is still unobservable. So identification of the control variables in the first stage is not sufficient to identify the structural parameters in the second stage. For instance, identification of the average structural function requires identification of $\mathbb{E}[Y \mid X = x, V = v]$. But this is not identified since the true X is unobserved.

3.3 Identification of Control Variables

We now discuss how the treatments in Matzkin (2003) and Imbens and Newey (2009) can be further extended to cover the unobservable true X in the presence of the measurement errors. We construct the identification results by imposing weak restrictions on the measurement errors as follows:

Assumption 4 *The measurement errors on X satisfy the following conditions:*
(i) $\mathbb{E}[\varepsilon_1 \mid X_2, Z] = 0$*;*
(ii) $\varepsilon_2 \perp\!\!\!\perp X \mid Z$*.*

Assumption 4 (i) imposes conditional mean zero of the first measurement error ε_1 given X_2 and Z. This is weaker than imposing $\mathbb{E}[\varepsilon_1 \mid X, \varepsilon_2, Z] = 0$ which is a form of classical measurement errors. Hausman, Newey, Ichimura, and Powell (1991), Schennach (2004) and Song, Schennach, and White (2015) have imposed $\mathbb{E}[\varepsilon_1 \mid X, \varepsilon_2] = 0$. In fact, Hausman, Newey, Ichimura, and Powell (1991) and Schennach (2004) do not require the instrumental variables Z since they assume that no such endogeneity exists in their nonlinear parametric models. Song, Schennach, and White (2015) also consider an endogeneity issue in nonseparable models, but they assume that the control variables V is observable to researchers, which excludes many interesting models. Thus the models in the aforementioned papers are special cases of the nonseparable simultaneous equations we consider.[8]

Assumption 4 (ii) is a conditional independence assumption which is weaker than full independence $\varepsilon_2 \perp\!\!\!\perp (X, Z)$. When the instrumental variables Z are constant, the condition is the same as

[8] Although their condition $\mathbb{E}[\varepsilon_1 \mid X, \varepsilon_2] = 0$ is much weaker than the full independence $X \perp\!\!\!\perp (\varepsilon_1, \varepsilon_2)$ in Horowitz and Markatou (1996), Li and Vuong (1998), Bonhomme and Robin (2010) and Evdokimov (2010), it still excludes the cases of $\mathbb{E}[\varepsilon_1 \mid X] \neq 0$ and $\mathbb{E}[\varepsilon_1 \mid \varepsilon_2] \neq 0$. Yet, Assumption 4 (i) embraces such a possibility. Indeed, this is an important improvement since, in many economic data, it could be the case that there is linear or nonlinear dependence among true variables and the measurement errors (Bound and Krueger (1991) and Chen, Hong, and Tamer (2005)).

the mutual independence, $\varepsilon_2 \perp\!\!\!\perp X$, imposed in Hausman, Newey, Ichimura, and Powell (1991) and Schennach (2004) who assume no endogeneity issue. As a matter of fact, $\varepsilon_2 \perp\!\!\!\perp (X, Z)$ is equivalent to imposing both $\varepsilon_2 \perp\!\!\!\perp Z$ and $\varepsilon_2 \perp\!\!\!\perp X \mid Z$, as stated in Lemma A.1. As a result, Assumption 4 (ii) requires weaker restriction than the full independence $\varepsilon_2 \perp\!\!\!\perp (X, Z)$ which has been imposed in Song, Schennach, and White (2015). The condition allows for some degree of dependence among variables.[9,10]

Thus, Assumption 4 allows for the linear dependence between the true X and the measurement error ε_1, the linear dependence between the measurement errors, and the nonlinear dependence between the true X and the measurement error ε_2. In terms of the application to consumer demand systems, the condition 4 (i) states that the measurement error in the first-year family expenditure is uncorrelated with the mismeasured second-year family expenditure and family income. The condition 4 (ii) requires that the measurement error in the second-year family expenditure is conditionally independent of the true family expenditure given family income. Since the conditional mean of the measurement error in the second-year family expenditure does not need to be zero, the assumption allows for a positive drift term in more recent family expenditure, which is commonly observed in real data.

We now provide an identification result for the control variable. Let $\mathbb{N} \equiv \{0, 1, ...\}$, $\overline{\mathbb{N}} \equiv \mathbb{N} \cup \{\infty\}$ and $\nabla_\alpha^\lambda \equiv (\partial^\lambda / \partial\alpha^\lambda)$ denote a partial derivative of degree $\lambda \in \{0, ..., \Lambda\}$ for $\Lambda \in \overline{\mathbb{N}}$ with respect to a generic variable α. Denote $T \in \mathcal{T} \equiv \{1, Y, \mathbb{1}(Y \leq y)\}$ be a generic function of Y hereafter.[11] Let $\Psi_T(\zeta, z) \equiv \mathbb{E}[Te^{\mathrm{i}\zeta X} \mid Z = z]$ and $D_{\zeta a} \equiv ae^{\mathrm{i}\zeta X_2}$, and also define

$$\varphi_a(\zeta, z) \equiv \frac{\mathbb{E}[D_{\zeta a} \mid Z = z]}{\mathbb{E}[D_{\zeta 1} \mid Z = z]}$$

for each ζ and $a \in \mathbf{a} \equiv \{1, T, X_1\}$.

In the presence of X observed with measurement errors, the main identification strategy for the control variable $V = F_{X|Z}(X \mid Z)$ is to use a useful property of the Fourier transform. Note that

[9] In particular, it allows for the dependence between instrumental variables (Z) and the second measurement error (ε_2). In addition, it does not exclude the nonlinear dependence between the true regressor (X) and the second measurement error (ε_2) since Z could be a common factor causing both X and ε_2, which is implied by the condition $\varepsilon_2 \perp\!\!\!\perp X \mid Z$.

[10] It would be possible for Assumption 4 (ii) to be replaced with a weaker condition. We use $\mathbb{E}[e^{\mathrm{i}\zeta X} e^{\mathrm{i}\zeta\varepsilon_2} \mid Z] = \mathbb{E}[e^{\mathrm{i}\zeta X} \mid Z]\mathbb{E}[e^{\mathrm{i}\zeta\varepsilon_2} \mid Z]$ in the proof of the identification results below. This is a conditional version of sub-independence discussed in Ebrahimi, Hamedani, Soofi, and Volkmer (2010) (see De Paula (2008) for a related discussion). When Z is constant, it is the same as their sub-independence concept. The concept of conditional sub-independence is much weaker than conditional independence. To see this, consider conditional independence of X and ε_2 given Z. It means that $\mathbb{E}[e^{\mathrm{i}(\zeta X + \xi\varepsilon_2)} \mid Z] = \mathbb{E}[e^{\mathrm{i}\zeta X} \mid Z]\mathbb{E}[e^{\mathrm{i}\xi\varepsilon_2} \mid Z]$ for all $(\zeta, \xi) \in \mathbb{R}^2$. Thus the conditional independence restricts the relation in two directional coordinates (ζ, ξ). In contrast, our condition means that $\mathbb{E}[e^{\mathrm{i}\zeta(X+\varepsilon_2)} \mid Z] = \mathbb{E}[e^{\mathrm{i}\zeta X} \mid Z]\mathbb{E}[e^{\mathrm{i}\zeta\varepsilon_2} \mid Z]$ for all $\zeta \in \mathbb{R}$, which only imposes a restriction on one directional coordinate ζ. We leave this possibility for future research.

[11] For notational simplicity, the argument Y of the function T is suppressed.

$\Psi_1(\zeta, z)$ can be rewritten as

$$\begin{aligned}\Psi_1(\zeta, z) &\equiv \mathbb{E}[e^{\mathrm{i}\zeta X} \mid Z = z] \\ &= \int \mathbb{E}[e^{\mathrm{i}\zeta X} \mid X = x, Z = z] f_{X|Z}(x \mid z)\mathrm{d}x \\ &= \int f_{X|Z}(x \mid z) e^{\mathrm{i}\zeta x}\mathrm{d}x\end{aligned}$$

which is the Fourier transform of $f_{X|Z}(x \mid z)$. We also note that

$$\frac{1}{2\pi}\int \Psi_1(\zeta, z)e^{-\mathrm{i}\zeta x}\mathrm{d}\zeta$$

is the inverse Fourier transform of $\Psi_1(\zeta, z)$ for $(x, z) \in \mathcal{X} \times \mathcal{Z}$. Then we get

$$f_{X|Z}(x \mid z) = \frac{1}{2\pi}\int \Psi_1(\zeta, z)e^{-\mathrm{i}\zeta x}\mathrm{d}\zeta$$

where the integral is defined over the coordinate ζ. By the inversion theorem (e.g., Gurland (1948) and Gil-Peraez (1951))

$$F_{X|Z}(x \mid z) = \frac{1}{2} + \frac{1}{2\pi}\int_0^\infty \frac{\Psi_1(-\zeta, z)e^{\mathrm{i}\zeta x} - \Psi_1(\zeta, z)e^{-\mathrm{i}\zeta x}}{\mathrm{i}\zeta}\mathrm{d}\zeta.$$

Thus $F_{X|Z}(x \mid z)$ is identified when $\Psi_1(\zeta, z)$ is identified. As fully provided in the Appendix, we identify $\Psi_1(\zeta, z)$ by characterizing it in terms of observables using Assumption 4 as follows: for each ζ,

$$\Psi_1(\zeta, z) = \exp\left(\int_0^\zeta \mathrm{i}\varphi_{X_1}(\xi, z)\mathrm{d}\xi\right),$$

where $\varphi_{X_1}(\xi, z)$ is a function of observed random variables.

The idea of identifying $\Psi_1(\zeta, z)$ rests on nullifying the measurement errors in the frequency domain. Given the restrictions on the measurement errors in Assumption 4, $\Psi_1(\zeta, z)$ can be rewritten in terms of observed quantities without the measurement errors in coordinate ζ (see proofs of Lemma A.2 and Theorem 3.1 for more details). As a result, for fixed x and z, $V = F_{X|Z}(x \mid z)$ is identified. The next theorem summarizes the result:

Theorem 3.1 *Suppose Assumptions 2 and 4 hold. Then, from the set of observables $\{X_1, X_2, Z\}$, the conditional cumulative distribution $F_{X|Z}(x \mid z)$ is identified as*

$$F_{X|Z}(x \mid z) = \frac{1}{2} + \frac{1}{2\pi}\int_0^\infty \frac{\Psi_1(-\zeta, z)e^{\mathrm{i}\zeta x} - \Psi_1(\zeta, z)e^{-\mathrm{i}\zeta x}}{\mathrm{i}\zeta}\mathrm{d}\zeta$$

with $\Psi_1(\zeta, z) = \exp\left(\int_0^\zeta \mathrm{i}\varphi_{X_1}(\zeta, z)\mathrm{d}\xi\right)$ for each real ζ and $(x, z) \in \mathcal{X} \times \mathcal{Z}$, provided that all expec-

tations exist and $\mathbb{E}[D_{\xi 1} \mid Z = z]$ *is nonvanishing for any real* ξ *and* $z \in \mathcal{Z}$.[12]

We note that (x, z) are fixed in the definition of V. Although the true X is unobserved, the control variable V is identified from the observed variables (X_1, X_2, Z) for fixed (x, z).[13]

3.4 Identification of Structural Functions

We next provide identification results for the structural parameters such as the covariate-conditioned average structural function, average structural function and quantile structural function. The following condition is imposed.

Assumption 5 *The measurement errors satisfy the following conditions:*
(i) $\mathbb{E}[\varepsilon_1 \mid X_2, V] = 0$*;*
(ii) $\varepsilon_2 \perp\!\!\!\perp X \mid V$*;*
(iii) $\mathbb{E}[T \mid X, \varepsilon_2, V] = \mathbb{E}[T \mid X, V]$*.*

Parts (i) and (ii) in the condition are similar to Assumption 4, except having V in place of Z inside the information set. Thus, interpretation of Assumption 4 is conveyed to Assumption 5 (i)–(ii).

Assumption 5 (iii) is new here and imposes that ε_2 does not provide further information on T than X and V do. It implies a limited information transmission of the measurement error ε_2 to the dependent variable Y. And the restriction could be applied to either mean (when $T = Y$) or quantile of Y (when $T = \mathbb{1}(Y \leq y)$). For instance, when one is interested in the average effects (i.e., $T = Y$), the condition is about a mean independence assumption. In this case, the condition is similar to the one considered by Song, Schennach, and White (2015) who primarily focus on the average effects. When one is interested in the quantile effects (i.e., $T = \mathbb{1}(Y \leq y)$), the condition can be rewritten as $\varepsilon_2 \perp\!\!\!\perp Y \mid (X, V)$. From a similar argument to Lemma A.1, we can claim that $\varepsilon_2 \perp\!\!\!\perp (Y, X, V)$ if and only if $\varepsilon_2 \perp\!\!\!\perp (X, V)$ and $\varepsilon_2 \perp\!\!\!\perp Y \mid (X, V)$. So Assumption 5 (iii) is a weaker restriction than the full independence $\varepsilon_2 \perp\!\!\!\perp (Y, X, V)$. This does not exclude some degree of

[12] Alternatively, the cumulative distribution can be expressed as

$$F_{X|Z}(x \mid z) = \frac{1}{2} - \lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{\Psi_1(\zeta, z) e^{-\mathrm{i}\zeta x}}{\mathrm{i}\zeta} \mathrm{d}\zeta$$

or can be recovered from

$$F_{X|Z}(x \mid z) = \int_{-\infty}^{x} f_{X|Z}(\tilde{x} \mid z) \mathrm{d}\tilde{x}, \quad \text{where} \quad f_{X|Z}(\tilde{x} \mid z) = \frac{1}{2\pi} \int \Psi_1(\zeta, z) e^{-\mathrm{i}\zeta\tilde{x}} \mathrm{d}\zeta.$$

[13] As shown in following sections, X is inside the conditioning sets, and structural parameters are identified and estimated at a fixed x. Thus whether or not X is observed is immaterial as long as structural parameters are identified from observables. This is different than the conventional methods such as Imbens and Newey (2009) where observed X is used in the nonparametric estimation.

dependence of ε_2 and (X, V), which is compatible with Assumption 5 (ii). In the application, this condition requires that the measurement error in the second-year family expenditure is conditionally independent of the budget share of a commodity given the true family expenditure (and the control variable). On the other hand, it still allows for the dependence between the measurement error and the true family expenditure (and the control variable).

Identification of the structural parameters is obtained by generalizing the result for the control variable in Section 3.3. Let $\Psi_T(\zeta, v) \equiv \mathbb{E}[Te^{\mathrm{i}\zeta X} \mid V = v]$, and define a quantity of the general form

$$\mathcal{A}_T(x, v) \equiv \mathbb{E}[T \mid X = x, V = v] f_{X|V}(x \mid v) \tag{3}$$

for $T \in \mathcal{T}$ and $(x, v) \in \mathcal{X} \times \mathcal{V}$. Note that $\Psi_T(\zeta, v)$ can be rewritten as

$$\begin{aligned}\Psi_T(\zeta, v) &\equiv \mathbb{E}[Te^{\mathrm{i}\zeta X} \mid V = v] \\ &= \int \mathbb{E}[Te^{\mathrm{i}\zeta X} \mid X = x, V = v] f_{X|V}(x \mid v) \mathrm{d}x \\ &= \int \mathbb{E}[T \mid X = x, V = v] f_{X|V}(x \mid v) e^{\mathrm{i}\zeta x} \mathrm{d}x \\ &= \int \mathcal{A}_T(x, v) e^{\mathrm{i}\zeta x} \mathrm{d}x\end{aligned}$$

which is the Fourier transform of $\mathcal{A}_T(x, v)$. Since

$$\frac{1}{2\pi} \int \Psi_T(\zeta, v) e^{-\mathrm{i}\zeta x} \mathrm{d}\zeta$$

is the inverse Fourier transform of $\Psi_T(\zeta, v)$ for $(x, v) \in \mathcal{X} \times \mathcal{V}$, we obtain

$$\mathcal{A}_T(x, v) = \frac{1}{2\pi} \int \Psi_T(\zeta, v) e^{-\mathrm{i}\zeta x} \mathrm{d}\zeta$$

where the integral is defined over the coordinate ζ. Thus $\mathcal{A}_T(x, v)$ is identified when $\Psi_T(\zeta, v)$ is identified.

