

# Nonparametric Errors in Variables Models with Measurement Errors on both sides of the Equation\*

Michele De Nadai  
University of New South Wales

Arthur Lewbel  
Boston College

First version July 2011, final version June 2015

## Abstract

Measurement errors are often correlated, as in surveys where respondent's biases or tendencies to err affect multiple reported variables. We extend Schennach (2007) to identify moments of the conditional distribution of a true  $Y$  given a true  $X$  when both are measured with error, the measurement errors in  $Y$  and  $X$  are correlated, and the true unknown model of  $Y$  given  $X$  has nonseparable model errors. After showing nonparametric identification, we provide a sieve generalized method of moments based estimator of the model, and apply it to nonparametric Engel curve estimation. In our application measurement errors on the expenditures of a good  $Y$  are by construction correlated with measurement errors in total expenditures  $X$ . This problem, which is present in many consumption data sets, has been ignored in most demand applications. We find that accounting for this problem casts doubt on Hildenbrand's (1994) "increasing dispersion" assumption.

**Keywords:** Engel curve; errors-in-variables model; Fourier transform; generalized function; sieve estimation.

**JEL Classification Numbers:** C10, C14, D12.

---

\*This paper benefited from helpful discussions with Erich Battistin and seminar participants at University of Mannheim and University of Padua. Address for correspondence: School of Economics, University of New South Wales, Australian School of Business, Sydney 2052, Australia and Department of Economics, 140 Commonwealth Avenue, 02467 Chestnut Hill, MA, USA. E-mail: m.denadai@unsw.edu.au and lewbel@bc.edu

# 1 Introduction

We consider identification and estimation of conditional moments of a dependent variable  $Y$  given a regressor  $X$  in nonparametric regression models where both  $Y$  and  $X$  are mismeasured, and the measurement errors in  $Y$  and  $X$  are correlated. For example, correlated measurement errors are likely in survey data where each respondent’s reporting biases or tendencies to err affect multiple variables that he or she self reports.

An example application that we consider empirically is consumer demand estimation, where  $Y$  is the quantity or expenditures demanded of some good or service, and  $X$  is total consumption expenditures on all goods. In most consumption data sets (e.g., the US Consumer Expenditure Survey or the UK Family Expenditure Survey), total consumption  $X$  is constructed as the sum of expenditures on individual goods, so by construction any measurement error in  $Y$  will also appear as a component of, and hence be correlated with, the measurement error in  $X$ . Similar problems arise in profit, cost, or factor demand equations in production, and in autoregressive or other dynamic models where sources of measurement error are not independent over time.

Our identification procedure allows us to distinguish measurement errors from other sources of error that are due to unobserved structural or behavioral heterogeneity. This is important in applications because many policies may depend on the distribution of structural unobserved heterogeneity, but not on measurement error. For example, the effects of an income tax on aggregate demand or savings depend on the distribution of income elasticities in the population. In contrast to our results, most empirical analyses implicitly or explicitly attribute either none or all of estimated model errors to unobserved heterogeneity.

In the consumer demand application, it has long been known that for most goods, empirical estimates of  $Var(Y|X)$  are increasing in  $X$ . For example, Hildenbrand (1994, figures 3.6 and 3.7) documents this phenomenon for a variety of goods in two different countries, calls it the “increasing dispersion” assumption, and exploits it as a behavioral feature that helps give rise to the aggregate law of demand. This property is also often used to justify estimating Engel curves in budget share instead of quantity form, to reduce the resulting error heteroskedasticity. However, in this paper we find empirically that while this phenomenon clearly holds in estimates of  $Var(Y|X)$  on raw data, after nonparametrically accounting for joint measurement error in  $Y$  and  $X$ , the evidence for

increasing dispersion becomes considerably weaker, suggesting that this well documented feature of Engel curve estimates may be in part an artifact of measurement errors rather than a feature of behavior.

Our identification strategy is an extension of Schennach (2007), who provides nonparametric identification of the conditional mean of  $Y$  given  $X$  (using instruments  $Q$ ) when  $X$  is a classically mismeasured regressor. We extend Schennach (2007) primarily by allowing for a measurement error term in  $Y$  that may be correlated with the measurement error in  $X$ . An additional extension is that we identify higher moments of the true  $Y$  given the true  $X$  instead of just the conditional mean. A further extension allows the measurement error in  $X$  to take a multiplicative form that is particularly well suited for our Engel curve application. Our proofs make use of recent machinery provided by Zinde-Walsh (2014).

Building on estimators like Newey (2001), Schennach (2007) bases identification on taking Fourier transforms of the conditional means of  $Y$  and of  $XY$  given instruments. Our main insight is that, if additive measurement errors in  $X$  and  $Y$  are correlated with each other but otherwise have some of the properties of classical measurement errors, then their presence will only affect the Fourier transform on a finite number of points, so identification will still be possible. Our further extensions exploit similar properties in different measurement error specifications, and our empirical application makes use of some special features of Engel curves to fully identify higher moments.