Similar to $\varphi_a(\zeta, z)$, we define

$$\varphi_a(\zeta, v) \equiv \frac{\mathbb{E}[D_{\zeta a} \mid V = v]}{\mathbb{E}[D_{\zeta 1} \mid V = v]},$$

where $D_{\zeta a} \equiv a e^{\mathrm{i}\zeta X_2}$, for each ζ and $a \in \mathbf{a} \equiv \{1, T, X_1\}$. We identify $\Psi_T(\zeta, v)$ by characterizing it in terms of observables using Assumption 5 as follows:

$$\Psi_T(\zeta, v) = \varphi_T(\zeta, v) \exp\left(\int_0^\zeta \mathrm{i}\varphi_{X_1}(\xi, v) \mathrm{d}\xi\right)$$

where $\varphi_T(\zeta, v)$ and $\varphi_{X_1}(\zeta, v)$ are functions of observables.

Note that, given Assumption 3, $\mu(x,v) \equiv \mathbb{E}[Y \mid X = x, V = v] = \mathbb{E}[g(X,U) \mid V = v]$ when $T = Y$. Here $\mu(x,v)$ is a conditional analog of the average structural function $\bar{g}(x)$ (Blundell and Powell (2003)). $\mu(x,v)$ can be identified from the following result:

Theorem 3.2 *Suppose Assumptions 2 and 5 hold. Then, from the set of observables $\{X_1, X_2, Y, V\}$, the quantity $\mathcal{A}_T(x,v)$ is identified as*

$$\mathcal{A}_T(x,v) = \frac{1}{2\pi}\int \Psi_T(\zeta,v)e^{-\mathrm{i}\zeta x}\mathrm{d}\zeta$$

for $T \in \mathcal{T}$, each real ζ and $(x,v) \in \mathcal{X} \times \mathcal{V}$, where

$$\Psi_T(\zeta,v) = \varphi_T(\zeta,v)\exp\left(\int_0^\zeta \mathrm{i}\varphi_{X_1}(\xi,v)\mathrm{d}\xi\right),$$

provided that all expectations exist and $\mathbb{E}[D_{\xi 1} \mid V = v]$ is nonvanishing for any real ξ and $v \in \mathcal{V}$.

We note that the conditions of the nonvanishing characteristic functions $\mathbb{E}[D_{\xi 1} \mid Z = z]$ and $\mathbb{E}[D_{\xi 1} \mid V = v]$ in Theorems 3.1 and 3.2, respectively, impose restrictions on the conditional support of X_2. These have been commonly imposed in the deconvolution literature (e.g., Fan (1991) and Fan and Truong (1993)).

Since $\mu(x,v) = \mathcal{A}_Y(x,v)/\mathcal{A}_1(x,v)$, we can identify $\mu(x,v)$, based on the identification of $\mathcal{A}_Y(x,v)$ and $\mathcal{A}_1(x,v)$ from Theorem 3.2. Let $p(x) \equiv \int_{\mathcal{V}\cap\mathcal{V}(x)^c} F_V(\mathrm{d}V)$ where $\mathcal{V}(x)$ denotes the support of V conditional on $X = x$. We assume that the nonparametric rank condition, $p(x) = 0$, is satisfied. The condition $p(x) = 0$ guarantees that the support of Z covers the support of V. This is equivalent to the common support condition in Imbens and Newey (2009), which requires that the instrumental variables Z vary sufficiently. It thus ensures that $\mu(x,v)$ is identified over the full support of the marginal distribution of V. With identified $\mu(x,v)$, the average structural function can be identified by integrating out the control variable V, for a fixed x:

$$\bar{g}(x) = \int \mathbb{E}[Y \mid X = x, V = v]F_V(\mathrm{d}v) = \int \frac{\mathcal{A}_Y(x,v)}{\mathcal{A}_1(x,v)}F_V(\mathrm{d}v)$$

where $F_V(v)$, the cumulative distribution of $V \equiv F_{X|Z}(x \mid z)$, is identified from Theorem 3.1. We have shown the following result:

Lemma 3.3 *In the model of equations (1) and (2) where Assumptions 1–5 hold and $p(x) = 0$, the covariate-conditioned average structural function $\mu(x,v)$ and the average structural function $\bar{g}(x)$ are identified from the set of observables $\{X_1, X_2, Y, Z\}$ for all $(x,v) \in \mathcal{X} \times \mathcal{V}$.*

Note that when $T = \mathbb{1}(Y \le y)$, we have $\mathcal{A}_{\mathbb{1}(Y\le y)}(x,v) = \mathbb{E}[\mathbb{1}(Y \le y) \mid X = x, V = v]f_{X|V}(x \mid v)$, and when $T = 1$, we have $\mathcal{A}_1(x,v) = f_{X|V}(x \mid v)$. Thus the quantile structural function can be also identified. Note that

$$\begin{aligned} F_{Y|X,V}(y \mid x, v) &= \int \mathbb{1}(g(x,u) \leq y) F_{U|V}(\mathrm{d}u \mid v) \\ &= \mathbb{E}[\mathbb{1}(Y \leq y) \mid X = x, V = v] \\ &= \frac{\mathcal{A}_{\mathbb{1}(Y \leq y)}(x,v)}{\mathcal{A}_1(x,v)}. \end{aligned}$$

Then under $p(x) = 0$, integrating over the marginal distribution of V gives

$$\begin{aligned} \mathcal{G}(y,x) &\equiv \int \mathbb{1}(g(x,u) \leq y) F_U(\mathrm{d}u) \\ &= \int F_{Y|X,V}(y \mid x, v) F_V(\mathrm{d}v) \\ &= \int \frac{\mathcal{A}_{\mathbb{1}(Y \leq y)}(x,v)}{\mathcal{A}_1(x,v)} F_V(\mathrm{d}v). \end{aligned}$$

From the definition of the quantile structural function, we obtain

$$\vartheta^\tau(x) = \mathcal{G}^{-1}(\tau, x).$$

By $p(x) = 0$, $F_{Y|X,V}(y \mid x, v)$ is unique on $\mathcal{X} \times \mathcal{V}$ because $\mathcal{V}(x)$ is equal to $\mathcal{V}$. The next lemma summarizes the result:

Lemma 3.4 *In the model of equations (1) and (2) where Assumptions 1–5 hold and $p(x) = 0$, the quantile structural function $\vartheta^\tau(x)$ is identified from the set of observables $\{X_1, X_2, Y, Z\}$ for all $x \in \mathcal{X}$.*

3.5 Partial Identification

For the point identification of the quantile structural function and average structural function, the common support condition, $p(x) = 0$, is necessary. It guarantees that the support of V is equal to the support of V conditional on $X = x$. If the condition is so strong in an empirical application that it is dropped, one can still bound the quantile structural function and average structural function. We give partial identification results in the following result:

Proposition 3.5 *In the model of equations (1) and (2) where Assumptions 1–5 hold, bounds for the distribution of the structural function are identified from the set of observables $\{X_1, X_2, Y, Z\}$ for all $x \in \mathcal{X}$:*

$$\int_{\mathcal{V}(x)} \frac{\mathcal{A}_{\mathbb{1}(Y \leq y)}(x,v)}{\mathcal{A}_1(x,v)} F_V(\mathrm{d}v) \leq \mathcal{G}(y,x) \leq \int_{\mathcal{V}(x)} \frac{\mathcal{A}_{\mathbb{1}(Y \leq y)}(x,v)}{\mathcal{A}_1(x,v)} F_V(\mathrm{d}v) + p(x)$$

and these bounds are sharp. Furthermore, suppose that $B_l \leq g(x,u) \leq B_u$ for all x in the support of X and u in the support of U. Then, bounds for the average structural function are identified:

$$\int_{\mathcal{V}(x)} \frac{\mathcal{A}_Y(x,v)}{\mathcal{A}_1(x,v)} F_V(\mathrm{d}v) + B_l p(x) \leq \bar{g}(x) \leq \int_{\mathcal{V}(x)} \frac{\mathcal{A}_Y(x,v)}{\mathcal{A}_1(x,v)} F_V(\mathrm{d}v) + B_u p(x)$$

and these bounds are sharp.

3.6 Marginal Effects

Researchers may be interested in marginal effects of economic policies. In the case, previous results can be further modified to show identification. We define the covariate-conditioned average marginal effects by $\beta(x,v) \equiv \mathbb{E}[\nabla_x^1 g(X=x,U)]$. Recall $\mu(x,v) \equiv \mathbb{E}[Y \mid X=x, V=v] = \mathbb{E}[g(X,U) \mid V=v]$. Given Assumption 3, the covariate-conditioned average marginal effects can be identified by $\mu(x,v)$ as follows:

$$\begin{aligned}
\beta(x,v) &\equiv \mathbb{E}[\nabla_x^1 g(X=x,U)] \\
&= \int \nabla_x^1 g(X=x,U) F_{U|V}(\mathrm{d}U \mid v) \\
&= \int \nabla_x^1 g(X=x,U) F_{U|X,V}(\mathrm{d}U \mid x,v) \\
&= \nabla_x^1 \int g(X=x,U) F_{U|X,V}(\mathrm{d}U \mid x,v) \\
&= \nabla_x^1 \mu(x,v),
\end{aligned}$$

provided that the integral and derivative are interchangeable. More generally, any kind of the covariate-conditioned marginal effects (e.g., quantile marginal effects) can be identified from $\mathbb{E}[T \mid X=x, V=v]$.

Recall $\nabla_x^\lambda \equiv (\partial^\lambda / \partial x^\lambda)$ denotes a partial derivative of degree $\lambda \in \{0,...,\Lambda\}$ for $\Lambda \in \overline{\mathbb{N}}$ with respect to x. For $T \in \mathcal{T}$, we define a general form of the function

$$\mathcal{A}_{T,\lambda}(x,v) \equiv \nabla_x^\lambda (\mathbb{E}[T \mid X=x, V=v] f_{X|V}(x \mid v)).$$

The marginal effects can be shown to be functionals of $\mathcal{A}_{T,\lambda}(x,v)$. To show identification of the marginal effects, identification of $\mathcal{A}_{T,\lambda}(x,v)$ is essential. As shown in the Appendix, identification of $\mathcal{A}_{T,\lambda}(x,v)$ is obtained by generalizing the results in Section 3.4 into the version of their derivatives. The following theorem provides the identification results for $\mathcal{A}_{T,\lambda}(x,v)$:

Theorem 3.6 *Suppose Assumptions 2 and 5 hold. Then, from the set of observables $\{X_1, X_2, Y, V\}$,* *the quantity $\mathcal{A}_{T,\lambda}(x, v)$ is identified as*

$$\mathcal{A}_{T,\lambda}(x, v) = \frac{1}{2\pi} \int (-\mathrm{i}\zeta)^{\lambda} \Psi_T(\zeta, v) e^{-\mathrm{i}\zeta x} \mathrm{d}\zeta$$

for $T \in \mathcal{T}$, each real ζ, $\lambda \in \{0, ..., \Lambda\}$ and $(x, v) \in \mathcal{X} \times \mathcal{V}$, where

$$\Psi_T(\zeta, v) = \varphi_T(\zeta, v) \exp\left(\int_0^{\zeta} \mathrm{i}\varphi_{X_1}(\xi, v) \mathrm{d}\xi \right),$$

provided that all expectations exist and $\mathbb{E}[D_{\xi 1} \mid V = v]$ is nonvanishing for any real ξ and $v \in \mathcal{V}$.

From the first-order derivative of $\mu(x, v) = \mathcal{A}_Y(x, v)/\mathcal{A}_1(x, v)$ with respect to x, we obtain the covariate-conditioned average marginal effects of X on Y at x given $V = v$ such as

$$\beta(x, v) = \frac{\mathcal{A}_{Y,1}(x, v)}{\mathcal{A}_{1,0}(x, v)} - \frac{\mathcal{A}_{Y,0}(x, v)}{\mathcal{A}_{1,0}(x, v)} \frac{\mathcal{A}_{1,1}(x, v)}{\mathcal{A}_{1,0}(x, v)}.$$

Thus, the marginal effects can be recovered from $\mathcal{A}_{T,\lambda}(x, v)$. We give a formal result as follows:

Lemma 3.7 *In the model of equations (1) and (2) where Assumptions 1–5 hold, $\beta(x, v)$ is identified from the set of observables $\{X_1, X_2, Y, Z\}$ for all $(x, v) \in \mathcal{X} \times \mathcal{V}$.*

3.7 Average Effects

Based on $\beta(x, v)$, many interesting average marginal effects (see, e.g., Heckman and Vytlacil (2005)) can be identified. The first example is the average treatment effects by Florens, Heckman, Meghir, and Vytlacil (2008):

$$\begin{aligned}
\gamma(x) &\equiv \int \mathbb{E}[\nabla_x^1 g(X = x, U)] f_V(v) \mathrm{d}v \\
&= \int \beta(x, v) f_V(v) \mathrm{d}v \\
&= \int \left(\frac{\mathcal{A}_{Y,1}(x, v)}{\mathcal{A}_{1,0}(x, v)} - \frac{\mathcal{A}_{Y,0}(x, v)}{\mathcal{A}_{1,0}(x, v)} \frac{\mathcal{A}_{1,1}(x, v)}{\mathcal{A}_{1,0}(x, v)} \right) f_V(v) \mathrm{d}v.
\end{aligned}$$

The second example is the local average response of Altonji and Matzkin (2005) and the effect of continuous treatment on the treated (e.g., Florens, Heckman, Meghir, and Vytlacil (2008)):

$$\tilde{\gamma}(x) \equiv \int \mathbb{E}[\nabla_x^1 g(X = x, U)] f_{V|X}(v \mid X) \mathrm{d}v = \int \beta(x, v) f_{V|X}(v \mid X) \mathrm{d}v.$$

Since $f_{V|X}(v \mid X)$ can be rewritten as

$$f_{V|X}(v \mid X) = \frac{f_{X|V}(x \mid v) f_V(v)}{\int f_{X|V}(x \mid v) f_V(v) dv},$$

we can rewrite $\tilde{\gamma}(x)$ as

$$\begin{aligned}
\tilde{\gamma}(x) &= \int \left(\frac{\mathcal{A}_{Y,1}(x,v)}{\mathcal{A}_{1,0}(x,v)} - \frac{\mathcal{A}_{Y,0}(x,v)}{\mathcal{A}_{1,0}(x,v)} \frac{\mathcal{A}_{1,1}(x,v)}{\mathcal{A}_{1,0}(x,v)} \right) \frac{\mathcal{A}_{1,0}(x,v) f_V(v)}{\int \mathcal{A}_{1,0}(x,v) f_V(v) \mathrm{d}v} \mathrm{d}v \\
&= \int \left(\mathcal{A}_{Y,1}(x,v) - \mathcal{A}_{Y,0}(x,v) \frac{\mathcal{A}_{1,1}(x,v)}{\mathcal{A}_{1,0}(x,v)} \right) \frac{f_V(v)}{\int \mathcal{A}_{1,0}(x,v) f_V(v) \mathrm{d}v} \mathrm{d}v.
\end{aligned}$$

The third example is the average derivative of Stoker (1986) and Powell, Stock, and Stoker (1989):

$$\begin{aligned}
\delta &\equiv \mathbb{E}[\nabla_x^1 g(X = x, U)] \\
&= \int \beta(x,v) \mathrm{d}(x,v) \\
&= \int \left(\frac{\mathcal{A}_{Y,1}(x,v)}{\mathcal{A}_{1,0}(x,v)} - \frac{\mathcal{A}_{Y,0}(x,v)}{\mathcal{A}_{1,0}(x,v)} \frac{\mathcal{A}_{1,1}(x,v)}{\mathcal{A}_{1,0}(x,v)} \right) \mathrm{d}(x,v).
\end{aligned}$$

The fourth example is some weighted average measure of the local average response studied by Altonji and Matzkin (2005):

$$\tilde{\delta} \equiv \mathbb{E}[\nabla_x^1 g(X = x, U) f_{X,V}(x,v)] = \int \beta(x,v) f_{X,V}(x,v) \mathrm{d}(x,v).$$

Since $f_{X,V}(x,v)$ can be rewritten as

$$f_{X,V}(x,v) = f_{X|V}(x \mid v) f_V(v),$$

this object can be rewritten as

$$\begin{aligned}
\tilde{\delta} &= \int \left(\frac{\mathcal{A}_{Y,1}(x,v)}{\mathcal{A}_{1,0}(x,v)} - \frac{\mathcal{A}_{Y,0}(x,v)}{\mathcal{A}_{1,0}(x,v)} \frac{\mathcal{A}_{1,1}(x,v)}{\mathcal{A}_{1,0}(x,v)} \right) \mathcal{A}_{1,0}(x,v) f_V(v) \mathrm{d}(x,v) \\
&= \int \left(\mathcal{A}_{Y,1}(x,v) - \mathcal{A}_{Y,0}(x,v) \frac{\mathcal{A}_{1,1}(x,v)}{\mathcal{A}_{1,0}(x,v)} \right) f_V(v) \mathrm{d}(x,v).
\end{aligned}$$

The next lemma summarizes the result:

Lemma 3.8 *In the model of equations (1) and (2) where Assumptions 1–5 hold, the average effects* $\gamma(x)$, $\tilde{\gamma}(x)$, δ, *and* $\tilde{\delta}$ *are identified from the set of observables* $\{X_1, X_2, Y, Z\}$ *for all* $x \in \mathcal{X}$.

4 Identification with Instrumental Variables

We now discuss identification of the structural parameters, based on the instrumental variables approach. We show that a structural restriction of the monotonicity of the reduced-form equation (2) in the instrumental variable Z provides interesting identification powers. This identification strategy reduces two steps in the control variables approach into one step at the expense of the additional restriction of the monotonicity in Z. As a result, estimators based on the identification results would be much simpler than those from the two-step control variables approach.

4.1 General Identification

We first assume the following condition.

Assumption 6 *The function $h(Z, \eta)$ is strictly monotone in $Z \in \mathcal{Z}$ for all η.*

This assumption imposes monotonicity of the individual decision rule $h(Z, \eta)$ in the instrumental variables. For example, In the consumer demand systems, the family expenditure should be an increasing function of the family income. By abuse of notation, we define $W := W(x) \equiv h^{-1}(x, \eta) \in \mathcal{W}$ for given η, where $\mathcal{W}$ denotes the support of W. We set W to be positive (or negative) infinity when there is no such $z \in \mathcal{Z}$ that satisfies the decision rule $x = h(z, \eta)$. By the monotonicity of $h(z, \eta)$ in z, $W \leq w$ if and only if $h(z, \eta) \leq x$. Thus, the cumulative distribution of W is

$$F_W(w) \equiv \mathbb{P}(W \leq w) = \mathbb{P}(h(z, \eta) \leq x) = \mathbb{P}(X \leq x \mid Z = z) = F_{X|Z}(x \mid z) \tag{4}$$

where the third equality follows by $Z \perp\!\!\!\perp \eta$.

Let $\tilde{p}(x) \equiv \int_{\mathcal{W} \cap \mathcal{Z}^c} F_W(\mathrm{d}w)$. For any integrable function $T(Y)$, we obtain the following result:

Theorem 4.1 *In the model of equations (1) and (2) where Assumptions 1 and 6 hold and $\tilde{p}(x) = 0$, we have*

$$\mathbb{E}[T(g(x, U))] = \int \mathbb{E}[T(Y) \mid X = x, Z = z] F_W(\mathrm{d}w),$$

where $W \equiv Z = h^{-1}(x, \eta)$ for given η.

The result is similar in form to Corollary 2 in Hoderlein, Holzmann, Kasy, and Meister (2016) when the true X is observable without the measurement errors. It says that variations of x in $\mathbb{E}[T(Y) \mid X = x, Z = z]$ correspond to variations of x in $T(g(x, U))$. It thus shows that identification of $\mathbb{E}[T(Y) \mid X = x, Z = z]$ and $F_W(w)$ is sufficient to identify the structural effects. The role of $\tilde{p}(x) = 0$ is equivalent to the one of $p(x) = 0$ in the control variables approach. The condition guarantees that the support of Z covers the support of W. It thus ensures that $\mathbb{E}[T(Y) \mid X = x, Z = z]$ is identified over the full support of the marginal distribution of W.

However, the identification strategy is not feasible if the true X is observed with the measurement errors. In addition to Assumption 4, we assume the following condition which is associated with the measurement errors.

Assumption 7 *The measurement errors on X satisfy the following condition:*
$\mathbb{E}[T \mid X, \varepsilon_2, Z] = \mathbb{E}[T \mid X, Z]$.

Assumption 7 is similar to Assumption 5 (iii) except including Z in place of V in the information set. It imposes that ε_2 does not provide further information on T than X and Z do, which implies a limited information transmission of the measurement error ε_2 to the dependent variable Y.

Recall $\Psi_T(\zeta, z) \equiv \mathbb{E}[Te^{\mathrm{i}\zeta X} \mid Z = z]$, $D_{\zeta a} \equiv ae^{\mathrm{i}\zeta X_2}$ and

$$\varphi_a(\zeta, z) \equiv \frac{\mathbb{E}[D_{\zeta a} \mid Z = z]}{\mathbb{E}[D_{\zeta 1} \mid Z = z]}$$

for each ζ and $a \in \mathbf{a} \equiv \{1, T, X_1\}$. Define a quantity of the general form

$$\mathcal{A}_T(x, z) \equiv \mathbb{E}[T \mid X = x, Z = z] f_{X|Z}(x \mid z) \tag{5}$$

for $T \in \mathcal{T}$ and $(x, z) \in \mathcal{X} \times \mathcal{Z}$. We establish identification of the quantity $\mathcal{A}_T(x, z)$ in the presence of the unobservable true X, and then obtain identification of the structural parameters. As in Section 3, the main identification strategy is to use a useful property of the Fourier transform. Note that $\Psi_T(\zeta, z)$ can be rewritten as

$$\Psi_T(\zeta, z) \equiv \mathbb{E}[Te^{\mathrm{i}\zeta X} \mid Z = z] = \int \mathcal{A}_T(x, z) e^{\mathrm{i}\zeta x} \mathrm{d}x.$$

So we get

$$\mathcal{A}_T(x, z) = \frac{1}{2\pi} \int \Psi_T(\zeta, z) e^{-\mathrm{i}\zeta x} \mathrm{d}\zeta$$

where the integral is defined over the frequency domain ζ. Thus $\mathcal{A}_T(x, z)$ is identified when $\Psi_T(\zeta, z)$ is identified. We then identify $\Psi_T(\zeta, z)$ by characterizing it in terms of observables using Assumptions 4 and 7 as follows:

$$\Psi_T(\zeta, z) = \varphi_T(\zeta, z) \exp\left(\int_0^\zeta \mathrm{i}\varphi_{X_1}(\xi, z) \mathrm{d}\xi \right).$$

As a result, the quantity $\mathcal{A}_T(x, z)$ in equation (5) is identified.

Note that $\mathcal{A}_1(x, z) = f_{X|Z}(x \mid z)$. Thus, given the identification of $\mathcal{A}_1(x, z)$, $F_{X|Z}(x \mid z)$ is also identified by the inversion theorem (e.g., Gurland (1948) and Gil-Peraez (1951)). Then by equation (4), $F_W(w) = F_{X|Z}(x \mid z)$ is also identified. The next theorem summarizes the result:

Theorem 4.2 *Suppose Assumptions 2, 4, and 7 hold. Then, from the set of observables* $\{X_1, X_2, Y, Z\}$*, the quantity* $\mathcal{A}_T(x,z)$ *is identified as*

$$\mathcal{A}_T(x,z) = \frac{1}{2\pi}\int \Psi_T(\zeta,z)e^{-\mathrm{i}\zeta x}\mathrm{d}\zeta$$

for $T \in \mathcal{T}$*, each real* ζ *and* $(x,z) \in \mathcal{X}\times\mathcal{Z}$*, where*

$$\Psi_T(\zeta,z) = \varphi_T(\zeta,z)\exp\left(\int_0^\zeta \mathrm{i}\varphi_{X_1}(\xi,z)\mathrm{d}\xi\right),$$

provided that all expectations exist and $\mathbb{E}[D_{\xi 1} \mid Z = z]$ *is nonvanishing for any real* ξ *and* $z \in \mathcal{Z}$*. Furthermore, the cumulative distribution of* W *is also identified as*

$$F_W(w) = \frac{1}{2} + \frac{1}{2\pi}\int_0^\infty \frac{\Psi_1(-\zeta,z)e^{\mathrm{i}\zeta x} - \Psi_1(\zeta,z)e^{-\mathrm{i}\zeta x}}{\mathrm{i}\zeta}\mathrm{d}\zeta$$

with $\Psi_1(\zeta,z) = \exp\left(\int_0^\zeta \mathrm{i}\varphi_{X_1}(\xi,z)\mathrm{d}\xi\right)$ *for each real* ζ *and* $(x,z) \in \mathcal{X}\times\mathcal{Z}$*.*

Theorem 4.2 is essential in identifying many structural effects below.

4.2 Identification of Structural Functions

We now provide identification results for the structural parameters such as the average structural function and quantile structural function, using the instrumental variables approach.

Recall that the average structural function (Blundell and Powell (2003)) is defined as

$$\bar{g}(x) \equiv \mathbb{E}[g(x,U)] = \int g(x,u)F_U(\mathrm{d}u).$$

To identify it, note that from Theorem 4.1 with $T(Y) = Y$, we have

$$\mathbb{E}[g(x,U)] = \int \mathbb{E}[Y \mid X = x, Z = z]F_W(\mathrm{d}w).$$

As a result, the average structural function $\bar{g}(x)$ can be identified from the identification of $\mathcal{A}_T(x,z)$ and $F_W(w)$. Since $\mathcal{A}_1(x,z) = f_{X|Z}(x \mid z)$ and $\mathcal{A}_Y(x,z) = \mathbb{E}[Y \mid X = x, Z = z]f_{X|Z}(x \mid z)$ from equation (5), $\mathbb{E}[Y \mid X = x, Z = z]$ can be identified through the equality

$$\mathbb{E}[Y \mid X = x, Z = z] = \frac{\mathcal{A}_Y(x,z)}{\mathcal{A}_1(x,z)}.$$

We thus have the average structural function as

$$\bar{g}(x) = \int \frac{\mathcal{A}_Y(x,z)}{\mathcal{A}_1(x,z)} F_W(\mathrm{d}w).$$

In addition, $F_W(w)$ is also identified from Theorem 4.2. Thus the average structural function can be identified, based on the identification of $\mathcal{A}_1(x,z)$, $\mathcal{A}_Y(x,z)$ and $F_W(w)$ from Theorem 4.1 and Theorem 4.2. We have therefore shown the following identification result:

Lemma 4.3 *In the model of equations (1) and (2) where Assumptions 1, 2, 4, 6 and 7 hold and $\tilde{p}(x) = 0$, the average structural function $\bar{g}(x)$ is identified from the set of observables $\{X_1, X_2, Y, Z\}$ for all $x \in \mathcal{X}$.*

We now discuss how to recover the quantile structural effects when the true endogenous regressors are unobservable. Note that from Theorem 4.1 with $T(Y) = \mathbb{1}(Y \leq y)$, we have

$$\begin{aligned}
\mathcal{G}(y,x) &\equiv \mathbb{P}(g(x,U) \leq y) \\
&= \mathbb{E}[\mathbb{1}(g(x,U) \leq y)] \\
&= \int \mathbb{E}[\mathbb{1}(Y \leq y) \mid X = x, Z = z] F_W(\mathrm{d}w) \\
&= \int F_{Y|X,Z}(y \mid x,z) F_W(\mathrm{d}w).
\end{aligned}$$

Since $\mathcal{A}_1(x,z) = f_{X|Z}(x \mid z)$ and $\mathcal{A}_{\mathbb{1}(Y \leq y)}(x,z) = \mathbb{E}[\mathbb{1}(Y \leq y) \mid X = x, Z = z] f_{X|Z}(x \mid z)$ from equation (5), $F_{Y|X,Z}(y \mid x,z)$ can be identified from

$$F_{Y|X,Z}(y \mid x,z) \equiv \mathbb{E}[\mathbb{1}(Y \leq y) \mid X = x, Z = z] = \frac{\mathcal{A}_{\mathbb{1}(Y \leq y)}(x,z)}{\mathcal{A}_1(x,z)},$$

and $F_W(w)$ is also identified from Theorem 4.2. The condition $\tilde{p}(x) = 0$ ensures that $F_{Y|X,Z}(y \mid x,z)$ is unique on $\mathcal{X} \times \mathcal{Z}$. We have then

$$\mathcal{G}(y,x) = \int \frac{\mathcal{A}_{\mathbb{1}(Y \leq y)}(x,z)}{\mathcal{A}_1(x,z)} F_W(\mathrm{d}w).$$

Because, by definition, the inverse of $\int \mathbb{E}[\mathbb{1}(Y \leq y) \mid X = x, Z = z] F_W(\mathrm{d}w)$ is the quantile structural function, we get

$$\vartheta^\tau(x) = \mathcal{G}^{-1}(\tau, x).$$

The following lemma summarizes the result:

Lemma 4.4 *In the model of equations (1) and (2) where Assumptions 1, 2, 4, 6 and 7 hold and $\tilde{p}(x) = 0$, the quantile structural function $\vartheta^\tau(x)$ is identified from the set of observables $\{X_1, X_2, Y, Z\}$ for all $x \in \mathcal{X}$.*

4.3 Partial Identification

The condition $\tilde{p}(x) = 0$ requires that the support of W is the same as the support of Z. If this assumption may only be satisfied on a small set $\mathcal{Z}$, one might want to drop the assumption $\tilde{p}(x) = 0$ and bound the quantile structural function. In addition, if the structural function $g(x, u)$ is bounded, one can also bound the average structural function. Note that $\mathcal{A}_T(x, z)$ and $F_W(w)$ are identified for $(x, z) \in \mathcal{X} \times \mathcal{Z}$ from Theorem 4.2.