There is a large literature on the estimation of measurement error models. In addition to Schennach (2007), more recent work on measurement errors in nonparametric regression models includes Delaigle, Fan, and Carroll (2009), Rummel, Augustin, and Küchenhof (2010), Carroll, Chen, and Hu (2010), Meister (2011), and Carroll, Delaigle, and Hall (2011). Recent surveys containing many earlier references include Carroll, Ruppert, Stefanski, and Crainiceanu (2006) and Chen, Hong, and Nekipelov (2011).<sup>1</sup>

In the literature we find several examples of Engel curve estimation in the presence of measurement errors. Hausman, Newey, Ichimura, and Powell (1991) and Hausman, Newey, and Powell (1995) provide estimators for polynomial Engel curves with classically mismeasured  $X$ , Newey

---

<sup>1</sup>Earlier econometric papers closely related to Schennach (2007), but exploiting repeated measurements, are Hausman, Newey, Ichimura, and Powell (1991), Schennach (2004) and Li (2002). Most of these assume two mismeasures of the true  $X$  are available, one of which could have errors correlated with the measurement error  $Y$ .

(2001) estimates a nonpolynomial parametric Engel curve with mismeasured  $X$ , Blundell, Chen, and Kristensen (2007) estimate a semi-parametric model of Engel curves that allows  $X$  to be endogenous and hence mismeasured, and Lewbel (1996) identifies and estimates Engel curves allowing for correlated measurement errors in  $X$  and  $Y$  as we do, but does so in the context of a parametric model of  $Y$  given  $X$ .<sup>2</sup>

The conditional distribution of the true  $Y$  given the true  $X$  in Engel curves corresponds to the distribution of preference heterogeneity parameters in the population, which can be of particular interest for policy analysis. For example, consider the effect on demand of introducing a tax cut or tax increase that shifts households' total expenditure levels. This will in general affect the entire distribution of demand, not just its mean, both because Engel curves are generally nonlinear and because preferences are heterogeneous. Recovering moments of the distribution of demand is useful because many policy indicators, such as the welfare implication of a tax change, will in turn depend on more features of the distribution of demand than just its mean.

The next two sections show identification of the model with standard additive measurement error and of the specification more specifically appropriate for Engel curve data. We then describe our sieve based estimator, and provide a simulation study. After that is an empirical application to estimating food and clothing expenditures in US Consumer Expenditure Survey data, followed by conclusions and an appendix providing proofs.

## 2 Overview

Suppose that scalar random variables  $Y^*$  and  $X^*$  are measured with error, so we only observe  $Y$  and  $X$  where:

$$\begin{aligned} Y &= Y^* + S, \\ X &= X^* + W, \end{aligned}$$

---

<sup>2</sup>More generally, within econometrics there is a large recent literature on nonparametric identification of models having nonseparable errors (e.g., Chesher 2003, Matzkin 2007, Hoderlein and Mammen 2007, and Imbens and Newey 2009), multiple errors (e.g. random coefficient models like Beran, Feuerverger, and Hall 1996 and generalizations like Hoderlein, Nesheim, and Simoni 2011 and Lewbel and Pendakur 2011) or both (e.g., Matzkin 2003). This paper contributes to that literature by identifying models that have both additive measurement error and structural nonseparable unobserved heterogeneity.

with  $S$  and  $W$  being unobserved measurement errors that we assume, for now, to have the classical property of being mean zero with  $S, W \perp Y^*, X^*$ . This assumption is just made here and now to ease exposition; our formal results will substantially relax these independence assumptions, replacing them with Assumption 1 below. We will later further generalize the model to include different specifications for the measurement errors. We explicitly allow  $S$  and  $W$  to be correlated with each other. This might be due to the nature of the variables involved, or caused by the way in which  $Y$  and  $X$  are collected, as is the case for consumption data as described in the Introduction, or when related reporting biases affect the collection of both  $Y$  and  $X$ .

The model considered might also arise because of nondifferential properties of the measurement error in  $X$ . In the statistics literature, a measurement error  $W$  in  $X$  is called “differential” if it affects the observed outcome  $Y$ , after conditioning on the true  $X^*$ , that is, if  $Y | X^*, W$  does not equal  $Y | X^*$  (see, e.g., Carroll, Ruppert, Stefanski, and Crainiceanu (2006)). An alternative application of our identification results would be for a model in which  $Y$  is not mismeasured, and the additive error  $S$  instead represents the effect of differential measurement error  $W$  on the true observed outcome  $Y$ . In this setup  $Y | X^*$  is in general different from  $Y | X^*, W$ , with the two distributions being equal only if  $S$  and  $W$  are independent, so that the amount of correlation between  $S$  and  $W$  could be thought of as the extent of the departure from the nondifferential assumption on  $W$ . Since the nature of  $S$  does not affect our identification result, to ease exposition, in the following we will refer to  $S$  only as measurement error in  $Y^*$ .