Proposition 4.5 *In the model of equations (1) and (2) where Assumptions 1, 2, 4, 6 and 7 hold, bounds for the distribution of the structural function are identified from the set of observables $\{X_1, X_2, Y, Z\}$ for all $x \in \mathcal{X}$:*

$$\int_{\mathcal{Z}} \frac{\mathcal{A}_{\mathbb{1}(Y \leq y)}(x, z)}{\mathcal{A}_1(x, z)} F_W(\mathrm{d}w) \leq \mathcal{G}(y, x) \leq \int_{\mathcal{Z}} \frac{\mathcal{A}_{\mathbb{1}(Y \leq y)}(x, z)}{\mathcal{A}_1(x, z)} F_W(\mathrm{d}w) + \tilde{p}(x)$$

and these bounds are sharp. Furthermore, suppose that $B_l \leq g(x, u) \leq B_u$ for all x in the support of X and u in the support of U. Then, bounds for the average structural function are identified:

$$\int_{\mathcal{Z}} \frac{\mathcal{A}_Y(x, z)}{\mathcal{A}_1(x, z)} F_W(\mathrm{d}w) + B_l \tilde{p}(x) \leq \bar{g}(x) \leq \int_{\mathcal{Z}} \frac{\mathcal{A}_Y(x, z)}{\mathcal{A}_1(x, z)} F_W(\mathrm{d}w) + B_u \tilde{p}(x)$$

and these bounds are sharp.

5 Identification with One-dimensional Heterogeneity in the Structural Equation

Previous results for the quantile structural effects are comparable to those in the nonseparable model with a scalar heterogeneity. We now discuss how the quantile structural effects can be identified only through the information on the model of the structural equation (1) (i.e., without modeling the relationship between X and Z through the reduced-form equation (2)). Assume U is a scalar unobservable and $g(X, U)$ is strictly monotone in U. We then have the quantile structural effects $\vartheta^\tau(x) = g(X = x, q_\tau)$ where q_τ is the τth quantile of the marginal distribution of U, which have been extensively studied in the literature (see, e.g., Chernozhukov and Hansen (2005), Chernozhukov, Imbens, and Newey (2007), Horowitz and Lee (2007), Komunjer and Santos (2010), Gagliardini and Scaillet (2012), and Torgovitsky (2015)). By the independence of U and Z, the conditional quantile restriction yields that for each τ with $0 < \tau < 1$,

$$\begin{aligned}
\mathbb{E}[\mathbb{1}(Y < g(X, q_\tau)) \mid Z] &= \mathbb{E}[\mathbb{E}[\mathbb{1}(g(X, U) < g(X = x, q_\tau)) \mid X, Z \mid Z] \\
&= \mathbb{E}[\mathbb{E}[\mathbb{1}(U < q_\tau) \mid X, Z \mid Z] = \mathbb{E}[\mathbb{1}(U < q_\tau) \mid Z] = \mathbb{E}[\mathbb{1}(U < q_\tau)] \\
&= \tau.
\end{aligned}$$

Then the quantile structural effects can be recovered by a conditional moment restriction such as

$$\mathbb{E}[\rho_\tau(Y, X, \vartheta_0^\tau)) \mid Z] = 0$$

where $\rho_\tau(Y, X, \vartheta^\tau)) \equiv \mathbb{1}(Y < \vartheta^\tau(X)) - \tau$. Thus identification of the quantile structural effects depends on identification of the conditional distribution of the endogenous variables conditioning on the instruments, which is unobservable since the true X is observed with the measurement errors. Define an integral operator $\tilde{\mathcal{G}}(y, z)$ by

$$\tilde{\mathcal{G}}(y, z) \equiv \int F_{Y|X,Z}(y \mid x, z) F_{X|Z}(\mathrm{d}x \mid z) = \int \mathbb{E}[\mathbb{1}(Y \le y) \mid X = x, Z = z] F_{X|Z}(\mathrm{d}x \mid z)$$

for all $z \in \mathcal{Z}$. We thus have

$$\tilde{\mathcal{G}}(y, z) = \int \frac{\mathcal{A}_{\mathbb{1}(Y \le y)}(x, z)}{\mathcal{A}_1(x, z)} F_{X|Z}(\mathrm{d}x \mid z)$$

where $\mathcal{A}_1(x, z)$, $\mathcal{A}_{\mathbb{1}(Y \le y)}(x, z)$, and $F_{X|Z}(x \mid z)$ are identified by Theorem 4.2. The quantile structural effects ϑ^τ is then a solution of the nonlinear functional equation:

$$\tilde{\mathcal{G}}(\vartheta^\tau, z) = \tau.$$

We formally states sufficient conditions which include global identification of ϑ_0^τ on the parameter space Θ.

Assumption 8 *The structural equation (1) satisfy the following conditions:*
(i) The endogenous regressors $X \mid Z = z$ and instrumental variables Z are continuously distributed in $\mathbb{R}$ for all z;
(ii) We have $Z \perp\!\!\!\perp U$;
(iii) The function $g(X, U)$ is strictly monotone in U for all X where U is a continuously distributed scalar with strictly increasing $F_U(u)$ on the support of U;
(iv) We have $\tilde{\mathcal{G}}(\vartheta^\tau, z) - \tau = 0$, $\vartheta^\tau \in \Theta$, if and only if $\vartheta^\tau = \vartheta_0^\tau$.

We have thus shown the following result:

Lemma 5.1 *In the model of equation (1) in the absence of the reduced-form equation (2) where Assumptions 2, 4, 7 and 8 hold, the quantile structural function $\vartheta^\tau(x)$ is identified from the set of observables $\{X_1, X_2, Y, Z\}$ for all $x \in \mathcal{X}$.*

One might want to model both the structural equation and reduced-form equation and impose one-dimensional heterogeneity in both equations. This additional information could provide broader identification results. For instance, Torgovitsky (2015) additionally imposes Assumption 1 (ii)-(iii) and allows for a model with binary or discrete instrumental variables, which is excluded by

Assumption 8 (i) (see also D'Haultfœuille and Février (2015)). This is an interesting extension of our analysis in a way to relax the full support condition for the instrumental variables. Based on the results above and similar arguments in Torgovitsky (2015), we get the following result:

Lemma 5.2 *In the model of equations (1) and (2) where the endogeneous regressors $X \mid Z = z$ are continuously distributed in $\mathbb{R}$ for all z and where Assumptions 1 (ii)-(iii), 2, 4, 7, and 8(iii) hold, the quantile structural function $\vartheta^{\tau}(x)$ is identified from the set of observables $\{X_1, X_2, Y, Z\}$ for all $x \in \mathcal{X}$.*

This lemma implies that the availability of both structural and reduced-form equations and restrictions on the dimensions of the heterogeneity provide useful informational contents for the identification of the model where instrumental variables have small support.[14]

Table 1 summarizes the restrictions on the dimensions of the unobserved heterogeneity and the shapes of the unknown functions for each approaches. The control variables approach gets the identification results by utilizing both the structural and reduced-from equations. It allows for the multi-dimensional unobserved heterogeneity in the structural equation, but imposes the one-dimensional heterogeneity in the reduced-form equation and the monotonicity of the function in the unobserved heterogeneity. The instrumental variables approach obtains informational contents from the additional restriction of the monotonicity of the reduced-form equation in both the instrumental variables and unobserved heterogeneity. On the other hand, the approach from conditional moment restrictions only uses the structural equation and thus needs to impose the one-dimensional heterogeneity in the equation and the monotonicity of the equation in the unobserved heterogeneity. For the identification from these three approaches, the instrumental variables need to be continuous. However, when the reduced-form equation is available and one-dimensional heterogeneity and monotonicity are imposed on the equation, we can allow for binary or discrete instrumental variables as in the approach from the instrumental variables with small support. Therefore, restrictions on the dimensions and shapes can provide a great deal of informational contents for identification, and available economic knowledge on the models of interest are important to decide which approach should be selected.

6 An Application

6.1 Data Description and Selection

This section considers an application to the estimation of the nonseparable consumer demand systems (Engel curves). The dependent variable is the budget share of a commodity and the

[14]In a nonseparable model with endogeneity where binary proxies for the unobserved heterogeneity are available, Williams (2015) provides identification results for the model in the limit as the number of proxies increases. The approach could be useful when excluded instrumental variables are not accessible but only binary proxies are available.

regressor is the log of total family expenditure. The log of total family income is used as an instrumental variable to control for the endogeneity. We estimate the consumer demand systems using the Panel Study of Income and Dynamics (PSID). The PSID data were collected annually until the 1996 wave and biennially starting with the 1997 wave. Since 1999, the PSID began collecting information on a larger number of commodities, besides food. Starting with the 2005 wave, a few additional categories of commodities such as clothing and leisure were gathered. So we primarily focus on the recent year 2009 and use total family expenditure in year 2011 as a repeated measurement to control for the measurement errors.

We drop the Latino, Immigrant, and SEO subsamples. The data are a subset of households where the head is aged between 20 and 65 and the head or spouse is employed. We exclude households with child. This leaves us with final sample of 1,006 observations. The PSID data does not provide total family expenditure. We calculate the total expenditure from the sum of food (food at home, food away from home, and food stamps), home owner insurance, electricity, heating, water, other utilities (phone and cable), car insurance, car repairs, gasoline, parking (and car pool), bus fares, taxi fares, other transportation, school tuition, other school related expenses, child care, health (health insurance and out-of-pocket health expenses), rent, clothing, and leisure (trip and recreation expenses). The total expenditure and income are deflated by dividing them by the CPI index and are transformed into the natural logarithm.

Table 2 reports descriptive statistics. Log total expenditure of 2009 and 2011 waves share very similar mean and standard deviation. Except rent share, food at home share is the highest proportion of the total expenditure. Electricity, health, and leisure are the next top high shares. We conjecture that most of the commodities suffer from measurement errors which in turn boost measurement errors on the total expenditure.

Figure 1 plots kernel estimates of densities of the family expenditure in 2009 and 2011 waves. The kernel densities are estimated using an Epanechnikov kernel with bandwidths selection from Silverman's rule of thumb. Two measurements of the family expenditure are almost identical, which shows that high number of households have little change in expenditure over two years. This supports that we can use two repeated measurements to control for the measurement errors on the family expenditure. Figure 2 plots kernel estimates of densities of the family income in 2009 wave and the average family expenditure (sample average of 2009 and 2011 waves). They show a similar shape of the patterns. They also get a strong positive correlation which is 0.6935. We thus assume that the family income is a legitimate instrumental variable for the endogenous family expenditure.

In the application, we estimate the average structural function (ASF) when Y is the share of expenditure on a commodity. X_1 and X_2 are the log total expenditures of 2009 and 2011 waves, respectively. Z is the log total income of 2009. All variables are standardized by subtracting the mean and dividing by the standard deviation, which are reported in Table 2. Three estimators are considered. Our proposed estimator uses data $\{Y, Z, X_1, X_2\}$ (see Section C of the Supplementary

Material for details of the estimator). The control variables and instrumental variables estimators[15] use data $\{Y, Z, \bar{X}\}$ where $\bar{X} \equiv (X_1+X_2)/2$. Thus these naive estimators simply take sample average of two mismeasured expenditures for the sake of a better measurement. They also control for the endogeneity of the family expenditure using the family income of 2009 as the instrumental variable. In all estimators, optimal bandwidths are selected in a way that the estimated ASF is not sensitive to small changes of the bandwidths.

6.2 Results

We estimate demand curves for six commodities such as food, electricity, other utilities, gasoline, health, and leisure. Figures 3–8 plot the ASFs from three estimators over the range of the standardized log total expenditure from -2 to 2. Estimated ASFs from the measurement error-corrected (MEC), instrumental variables (IV), and control variables (CV) estimators are presented. 95% confidence bands which are constructed by the bootstrap method are also reported.[16] The estimated results for the demand curves are varying across different estimation methods.

The estimated values for the ASF show the measurement errors on the family expenditure cannot be neglected. For instance, in Figure 4, electricity demand from the MEC estimator shows a reversed S-shape. On the other hand, the estimates from the IV and CV estimators are very close to each other and they present linear patterns. In Figure 5, gasoline demand from the MEC estimator shows a flat reversed U-shape over the level of the family expenditure, but other two inconsistent estimators get steeper reversed U-shapes. Estimates of health demand from the inconsistent IV and CV estimators in Figure 6 are decreasing at the very top of the family expenditure, showing a reversed U-shape. However, estimated health demand from the MEC estimator shows a S-shape. Given that all three estimators control for the endogeneity of the family expenditure, the discrepancy between estimators would come from the incorrect measurement of the family expenditure. Thus, the results show that taking into account both the endogeneity and measurement errors in the family expenditure is substantial in analyzing the consumer demand systems.[17]

6.3 Sensitivity Analyses

We conduct a sensitivity analysis by exchanging the roles of X_1 and X_2. Assumption 4 imposes asymmetric restrictions on the measurement errors in two measurements. In particular, part (ii) requires stronger restriction than part (i). From this analysis, we can check if the asymmetry

[15] For brevity, we do not present two estimators here. See Imbens and Newey (2009) and Kasy (2014) for detailed discussions on the estimators.

[16] We need to derive a pointwise asymptotic distribution of the ASF for justifying the bootstrap, but it is beyond the scope of the paper. See Blundell, Chen, and Kristensen (2007) who similarly adopt the bootstrap.

[17] To address the possibility that the dependent variable, commodity share, is also mismeasured because of mismeasured denominator of the share, we have multiplied both sides by the family expenditure and estimated the average structural function. This resulting equation also satisfies the conditions imposed and the modification did not alter the qualitative conclusion.

between the assumptions on the two measurements is essential in analyzing the PSID. The estimated results are presented in Figure 9. It shows that estimated ASFs for all six commodities from the MEC estimator are very close to those in Figures 3–8. This implies that the asymmetry is not necessary to obtain the same substantive conclusion in the 2009 wave of the PSID and it is indeed imposed to derive our theoretical identification results under the weakest possible assumptions. So one could get more efficient estimates of ASFs by constructing a weighted average of the two results.

We perform an additional sensitivity analysis by estimating ASF in different years such as 2007 and 2005 waves (presented in Section F of the Supplementary Material). The results are qualitatively similar to those in the 2009 wave. In both years, estimated ASFs in six commodities from the MEC estimator are apparently different than those from the IV and CV estimators which only control for the endogeneity. Thus, these sensitivity analyses confirm that the proposed estimator is useful to correct for both the endogeneity and measurement errors in the family income from the PSID data.

7 Conclusion

This paper studies nonparametric identification and estimation of models with the nonseparable triangular simultaneous equations when the true endogenous variables are mismeasured. We impose weak restrictions on the measurement errors and provide a unified framework to identify the structural parameters, which can be applied to popular methods such as the instrumental variables and control variables approaches. We document that restrictions on the unobserved heterogeneity and shapes of the unknown functions contain useful informational contents for the identification. Our identification strategy is comprehensive but simple, so various effects such as the average effects, marginal effects, and quantile structural effects can be recovered. We provide bounds on the structural effects when the nonparametric rank condition is not satisfied. We propose easy-to-implement estimators which have closed-form expressions and require no numerical optimization, and derive their uniform convergence rates. An application to estimating the consumer demand systems using the PSID data confirms practical usefulness of the proposed methods.

There are many promising avenues for future work. First, general conditions for the nonparametric identification in systems of equations are given in Matzkin (2008) and Matzkin (2015). Further research is required to study what kind of restrictions should be imposed when true endogenous variables are unobservable due to measurement errors. Second, this paper focuses on the point identification of the structural effects. It would be a natural extension to provide conditions for partial identification of the structural parameters in the spirit of Manski (2007), when some of the necessary conditions for the point identification are relaxed or removed, and develop methods for hypothesis testing within the partial identification framework (see, e.g., Santos (2012)).

A Proofs

Recall $\nabla_\alpha^\lambda \equiv (\partial^\lambda/\partial\alpha^\lambda)$ denotes a partial derivative of degree $\lambda \in \{0, ..., \Lambda\}$ for $\Lambda \in \overline{\mathbb{N}}$ with respect to a generic variable α. And also recall $D_{\zeta a} \equiv a e^{\mathrm{i}\zeta X_2}$ and

$$\varphi_a(\zeta, v) \equiv \frac{\mathbb{E}[D_{\zeta a} \mid V = v]}{\mathbb{E}[D_{\zeta 1} \mid V = v]}, \qquad \varphi_a(\zeta, z) \equiv \frac{\mathbb{E}[D_{\zeta a} \mid Z = z]}{\mathbb{E}[D_{\zeta 1} \mid Z = z]}$$

for each ζ and $a \in \mathbf{a} \equiv \{1, T, X_1\}$.

Lemma A.1 *We have $\varepsilon_2 \perp\!\!\!\perp (X, Z)$ if and only if $\varepsilon_2 \perp\!\!\!\perp Z$ and $\varepsilon_2 \perp\!\!\!\perp X \mid Z$.*

Proof of Lemma A.1 To show the 'if' part, we note

$$\begin{aligned}
\mathbb{P}(\varepsilon_2 \le e_2, X \le x, Z \le z) =& \mathbb{P}(\varepsilon_2 \le e_2 \mid X \le x, Z \le z)\mathbb{P}(X \le x \mid Z \le z)\mathbb{P}(Z \le z) \\
=& \mathbb{P}(\varepsilon_2 \le e_2 \mid Z \le z)\mathbb{P}(X \le x \mid Z \le z)\mathbb{P}(Z \le z) \\
=& \mathbb{P}(\varepsilon_2 \le e_2)\mathbb{P}(X \le x \mid Z \le z)\mathbb{P}(Z \le z) \\
=& \mathbb{P}(\varepsilon_2 \le e_2)\mathbb{P}(X \le x, Z \le z),
\end{aligned}$$

where definition of conditional distribution are used in the first and fourth equalities, $\varepsilon_2 \perp\!\!\!\perp X \mid Z$ is used in the second equality, and $\varepsilon_2 \perp\!\!\!\perp Z$ is used in the third equality. To show the 'only if' part, we observe that $\varepsilon_2 \perp\!\!\!\perp (X, Z)$ straightforwardly implies $\varepsilon_2 \perp\!\!\!\perp Z$. Moreover, we note that $\varepsilon_2 \perp\!\!\!\perp (X, Z)$ implies $\varepsilon_2 \perp\!\!\!\perp (X, Z) \mid Z$ by Lemma 4.2 (ii) in Dawid (1979) and $\varepsilon_2 \perp\!\!\!\perp (X, Z) \mid Z$ implies $\varepsilon_2 \perp\!\!\!\perp X \mid Z$ by the symmetry property of the conditional independence and Lemma 4.2 (i) in Dawid (1979). As a result, the claim follows immediately. *Q.E.D.*

Lemma A.2 *Under Assumption 4, for any $z \in \mathcal{Z}$ we have*

$$\mathbb{E}[e^{\mathrm{i}\zeta X} \mid Z = z] = \exp\left(\int_0^\zeta \mathrm{i}\varphi_{X_1}(\xi, z)\mathrm{d}\xi\right),$$

provided that $\mathbb{E}[D_{\xi 1} \mid Z = z]$ is nonvanishing for any real ξ and $z \in \mathcal{Z}$.

Proof of Lemma A.2 We note that $\mathbb{E}[D_{\xi 1} \mid Z = z] \equiv \mathbb{E}[e^{\mathrm{i}\xi X_2} \mid Z = z] = \mathbb{E}[e^{\mathrm{i}\xi X} \mid Z = z]\mathbb{E}[e^{\mathrm{i}\xi\varepsilon_2} \mid Z = z]$ by $\varepsilon_2 \perp\!\!\!\perp X \mid Z$. Thus the nonvanishing $\mathbb{E}[D_{\xi 1} \mid Z = z]$ guarantees that both $\mathbb{E}[e^{\mathrm{i}\xi X} \mid Z = z]$ and $\mathbb{E}[e^{\mathrm{i}\xi\varepsilon_2} \mid Z = z]$ are nonvanishing. Then we have

$$\begin{aligned}
\mathbb{E}[e^{\mathrm{i}\zeta X} \mid Z = z] =& \exp(\ln(\mathbb{E}[e^{\mathrm{i}\zeta X} \mid Z = z] - \ln 1) \\
=& \exp\left(\int_0^\zeta \nabla_\xi^1 \ln(\mathbb{E}[e^{\mathrm{i}\xi X} \mid Z = z])\mathrm{d}\xi\right)
\end{aligned}$$

$$
\begin{aligned}
&= \exp\left(\int_0^\zeta \frac{\mathrm{i}\mathbb{E}[Xe^{\mathrm{i}\xi X} \mid Z=z]}{\mathbb{E}[e^{\mathrm{i}\xi X} \mid Z=z]}\mathrm{d}\xi\right)\\
&= \exp\left(\int_0^\zeta \frac{\mathrm{i}\mathbb{E}[Xe^{\mathrm{i}\xi X} \mid Z=z]\mathbb{E}[e^{\mathrm{i}\xi\varepsilon_2} \mid Z=z]}{\mathbb{E}[e^{\mathrm{i}\xi X} \mid Z=z]\mathbb{E}[e^{\mathrm{i}\xi\varepsilon_2} \mid Z=z]}\mathrm{d}\xi\right)\\
&= \exp\left(\int_0^\zeta \frac{\mathrm{i}\{\mathbb{E}[Xe^{\mathrm{i}\xi(X+\varepsilon_2)} \mid Z=z]+\mathbb{E}[\mathbb{E}(\varepsilon_1 \mid X_2, Z)e^{\mathrm{i}\xi X_2} \mid Z=z]\}}{\mathbb{E}[e^{\mathrm{i}\xi(X+\varepsilon_2)} \mid Z=z]}\mathrm{d}\xi\right)\\
&= \exp\left(\int_0^\zeta \frac{\mathrm{i}\{\mathbb{E}[Xe^{\mathrm{i}\xi X_2} \mid Z=z]+\mathbb{E}[\mathbb{E}(\varepsilon_1 e^{\mathrm{i}\xi X_2} \mid X_2, Z) \mid Z=z]\}}{\mathbb{E}[e^{\mathrm{i}\xi X_2} \mid Z=z]}\mathrm{d}\xi\right)\\
&= \exp\left(\int_0^\zeta \frac{\mathrm{i}\{\mathbb{E}[Xe^{\mathrm{i}\xi X_2} \mid Z=z]+\mathbb{E}[\varepsilon_1 e^{\mathrm{i}\xi X_2} \mid Z=z]\}}{\mathbb{E}[e^{\mathrm{i}\xi X_2} \mid Z=z]}\mathrm{d}\xi\right)\\
&= \exp\left(\int_0^\zeta \frac{\mathrm{i}\mathbb{E}[X_1e^{\mathrm{i}\xi X_2} \mid Z=z]}{\mathbb{E}[e^{\mathrm{i}\xi X_2} \mid Z=z]}\mathrm{d}\xi\right)\\
&= \exp\left(\int_0^\zeta \frac{\mathrm{i}\mathbb{E}[D_{\xi X_1} \mid Z=z]}{\mathbb{E}[D_{\xi 1} \mid Z]}\mathrm{d}\xi\right)\\
&= \exp\left(\int_0^\zeta \mathrm{i}\varphi_{X_1}(\xi, z)\mathrm{d}\xi\right)
\end{aligned}
$$

where $\mathbb{E}(\varepsilon_1 \mid X_2, Z) = 0$ and $\varepsilon_2 \perp\!\!\!\perp X \mid Z$ are used in the fifth equality and the law of iterative expectation is used in the seventh equality. *Q.E.D.*

Proof of Theorem 3.1 We note that

$$
\begin{aligned}
\Psi_1(\zeta, z) &\equiv \mathbb{E}[e^{\mathrm{i}\zeta X} \mid Z=z]\\
&= \int \mathbb{E}[e^{\mathrm{i}\zeta X} \mid X=x, Z=z]f_{X|Z}(x \mid z)\mathrm{d}x\\
&= \int f_{X|Z}(x \mid z)e^{\mathrm{i}\zeta x}\mathrm{d}x
\end{aligned}
$$

which is the Fourier transform of $f_{X|Z}(x \mid z)$. We also note that

$$
\frac{1}{2\pi}\int \Psi_1(\zeta, z)e^{-\mathrm{i}\zeta x}\mathrm{d}\zeta
$$

is the inverse Fourier transform of $\Psi_1(\zeta, z)$ for $(x, z) \in \mathcal{X} \times \mathcal{Z}$. Then we get

$$
f_{X|Z}(x \mid z) = \frac{1}{2\pi}\int \Psi_1(\zeta, z)e^{-\mathrm{i}\zeta x}\mathrm{d}\zeta.
$$

Thus by the inversion theorem

$$
F_{X|Z}(x \mid z) = \frac{1}{2} + \frac{1}{2\pi}\int_0^\infty \frac{\Psi_1(-\zeta, z)e^{\mathrm{i}\zeta x} - \Psi_1(\zeta, z)e^{-\mathrm{i}\zeta x}}{\mathrm{i}\zeta}\mathrm{d}\zeta.
$$

We have for each ζ

$$\Psi_1(\zeta, z) = \exp\left(\int_0^{\zeta} \mathrm{i}\varphi_{X_1}(\xi, z)\mathrm{d}\xi\right)$$

which is immediate by Lemma A.2. *Q.E.D.*

Lemma A.3 *$X \perp\!\!\!\perp \varepsilon_2 \mid V$ if and only if $(X, V) \perp\!\!\!\perp \varepsilon_2 \mid V$.*

Proof of Lemma A.3 We note that $X \perp\!\!\!\perp \varepsilon_2 \mid Z$ if and only if $(X, Z) \perp\!\!\!\perp (\varepsilon_2, Z) \mid Z$ from Lemma 4.1 in Dawid (1979). We also note that $(X, Z) \perp\!\!\!\perp (\varepsilon_2, Z) \mid Z$ implies $(X, Z) \perp\!\!\!\perp \varepsilon_2 \mid Z$ from Lemma 4.2 (ii) in Dawid (1979). Thus the 'if' statement follows immediately. The 'only if' statement can be proved since $(X, Z) \perp\!\!\!\perp \varepsilon_2 \mid Z$ implies $Z \perp\!\!\!\perp \varepsilon_2 \mid Z$ and $X \perp\!\!\!\perp \varepsilon_2 \mid Z$. *Q.E.D.*

Lemma A.4 *Under Assumption 5 (i)–(ii), for any $v \in \mathcal{V}$ we have*

$$\mathbb{E}[e^{\mathrm{i}\zeta X} \mid V = v] = \exp\left(\int_0^{\zeta} \mathrm{i}\varphi_{X_1}(\xi, v)\mathrm{d}\xi\right),$$

provided that $\mathbb{E}[D_{\xi 1} \mid V = v]$ is nonvanishing for any real ξ and $v \in \mathcal{V}$.

Proof of Lemma A.4 The result is immediate from the proof of Lemma A.2 by replacing Z with V. *Q.E.D.*

Proof of Theorem 3.2 We note that

$$\begin{aligned}
\Psi_T(\zeta, v) &\equiv \mathbb{E}[Te^{\mathrm{i}\zeta X} \mid V = v] \\
&= \int \mathbb{E}[Te^{\mathrm{i}\zeta X} \mid X = x, V = v] f_{X|V}(x \mid v)\mathrm{d}x \\
&= \int \mathbb{E}[T \mid X = x, V = v] f_{X|V}(x \mid v)e^{\mathrm{i}\zeta x}\mathrm{d}x \\
&= \int \mathcal{A}_T(x, v)e^{\mathrm{i}\zeta x}\mathrm{d}x
\end{aligned}$$

which is the Fourier transform of $\mathcal{A}_T(x, v)$. Since

$$\frac{1}{2\pi}\int \Psi_T(\zeta, v)e^{-\mathrm{i}\zeta x}\mathrm{d}\zeta$$

is the inverse Fourier transform of $\Psi_T(\zeta, v)$ for $(x, v) \in \mathcal{X} \times \mathcal{V}$, we get

$$\mathcal{A}_T(x, v) = \frac{1}{2\pi}\int \Psi_T(\zeta, v)e^{-\mathrm{i}\zeta x}\mathrm{d}\zeta$$

where the integral is defined over the coordinate ζ. Thus $\mathcal{A}_T(x, v)$ is identified when $\Psi_T(\zeta, v)$ is identified.

We now prove

$$\Psi_T(\zeta,v) = \varphi_T(\zeta,v)\exp\left(\int_0^\zeta \mathrm{i}\varphi_{X_1}(\xi,v)\mathrm{d}\xi\right).$$

By the definition of $\varphi_a(\zeta,v)$, for each ζ, we have

$$\begin{aligned}
\varphi_T(\zeta,v)\exp\left(\int_0^\zeta \mathrm{i}\varphi_{X_1}(\xi,v)\mathrm{d}\xi\right) &= \frac{\mathbb{E}[D_{\zeta T}\mid V=v]}{\mathbb{E}[D_{\zeta 1}\mid V=v]}\exp\left(\int_0^\zeta \frac{\mathrm{i}\mathbb{E}[D_{\xi X_1}\mid V=v]}{\mathbb{E}[D_{\xi 1}\mid V=v]}\mathrm{d}\xi\right)\\
&= \frac{\mathbb{E}[Te^{\mathrm{i}\zeta X_2}\mid V=v]}{\mathbb{E}[e^{\mathrm{i}\zeta X_2}\mid V=v]}\mathbb{E}[e^{\mathrm{i}\zeta X}\mid V=v]\\
&= \frac{\mathbb{E}[Te^{\mathrm{i}\zeta X_2}\mid V=v]}{\mathbb{E}[e^{\mathrm{i}\zeta X}\mid V=v]\mathbb{E}[e^{\mathrm{i}\zeta\varepsilon_2}\mid V=v]}\mathbb{E}[e^{\mathrm{i}\zeta X}\mid V=v]\\
&= \frac{\mathbb{E}[\mathbb{E}[Te^{\mathrm{i}\zeta X_2}\mid X,\varepsilon_2,V]\mid V=v]}{\mathbb{E}[e^{\mathrm{i}\zeta\varepsilon_2}\mid V=v]}\\
&= \frac{\mathbb{E}[\mathbb{E}[T\mid X,\varepsilon_2,V]e^{\mathrm{i}\zeta X_2}\mid V=v]}{\mathbb{E}[e^{\mathrm{i}\zeta\varepsilon_2}\mid V=v]}\\
&= \frac{\mathbb{E}[\mathbb{E}[T\mid X,V]e^{\mathrm{i}\zeta X}e^{\mathrm{i}\zeta\varepsilon_2}\mid V=v]}{\mathbb{E}[e^{\mathrm{i}\zeta\varepsilon_2}\mid V=v]}\\
&= \frac{\mathbb{E}[\mathbb{E}[Te^{\mathrm{i}\zeta X}\mid X,V]e^{\mathrm{i}\zeta\varepsilon_2}\mid V=v]}{\mathbb{E}[e^{\mathrm{i}\zeta\varepsilon_2}\mid V=v]}\\
&= \frac{\mathbb{E}[\mathbb{E}[Te^{\mathrm{i}\zeta X}\mid X,V]\mid V=v]\mathbb{E}[e^{\mathrm{i}\zeta\varepsilon_2}\mid V=v]}{\mathbb{E}[e^{\mathrm{i}\zeta\varepsilon_2}\mid V=v]}\\
&= \mathbb{E}[\mathbb{E}[Te^{\mathrm{i}\zeta X}\mid X,V]\mid V=v]\\
&= \mathbb{E}[Te^{\mathrm{i}\zeta X}\mid V=v]\\
&\equiv \Psi_T(\zeta,v)
\end{aligned}$$

where Lemma A.4, $\varepsilon_2 \perp\!\!\!\perp X \mid V=v$, $\mathbb{E}[T\mid X,\varepsilon_2,V]=\mathbb{E}[T\mid X,V]$, $(X,V)\perp\!\!\!\perp \varepsilon_2\mid V$ (from Lemma A.3) and the law of iterated expectation are used in the second, third, sixth, eighth and tenth equalities, respectively. This completes the proof. *Q.E.D.*

Proof of Proposition 3.5 To show the bounds, we use a similar argument to the one in Imbens and Newey (2009). By $\mathbb{P}(g(x,U)\le y\mid V)\ge 0$ and equation 3, we have

$$\begin{aligned}
\mathcal{G}(y,x) &\equiv \mathbb{P}(g(x,U)\le y)\\
&= \int \mathbb{P}(g(x,U)\le y\mid V=v)F_V(\mathrm{d}v)\\
&\ge \int_{\mathcal{V}(x)} \mathbb{P}(g(x,U)\le y\mid V=v)F_V(\mathrm{d}v)\\
&= \int_{\mathcal{V}(x)} \mathbb{P}(g(x,U)\le y\mid X=x,V=v)F_V(\mathrm{d}v)
\end{aligned}$$

$$= \int_{\mathcal{V}(x)} \mathbb{E}[\mathbb{1}(Y \le y) \mid X = x, V = v] F_V(\mathrm{d}v)$$
$$= \int_{\mathcal{V}(x)} \frac{\mathcal{A}_{\mathbb{1}(Y \le y)}(x, v)}{\mathcal{A}_1(x, v)} F_V(\mathrm{d}v)$$

where $F_V(v)$ is identified from Theorem 3.1 and $\mathcal{A}_{1(Y \le y)}(x, v)$ and $\mathcal{A}_1(x, v)$ are identified from Theorem 3.2. By $\mathbb{P}(g(x, U) \le y \mid V) \le 1$ and $p(x) \equiv \int_{\mathcal{V} \cap \mathcal{V}(x)^c} F_V(\mathrm{d}v)$, we have

$$\begin{aligned}
\mathcal{G}(y, x) &= \int_{\mathcal{V}(x)} \mathbb{P}(g(x, U) \le y \mid V = v) F_V(\mathrm{d}v) + \int_{\mathcal{V} \cap \mathcal{V}(x)^c} \mathbb{P}(g(x, U) \le y \mid V = v) F_V(\mathrm{d}v) \\
&\le \int_{\mathcal{V}(x)} \mathbb{E}[\mathbb{1}(Y \le y) \mid X = x, V = v] F_V(\mathrm{d}v) + \int_{\mathcal{V} \cap \mathcal{V}(x)^c} F_V(\mathrm{d}v) \\
&= \int_{\mathcal{V}(x)} \frac{\mathcal{A}_{1(Y \le y)}(x, v)}{\mathcal{A}_1(x, v)} F_V(\mathrm{d}v) + p(x).
\end{aligned}$$

The bounds on $\mathcal{G}(y, x)$ are sharp since there is no restriction on $\mathbb{E}[\mathbb{1}(Y \le y) \mid X = x, V = v]$ imposed by the data.

To show the second part, recall that $\mu(x, v) \equiv \mathbb{E}[Y \mid X = x, V = v]$. We then have

$$\begin{aligned}
\int_{\mathcal{V}(x)} \mu(x, v) F_V(\mathrm{d}v) &= \int_{\mathcal{V}(x)} \mathbb{E}[Y \mid X = x, V = v] F_V(\mathrm{d}v) \\
&= \int_{\mathcal{V}(x)} \int g(x, u) F_{U|X,V}(\mathrm{d}u \mid X, V) F_V(\mathrm{d}v) \\
&= \int_{\mathcal{V}(x)} \int g(x, u) F_{U|V}(\mathrm{d}u \mid V) F_V(\mathrm{d}v)
\end{aligned}$$

by Assumption 3. Note that $p(x) \equiv \int_{\mathcal{V} \cap \mathcal{V}(x)^c} F_V(\mathrm{d}v)$ and $B_l \le g(x, u) \le B_u$. We then have

$$B_l p(x) \le \int_{\mathcal{V} \cap \mathcal{V}(x)^c} \int g(x, u) F_{U|V}(\mathrm{d}u \mid V) F_V(\mathrm{d}V) \le B_u p(x).$$

Thus, by adding up two equations, we get

$$\begin{aligned}
\int_{\mathcal{V}(x)} \mathbb{E}[Y \mid X = x, V = v] F_V(\mathrm{d}v) + B_l p(x) &\le \int \int_{\mathcal{V}} g(x, u) F_{U|V}(\mathrm{d}u \mid V) F_V(\mathrm{d}V) \\
&\le \int_{\mathcal{V}(x)} \mathbb{E}[Y \mid X = x, V = v] F_V(\mathrm{d}v) + B_u p(x).
\end{aligned}$$

Since

$$\bar{g}(x) = \int \int_{\mathcal{V}} g(x, u) F_{U|V}(\mathrm{d}u \mid V) F_V(\mathrm{d}V)$$

and

$$\mathbb{E}[Y \mid X = x, V = v] = \frac{\mathcal{A}_Y(x,v)}{\mathcal{A}_1(x,v)},$$

the conclusion follows by the identification of $F_V(v)$ from Theorem 3.1 and the identification of $\mathcal{A}_Y(x,v)$ and $\mathcal{A}_1(x,v)$ from Theorem 3.2. To show sharpness of these bounds, let $u = V$ and

$$g^l(x,u) = \begin{cases} \frac{\mathcal{A}_Y(x,V)}{\mathcal{A}_1(x,V)}, & V \in \mathcal{V}(x) \\ B_l, & V \in \mathcal{V}(x)^c \end{cases}$$

Then $\bar{g}(x) = \int_{\mathcal{V}(x)} \frac{\mathcal{A}_Y(x,v)}{\mathcal{A}_1(x,v)} F_V(\mathrm{d}v) + B_l p(x)$. Similarly, defining

$$g^u(x,u) = \begin{cases} \frac{\mathcal{A}_Y(x,V)}{\mathcal{A}_1(x,V)}, & V \in \mathcal{V}(x) \\ B_u, & V \in \mathcal{V}(x)^c \end{cases}$$

gives $\bar{g}(x) = \int_{\mathcal{V}} \frac{\mathcal{A}_Y(x,v)}{\mathcal{A}_1(x,v)} F_V(\mathrm{d}v) + B_u p(x)$. *Q.E.D.*

Proof of Theorem 3.6 We note that for each $\lambda \in \{0, ..., \Lambda\}$

$$\begin{aligned}
(-\mathrm{i}\zeta)^\lambda \Psi_T(\zeta, v) &\equiv (-\mathrm{i}\zeta)^\lambda \mathbb{E}[Te^{\mathrm{i}\zeta X} \mid V = v] \\
&= (-\mathrm{i}\zeta)^\lambda \int \mathbb{E}[Te^{\mathrm{i}\zeta X} \mid X = x, V = v] f_{X|V}(x \mid v)\mathrm{d}x \\
&= (-\zeta)^\lambda \int \mathbb{E}[T \mid X = x, V = v] f_{X|V}(x \mid v) \nabla_x^\lambda e^{\mathrm{i}\zeta x}\mathrm{d}x \\
&= \int \nabla_x^\lambda (\mathbb{E}[T \mid X = x, V = v] f_{X|V}(x \mid v)) e^{\mathrm{i}\zeta x}\mathrm{d}x \\
&= \int \mathcal{A}_{T,\lambda}(x,v) e^{\mathrm{i}\zeta x}\mathrm{d}x
\end{aligned}$$

which is the Fourier transform of $\mathcal{A}_{T,\lambda}(x,v)$, where integral by parts is used in the fourth equality. Since

$$\frac{1}{2\pi} \int (-\mathrm{i}\zeta)^\lambda \Psi_T(\zeta, v) e^{-\mathrm{i}\zeta x}\mathrm{d}\zeta$$

is the inverse Fourier transform of $(-\mathrm{i}\zeta)^\lambda \Psi_T(\zeta, v)$ for $(x,v) \in \mathcal{X} \times \mathcal{V}$, we get

$$\mathcal{A}_{T,\lambda}(x,v) = \frac{1}{2\pi} \int (-\mathrm{i}\zeta)^\lambda \Psi_T(\zeta, v) e^{-\mathrm{i}\zeta x}\mathrm{d}\zeta.$$

By Lemma A.4 and a similar argument for the identification of $\Psi_T(\zeta, v)$ in the proof of Theorem 3.2, $\Psi_T(\zeta, v)$ is identified. This completes the proof. *Q.E.D.*

Proof of Theorem 4.1 Recall that $W = h^{-1}(x, \eta)$ denotes the inverse function for $h(z, \eta)$

in the first argument. We have for any integrable function $T(y)$

$$\begin{aligned}
\mathbb{E}[T(g(x,U))] &= \int \mathbb{E}[T(g(x,U)) \mid W = w]F_W(\mathrm{d}w) \\
&= \int \mathbb{E}[T(g(x,U)) \mid W = w, Z = z]F_W(\mathrm{d}w) \\
&= \int \mathbb{E}[T(g(x,U)) \mid X = x, Z = z]F_W(\mathrm{d}w) \\
&= \int \mathbb{E}[T(Y) \mid X = x, Z = z]F_W(\mathrm{d}w)
\end{aligned}$$

where $Z \perp\!\!\!\perp (U, \eta)$ is used in the second equality. The third equality follows by the fact that $W = w$ and $Z = z$ if and only if $X = x$ and $Z = z$, given that η is one-dimensional, invoking Corollary 2 of Hoderlein, Holzmann, Kasy, and Meister (2016). *Q.E.D.*

Proof of Theorem 4.2 The proof of the first part is similar to the proof of Theorem 3.2. Replacing V with Z yields the desired result, by Lemmas A.3 and A.4.

To prove the second part, note that since

$$\mathcal{A}_1(x,z) = f_{X|Z}(x \mid z) = \frac{1}{2\pi}\int \Psi_1(\zeta, z)e^{-\mathrm{i}\zeta x}\mathrm{d}\zeta,$$

we get

$$F_W(w) = F_{X|Z}(x \mid z) = \frac{1}{2} + \frac{1}{2\pi}\int_0^\infty \frac{\Psi_1(-\zeta, z)e^{\mathrm{i}\zeta x} - \Psi_1(\zeta, z)e^{-\mathrm{i}\zeta x}}{\mathrm{i}\zeta}\mathrm{d}\zeta$$

by equation (4) and the inversion theorem (e.g., Gil-Peraez (1951)). This completes the proof.

Q.E.D.

Proof of Proposition 4.5 To show the first part, we note that

$$\begin{aligned}
\mathcal{G}(y,x) &\equiv \mathbb{P}(g(x,U) \le y) \\
&= \int \mathbb{P}(g(x,U) \le y \mid W = w)F_W(\mathrm{d}w) \\
&= \int \mathbb{P}(Y \le y \mid X = x, Z = z]F_W(\mathrm{d}w) \\
&\ge \int_{\mathcal{Z}} \mathbb{P}(Y \le y \mid X = x, Z = z)F_W(\mathrm{d}w) \\
&= \int_{\mathcal{Z}} \mathbb{E}[\mathbb{1}(Y \le y) \mid X = x, Z = z]F_W(\mathrm{d}w)
\end{aligned}$$

by $\mathbb{P}(g(x,U) \le y \mid W) \ge 0$. Since $\mathcal{A}_1(x,z) = f_{X|Z}(x \mid z)$ and $\mathcal{A}_{\mathbb{1}(Y \le y)}(x,z) = \mathbb{E}[\mathbb{1}(Y \le y) \mid X =$

$x, Z = z]f_{X|Z}(x \mid z)$ from equation (5), We get

$$\mathbb{E}[\mathbb{1}(Y \leq y) \mid X = x, Z = z] = \frac{\mathcal{A}_{\mathbb{1}(Y \leq y)}(x, z)}{\mathcal{A}_1(x, z)}.$$

So we get

$$\mathcal{G}(y, x) \geq \int_{\mathcal{Z}} \frac{\mathcal{A}_{\mathbb{1}(Y \leq y)}(x, z)}{\mathcal{A}_1(x, z)} F_W(\mathrm{d}w)$$

where $\mathcal{A}_{\mathbb{1}(Y \leq y)}(x, z)$, $\mathcal{A}_1(x, z)$, and $F_W(w)$ are identified from Theorem 4.2.

We also note by $\mathbb{P}(g(x, U) \leq y \mid W) \leq 1$ and Theorem 4.2 that

$$\begin{aligned}
\mathcal{G}(y, x) &= \int_{\mathcal{Z}} \mathbb{P}(g(x, U) \leq y \mid W = w) F_W(\mathrm{d}w) + \int_{\mathcal{W} \cap \mathcal{Z}^c} \mathbb{P}(g(x, U) \leq y \mid W = w) F_W(\mathrm{d}w) \\
&\leq \int_{\mathcal{Z}} \mathbb{P}(Y \leq y \mid X = x, Z = z] F_W(\mathrm{d}w) + \int_{\mathcal{W} \cap \mathcal{Z}^c} F_W(\mathrm{d}w) \\
&= \int_{\mathcal{Z}} \mathbb{E}[\mathbb{1}(Y \leq y) \mid X = x, Z = z] F_W(\mathrm{d}w) + \tilde{p}(x) \\
&= \int_{\mathcal{Z}} \frac{\mathcal{A}_{\mathbb{1}(Y \leq y)}(x, z)}{\mathcal{A}_1(x, z)} F_W(\mathrm{d}w) + \tilde{p}(x).
\end{aligned}$$

The bounds on $\mathcal{G}(y, x)$ are sharp since there is no restriction on $\mathbb{E}[\mathbb{1}(Y \leq y) \mid X = x, Z = z]$ imposed by the data.

To show the second part, we note that

$$\int_{\mathcal{Z}} \mathbb{E}[Y \mid X = x, Z = z] F_W(\mathrm{d}w) = \int_{\mathcal{Z}} \int g(x, u) F_{U|X,Z}(\mathrm{d}u \mid X, Z) F_W(\mathrm{d}w)$$

by Assumption $Z \perp\!\!\!\perp (U, \eta)$. Since $\tilde{p}(x) \equiv \int_{\mathcal{W} \cap \mathcal{Z}^c} F_W(\mathrm{d}w)$ and $B_l \leq g(x, u) \leq B_u$, we get

$$B_l \tilde{p}(x) \leq \int_{\mathcal{W} \cap \mathcal{Z}^c} \int g(x, u) F_{U|X,Z}(\mathrm{d}u \mid X, Z) F_W(\mathrm{d}W) \leq B_u \tilde{p}(x).$$

Thus, by adding up two equations, it follows that

$$\begin{aligned}
\int_{\mathcal{Z}} \mathbb{E}[Y \mid X = x, Z = z] F_W(\mathrm{d}w) + B_l \tilde{p}(x) &\leq \int \int_{\mathcal{Z}} g(x, u) F_{U|X,Z}(\mathrm{d}u \mid X, Z) F_W(\mathrm{d}W) \\
&\leq \int_{\mathcal{Z}} \mathbb{E}[Y \mid X = x, Z = z] F_W(\mathrm{d}w) + B_u \tilde{p}(x).
\end{aligned}$$

Since

$$\bar{g}(x) = \int \int_{\mathcal{Z}} g(x, u) F_{U|X,Z}(\mathrm{d}u \mid X, Z) F_W(\mathrm{d}W)$$

and

$$\mathbb{E}[Y \mid X = x, Z = z] = \frac{\mathcal{A}_Y(x, z)}{\mathcal{A}_1(x, z)},$$

the conclusion follows by the identification of $\mathcal{A}_Y(x, z)$, $\mathcal{A}_1(x, z)$, and $F_W(w)$ in Theorem 4.2. To show sharpness of these bounds, let $u = Z$ and

$$g^l(x, u) = \begin{cases} \frac{\mathcal{A}_Y(x,Z)}{\mathcal{A}_1(x,Z)}, & Z \in \mathcal{Z} \\ B_l, & Z \in \mathcal{Z}^c \end{cases}$$

Then $\bar{g}(x) = \int_{\mathcal{Z}} \frac{\mathcal{A}_Y(x,z)}{\mathcal{A}_1(x,z)} F_W(\mathrm{d}w) + B_l \tilde{p}(x)$. Similarly, defining

$$g^u(x, u) = \begin{cases} \frac{\mathcal{A}_Y(x,Z)}{\mathcal{A}_1(x,Z)}, & Z \in \mathcal{Z} \\ B_u, & Z \in \mathcal{Z}^c \end{cases}$$

gives $\bar{g}(x) = \int_{\mathcal{Z}} \frac{\mathcal{A}_Y(x,z)}{\mathcal{A}_1(x,z)} F_W(\mathrm{d}w) + B_u \tilde{p}(x)$. *Q.E.D.*

References

Altonji, J. G., and R. L. Matzkin (2005): “Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors,” *Econometrica*, 73, 1053–1102.

Amemiya, T. (1985): “Instrumental Variable Estimator for the Nonlinear Error-in-variables Model,” *Journal of Econometrics*, 28, 273–289.

Banks, J., R. Blundell, and A. Lewbel (1997): “Quadratic Engel Curves and Consumer Demand,” *Review of Economics and Statistics*, 79, 527–539.

Berry, S., J. Levinsohn, and A. Pakes (1995): “Automobile Prices in Market Equilibrium,” *Econometrica*, 63, 841–890.

Blundell, R., M. Browning, and I. Crawford (2003): “Nonparametric Engel Curves and Revealed Preference,” *Econometrica*, 71, 205–240.

Blundell, R., X. Chen, and D. Kristensen (2007): “Semi-nonparametric IV Estimation of Shape-invariant Engel Curves,” *Econometrica*, 75, 1613–1669.

Blundell, R., and J. L. Powell (2003): “Endogeneity in Nonparametric and Semiparametric Regression Models,” *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, Vol. II (Cambridge: Cambridge University Press), 312–357.

Bonhomme, S., and J.-M. Robin (2010): “Generalized Non-Parametric Deconvolution with an Application to Earnings Dynamics,” *Review of Economic Studies*, 77, 491–533.

Bound, J., and A. Krueger (1991): “The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?,” *Journal of Labor Economics*, 9, 1–24.

Carneiro, P., J. J. Heckman, and E. Vytlacil (2010): “Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin,” *Econometrica*, 78, 377–394.

Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu (2006): “Measurement Error in Nonlinear Models,” Chapman & Hall.

Chen, X., H. Hong, and D. Nekipelov (2011): “Nonlinear Models of Measurement Errors,” *Journal of Economic Literature*, 49, 901–937.

Chen, X., H. Hong, and E. Tamer (2005): “Measurement Error Models with Auxiliary Data,” *Review of Economic Studies*, 72, 343–366.

Chernozhukov, V., and C. Hansen (2005): “An IV Model of Quantile Treatment Effect,” *Econometrica*, 73, 245–261.

Chernozhukov, V., G. Imbens, and W. Newey (2007): "Instrumental Variable Estimation of Nonseparable Models," *Journal of Econometrics*, 139, 4–14.

Chesher, A. (2005): "Nonparametric Identification under Discrete Variation," *Econometrica*, 73, 1525–1550.

Chiappori, P.-A., I. Komunjer, and D. Kristensen (2015): "Nonparametric Identification and Estimation of Transformation Models," *Journal of Econometrics*, 188, 22–39.

Dawid, A. P. (1979): "Conditional Independence in Statistical Theory," *Journal of the Royal Statistical Society, Series B*, 41, 1–31.

De Paula, A. (2008): "Conditional Moments and Independence," *The American Statistician*, 62, 219–221.

Delaigle, A., P. Hall, and A. Meister (2008): "On Deconvolution with Repeated Measurements," *Annals of Statistics*, 36, 665–685.

D'Haultfœuille, X., and P. Février (2015): "Identification of Nonseparable Triangular Models with Discrete Instruments," *Econometrica*, 83, 1199–1210.

Ebrahimi, N., G. G. Hamedani, E. S. Soofi, and H. Volkmer (2010): "A Class of Models for Uncorrelated Random Variables," *Journal of Multivariate Analysis*, 101, 1859–1871.

Engel, E. (1895): "Die Lebenskosten Belgischer Arbeiter-Familien Früher and Jetzt," *International Statistical Institute Bulletin*, 9, 1–74.

Evdokimov, K. (2010): "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity," *Unpublished manuscript.*

Fan, J. (1991): "On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems," *Annals of Statistics*, 19, 1257–1272.

Fan, J., and Y. K. Truong (1993): "Nonparametric Regression with Errors in Variables," *Annals of Statistics*, 21, 1900–1925.

Firpo, S. (2007): "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica*, 75, 259–276.

Firpo, S., N. M. Fortin, and T. Lemieux (2009): "Unconditional Quantile Regressions," *Econometrica*, 77, 953–973.

Florens, J. P., J. Heckman, C. Meghir, and E. Vytlacil (2008): "Identification of Treatment Effects using Control Functions in Models with Continuous, Endogenous Treatment and Heterogeneous Effects," *Econometrica*, 76, 1191–1206.

Gagliardini, P., and O. Scaillet (2012): "Nonparametric Instrumental Variable Estimation of Structural Quantile Effects," *Econometrica*, 80, 1533–1562.

Gandhi, A., K. i. Kim, and A. Petrin (2011): "Identification and Estimation in Discrete Choice Demand Models when Endogenous Variables Interact with the Error," *Unpublished manuscript.*

Gil-Peraez, J. (1951): "Note on the Inversion Theorem," *Biometrika*, 38, 481–482.

Griliches, Z. (1974): "Errors in Variables and Other Unobservables," *Econometrica*, 42, 971–998.

Gurland, J. (1948): "Inversion Formulae for the Distribution of Ratios," *Annals of Mathematical Statistics*, 19, 228–237.

Hausman, J. A. (1977): "Errors in Variables in Simultaneous Equation Models," *Journal of Econometrics*, 5, 389–401.

Hausman, J. A., W. K. Newey, H. Ichimura, and J. L. Powell (1991): "Identification and Estimation of Polynomial Errors-in-variables Models," *Journal of Econometrics*, 50, 273–295.

Heckman, J. J., R. L. Matzkin, and L. Nesheim (2010): "Nonparametric Identification and Estimation of Nonadditive Hedonic Models," *Econometrica*, 78, 1569–1591.

Heckman, J. J., and E. Vytlacil (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73, 669–738.

Hoderlein, S., H. Holzmann, M. Kasy, and A. Meister (2016): "Erratum Instrumental Variables with Unrestricted Heterogeneity and Continuous Treatment," *Review of Economic Studies (forthcoming).*

Hoderlein, S., and E. Mammen (2007): "Identification of Marginal Effects in Nonseparable Models Without Monotonicity," *Econometrica*, 75, 1513–1518.

Horowitz, J., and S. Lee (2007): "Nonparametric Instrumental Variables Estimation of a Quantile Regression Model," *Econometrica*, 75, 1191–1208.

Horowitz, J., and M. Markatou (1996): "Semiparametric Estimation of Regression Models for Panel Data," *Review of Economic Studies*, 63, 145–168.

Hsiao, C. (1989): "Consistent Estimation for Some Nonlinear Errors-in-variables Models," *Journal of Econometrics*, 41, 159–185.

Hu, Y., and Y. Sasaki (2015): "Closed-Form Estimation of Nonparametric Models with Non-Classical Measurement Errors," *Journal of Econometrics*, 185, 392–408.

Hu, Y., and S. M. Schennach (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76, 195–216.

Imbens, G. W., and W. K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77, 1481–1512.

Kasy, M. (2014): "Instrumental Variables with Unrestricted Heterogeneity and Continuous Treatment," *Review of Economic Studies*, 81, 1614–1636.

Khan, S., and E. Tamer (2010): "Irregular Identification, Support Conditions, and Inverse Weight Estimation," *Econometrica*, 78, 2021–2042.

Komunjer, I., and A. Santos (2010): "Semi-parametric Estimation of Non-separable Models: a Minimum Distance from Independence Approach," *Econometrics Journal*, 13, S28–S55.

Lewbel, A. (1996): "Demand Estimation with Expenditure Measurement Errors on the Left and Right Hand Side," *Review of Economics and Statistics*, 78, 718–725.

Li, T. (2002): "Robust and Consistent Estimation of Nonlinear Errors-in-variables Models," *Journal of Econometrics*, 110, 1–26.

Li, T., and Q. Vuong (1998): "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators," *Journal of Multivariate Analysis*, 65, 139–165.

Linton, O., and Y.-J. Whang (2002): "Nonparametric Estimation with Aggregated Data," *Econometric Theory*, 18, 420–468.

Liviatan, N. (1961): "Errors in Variables and Engel Curve Analysis," *Econometrica*, 29, 336–362.

Lukacs, E. (1970): "Characteristic Functions," Griffin.

Manski, C. (2007): "Identification for Prediction and Decision," Harvard University Press.

Matzkin, R. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339–1375.

——— (2008): "Identification in Nonparametric Simultaneous Equations Models," *Econometrica*, 76, 945–978.

——— (2015): "Estimation of Nonparametric Models with Simultaneity," *Econometrica*, 83, 1–66.

Olley, S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1298.

Powell, J., J. Stock, and T. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 1403–1430.

Rothe, C. (2010): "Nonparametric Estimation of Distributional Policy Effects," *Journal of Econometrics*, 155, 56–70.

——— (2012): “Partial Distributional Policy Effects,” *Econometrica*, 80, 2269–2301.

Santos, A. (2012): “Inference in Nonparametric Instrumental Variables with Partial Identification,” *Econometrica*, 80, 213–275.

Sasaki, Y. (2015): “What Do Quantile Regressions Identify for General Structural Functions,” *Econometric Theory*, 31, 1102–1116.

Schennach, S. M. (2004): “Estimation of Nonlinear Models with Measurement Error,” *Econometrica*, 72, 33–75.

Song, S., S. M. Schennach, and H. White (2015): “Estimating Nonseparable Models with Mismeasured Endogenous Variables,” *Quantitative Economics*, 6, 749–794.

Stoker, T. (1986): “Consistent Estimation of Scaled Coefficients,” *Econometrica*, 54, 1461–1481.

Torgovitsky, A. (2015): “Identification of Nonseparable Models Using Instruments with Small Support,” *Econometrica*, 83, 1185–1197.

Williams, B. (2015): “Identification of a Nonseparable Model under Endogeneity using Binary Proxies for Unobserved Heterogeneity,” *Unpublished manuscript*.

Working, H. (1943): “Statistical Laws of Family Expenditure,” *Journal of the American Statistical Association*, 38, 43–56.

Table 1: Summary of restrictions for each approaches

approach	*multi-dimensional*	*one-dimensional*	*monotone*
CV	U	η	$h(\cdot,\cdot)$ in η
IV	U	η	$h(\cdot,\cdot)$ in Z and η
CMR		U	$g(\cdot,\cdot)$ in U
IVSS		U and η	$g(\cdot,\cdot)$ in U and $h(\cdot,\cdot)$ in η

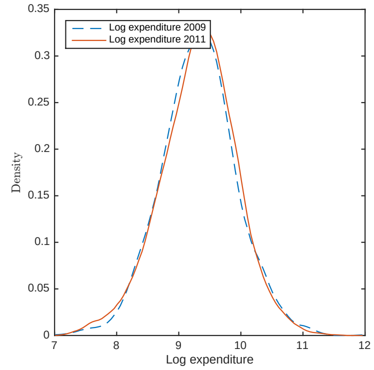
Note: The table summarizes restrictions for each approaches from Control Variables (CV), Instrumental Variables (IV), Conditional Moment Restrictions (CMR), and Instrumental Variables with Small Support (IVSS).

Table 2: Descriptive statistics in year 2009

Variable	Mean	Median	Std. Dev.	Min.	Max.
Log expenditure (2009)	9.3226	9.3196	0.6015	6.8503	12.1691
Log expenditure (2011)	9.3244	9.3475	0.6062	7.3491	12.7196
Log income	9.9497	10.0589	0.9181	5.8502	13.1580
Food-in share	0.1333	0.1239	0.0728	0.0000	0.4972
Food-out share	0.0602	0.0495	0.0486	0.0000	0.3123
Food stamp share	0.0044	0.0000	0.0371	0.0000	0.7229
Home insurance share	0.0139	0.0110	0.0165	0.0000	0.1204
Electricity share	0.0933	0.0751	0.0910	0.0000	0.8056
Heating share	0.0226	0.0147	0.0320	0.0000	0.4864
Water share	0.0098	0.0067	0.0122	0.0000	0.0986
Other utilities share	0.0614	0.0560	0.0358	0.0000	0.2571
Car insurance share	0.0395	0.0345	0.0314	0.0000	0.2700
Car repair share	0.0326	0.0000	0.0795	0.0000	0.7833
Gasoline share	0.0500	0.0414	0.0401	0.0000	0.3049
Parking and car pool share	0.0011	0.0000	0.0051	0.0000	0.0602
Transportation share	0.0087	0.0000	0.0322	0.0000	0.3233
Tuition share	0.0351	0.0000	0.0990	0.0000	0.7072
Other school expenses share	0.0008	0.0000	0.0090	0.0000	0.2037
Childcare share	0.0002	0.0000	0.0041	0.0000	0.1036
Health share	0.0758	0.0524	0.0812	0.0000	0.7725
Clothing share	0.0352	0.0234	0.0436	0.0000	0.6241
Leisure share	0.0713	0.0521	0.0748	0.0000	0.6206
Rent share	0.3126	0.3065	0.1375	0.0051	0.8103
Number of observations	1,006				

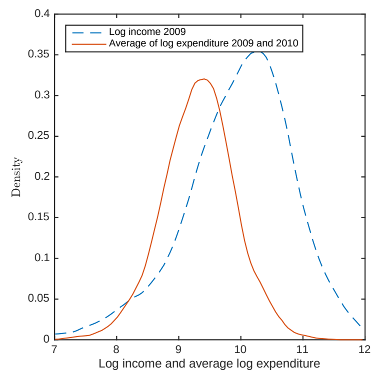
Note: The source of the data is the PSID. All commodity shares and the log income are in 2009 year wave. Two repeated measures of the log expenditure are in 2009 and 2011 waves.

Figure 1: Densities of log expenditure 2009 and 2011



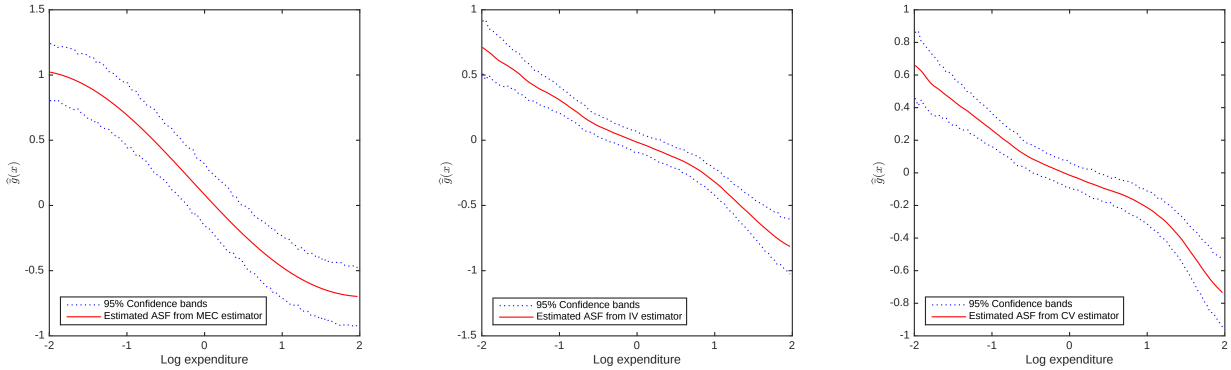
Note: The figure plots kernel estimates of the densities of log total expenditure in 2009 and 2011 waves. The kernel densities are estimated using an Epanechnikov kernel with the bandwidths selection from Silverman's rule of thumb.

Figure 2: Densities of the log income and average of the log expenditures in 2009 and 2011



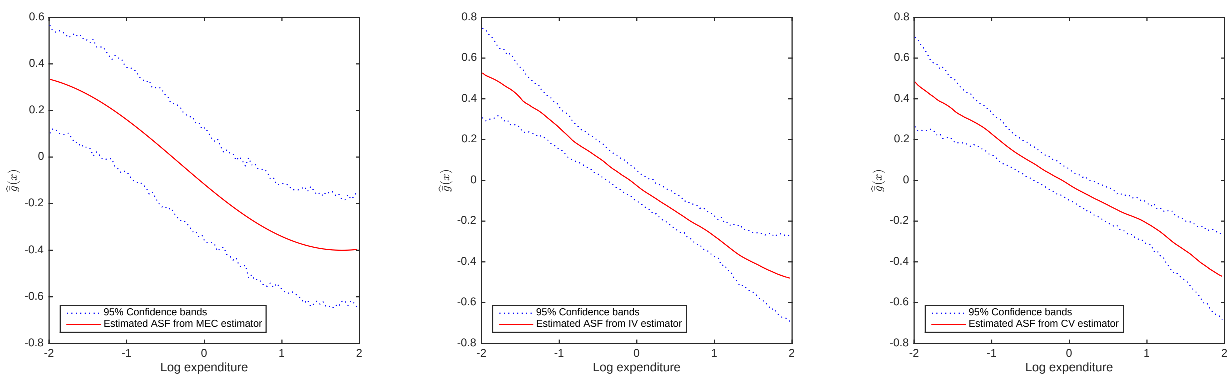
Note: The figure plots kernel estimates of the densities of log total income in 2009 and the average log total expenditure in 2009 and 2011 waves. The kernel densities are estimated using an Epanechnikov kernel with the bandwidths selection from Silverman's rule of thumb.

Figure 3: 2009 Food



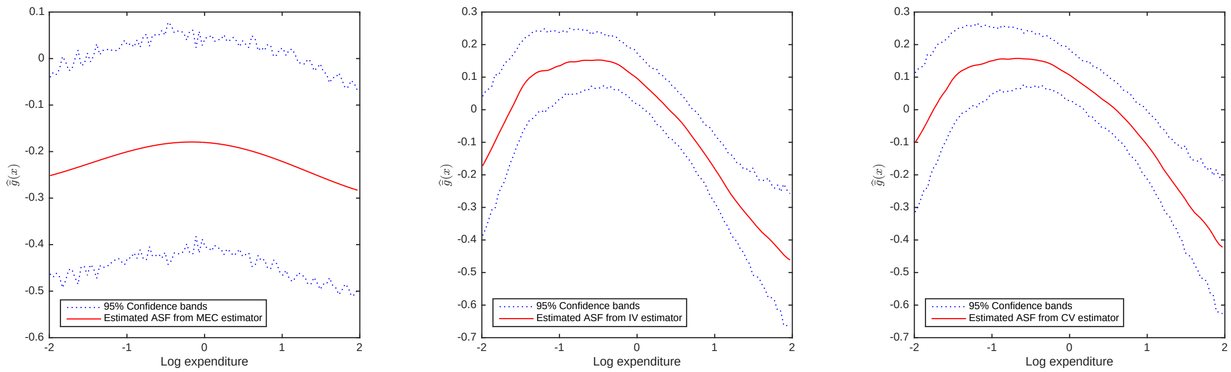
Note: The figures report the estimated average structural function (ASF) and confidence bands for the food share in year 2009. The food share is calculated as the sum of food-in, food-out, and food stamp. All variables are standardized by subtracting the mean and dividing by the standard deviation, which are reported in Table 2. The ASF is estimated by the measurement error-corrected (MEC), instrumental variables (IV), and control variables (CV) estimators. Confidence bands are constructed by performing the bootstrap method.

Figure 4: 2009 Electricity



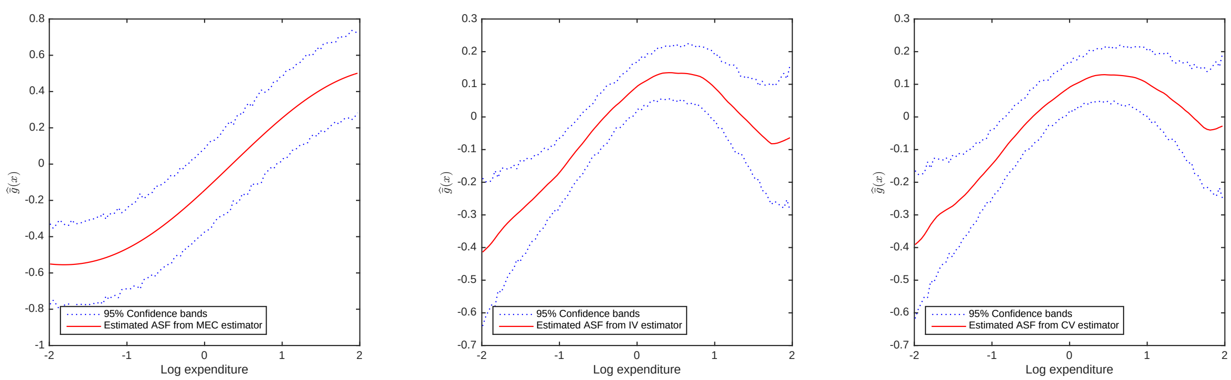
Note: The figures report the estimated average structural function (ASF) and confidence bands for the electricity share in year 2009. All variables are standardized by subtracting the mean and dividing by the standard deviation, which are reported in Table 2. The ASF is estimated by the measurement error-corrected (MEC), instrumental variables (IV), and control variables (CV) estimators. Confidence bands are constructed by performing the bootstrap method.

Figure 5: 2009 Gasoline



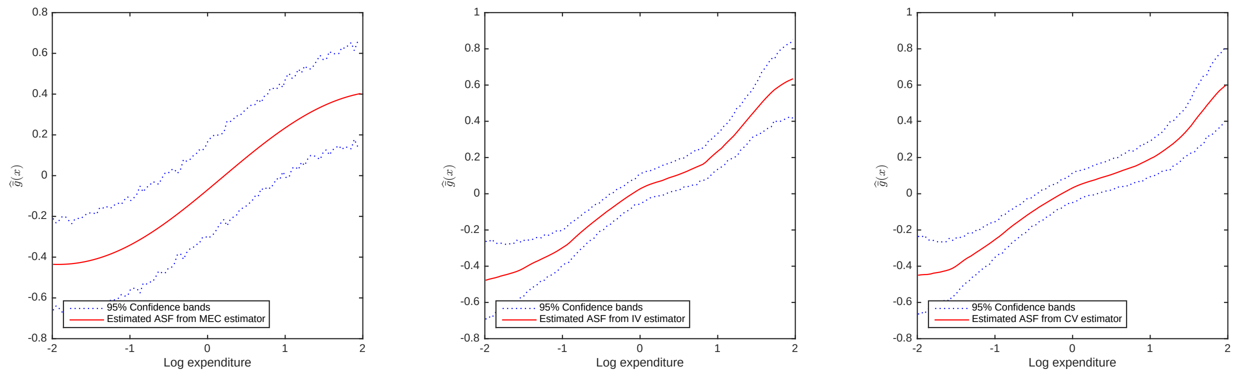
Note: The figures report the estimated average structural function (ASF) and confidence bands for the gasoline share in year 2009. All variables are standardized by subtracting the mean and dividing by the standard deviation, which are reported in Table 2. The ASF is estimated by the measurement error-corrected (MEC), instrumental variables (IV), and control variables (CV) estimators. Confidence bands are constructed by performing the bootstrap method.

Figure 6: 2009 Health



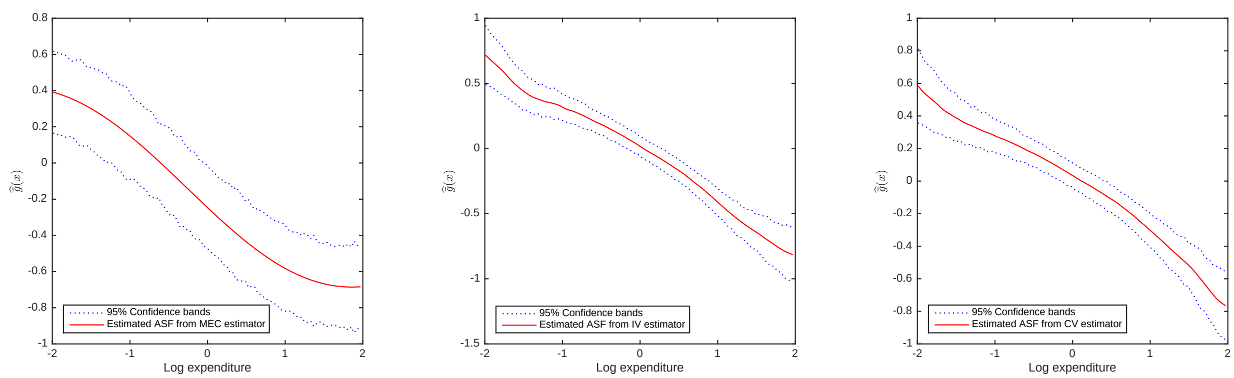
Note: The figures report the estimated average structural function (ASF) and confidence bands for the health share in year 2009. All variables are standardized by subtracting the mean and dividing by the standard deviation, which are reported in Table 2. The ASF is estimated by the measurement error-corrected (MEC), instrumental variables (IV), and control variables (CV) estimators. Confidence bands are constructed by performing the bootstrap method.

Figure 7: 2009 Leisure



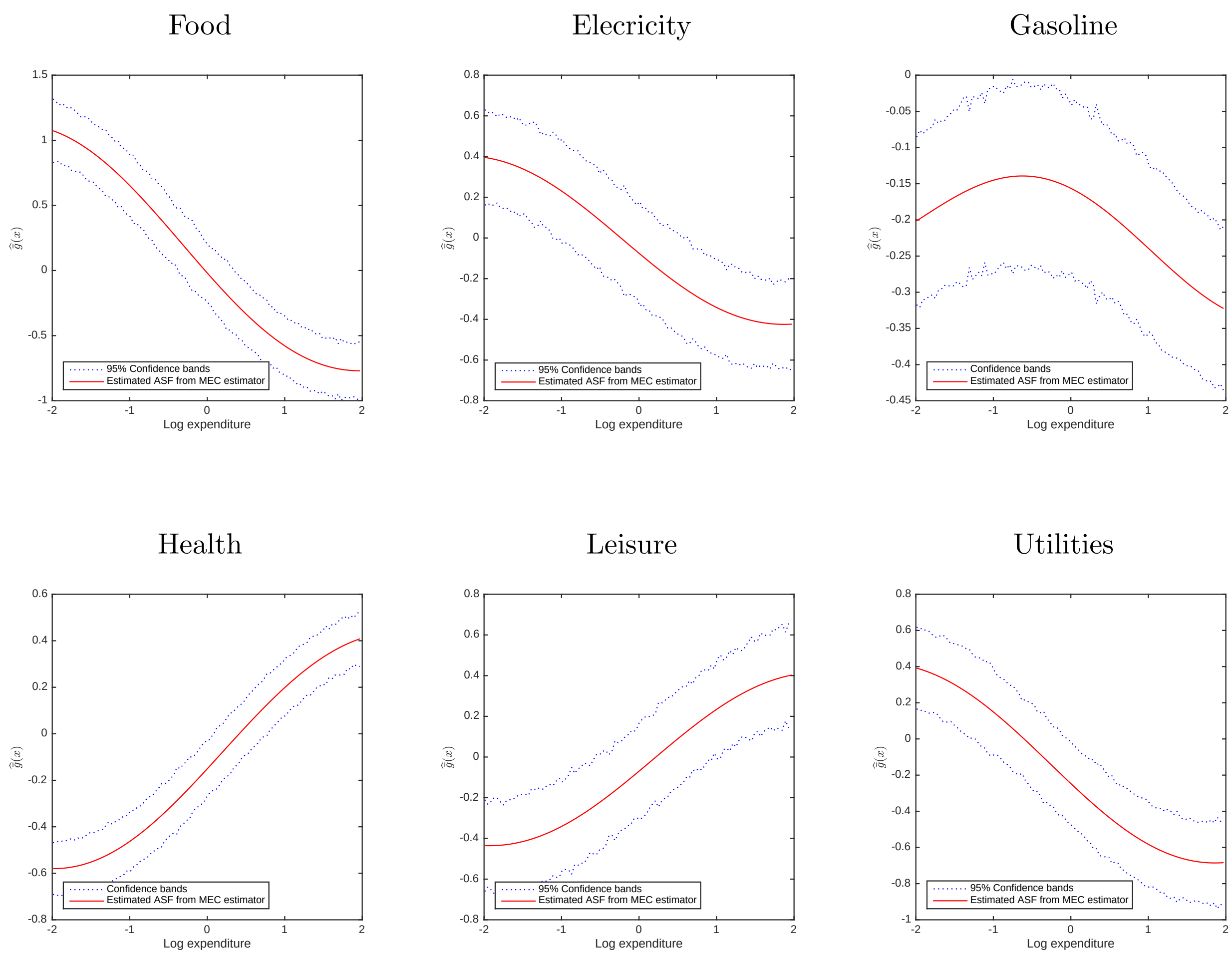
Note: The figures report the estimated average structural function (ASF) and confidence bands for the leisure share in year 2009. All variables are standardized by subtracting the mean and dividing by the standard deviation, which are reported in Table 2. The ASF is estimated by the measurement error-corrected (MEC), instrumental variables (IV), and control variables (CV) estimators. Confidence bands are constructed by performing the bootstrap method.

Figure 8: 2009 Utilities



Note: The figures report the estimated average structural function (ASF) and confidence bands for the other utilities share except electricity, heating, and water in year 2009. All variables are standardized by subtracting the mean and dividing by the standard deviation, which are reported in Table 2. The ASF is estimated by the measurement error-corrected (MEC), instrumental variables (IV), and control variables (CV) estimators. Confidence bands are constructed by performing the bootstrap method.

Figure 9: Sensitivity to exchanging X_1 and X_2



Note: The figures report the estimated average structural function (ASF) and confidence bands for the shares of the six commodities in year 2009 by the measurement error-corrected (MEC) estimator. The estimator is obtained by exchanging the roles of X_1 and X_2. All variables are standardized by subtracting the mean and dividing by the standard deviation, which are reported in Table 2. Confidence bands are constructed by performing the bootstrap method.