Without loss of generality we specify  $Y^*$  as

$$Y^* = H(X^*, U),$$

where  $H(\cdot, \cdot)$  is an unknown function of a scalar random regressor  $X^*$ , and a random scalar or vector of nonseparable unobservables  $U$ , which can be interpreted as regression model errors or unobserved heterogeneity in the population. The extension to inclusion of other (observed) covariates will be straightforward, so we drop them for now.

In this setup our primary goal is identification (and later estimation) of the nonparametric regression function  $E[H(X^*, U) | X^*]$ , but we more generally consider identification of conditional moments  $E[H(X^*, U)^k | X^*]$  for integers  $k$ . Thus our results can be interpreted as separating the

impact of unobserved heterogeneity  $U$  from the effects of measurement errors on the relationship of  $Y$  to  $X$ . We do not deal directly with estimation of  $H$  and of  $U$ , but these could be recovered with some additional assumptions given our results. This, along with identification of the conditional distribution of  $Y^*$  and  $X^*$ , is discussed at the end of section 3.

The two main complications in our setup compared to a classical measurement error problem are that the true regression model  $H$  is an unknown, generally nonlinear, function containing nonseparable model errors  $U$ , and that measurement errors in  $Y^*$  and  $X^*$  may be correlated. To aid identification, we assume throughout the availability of additional information provided by instruments  $Q$  satisfying the following relationship

$$X^* = m(Q) + V,$$

for some unknown function  $m(Q) = E(X^* | Q)$ , with  $V \perp Q$ . This independence assumption is standard in the measurement error literature, particularly for nonlinear models.

In the special case where  $S$  is identically zero, that is  $Y^* \equiv Y$ , Schennach (2007) shows identification of the conditional mean function  $E(Y^* | X^*)$ , relying on features of the Fourier transforms of  $E[Y | Q]$  and of  $E[YX | Q]$ . These transforms are in general not ordinary point-wise defined functions, but generalized functions, being the Fourier transforms of not-absolutely integrable functions.

The intuition behind our extension is that when  $S$  is not identically zero, and is in fact correlated with  $W$ , the conditional expectation  $E[YX | Q]$  is shifted by a quantity which is proportional to  $E[WS | Q]$ , while  $E[Y | Q]$  remains unaffected. Under the identifying assumption that the covariance between  $S$  and  $W$  does not depend on the instruments  $Q$  (which is an implication of classical measurement error), this additional term  $E(SW)$  is a constant. Since the Fourier transform of a constant is a function which has a single point of support, i.e. the Dirac's delta generalized function, this only affects the Fourier transform of  $E[YX | Q]$  on one point, and the identification of the conditional mean function of interest  $E[Y^* | X^*]$  can proceed as before. We generalize this argument to also identify higher order conditional moments of the form  $E[Y^k | X^*]$ , with  $k \geq 2$ .

The same intuition can be applied to different specifications of the measurement error. An alternative specification we consider which as we show is particularly appropriate for Engel curves

(see also Lewbel 1996) is:

$$\begin{aligned} Y &= Y^* + X^*S, \\ X &= X^*W. \end{aligned}$$

This model is consistent with empirical evidence that measurement errors in expenditures increase with total expenditure  $X^*$ , and is consistent with the standard survey data generating process in which total reported expenditures  $X$  are constructed by summing the reported expenditures on individual goods.

In this Engel curve model we identify  $E[Y^{*k} | X^*] = E[H(X^*, U)^k | X^*]$  for an arbitrary integer  $k$ , thereby separating the effects on  $Y$  of observed and unobserved heterogeneity ( $X^*$  and  $U$ ) from the measurement errors  $S$  and  $W$ . This is done by relying on two different identification strategies, depending on different assumptions regarding the structure of measurement errors. The first approach builds on the fact that for utility derived Engel curves  $H(0, U) \equiv 0$ , while the second approach makes use of the specific dependence structure between  $W$  and  $S$  implied by the definition of  $Y$  and the construction of  $X$  in the Engel curve framework. We then find that this second approach has some features that make it more appropriate for our data, and we use it in our empirical application.

### 3 Identification

As discussed in the previous section, we begin by writing the unobserved  $Y^*$  and  $X^*$  as

$$\begin{aligned} Y^* &= H(X^*, U), \\ X^* &= m(Q) + V, \end{aligned} \tag{1}$$

where  $Q$  is a vector of instruments,  $V$  is a scalar unobserved random variable independent of  $Q$ ,  $U$  is a vector of unobserved disturbances, and the function  $H(\cdot, \cdot)$  is unknown. The scalar random variable  $Y^*$  is unobserved, but, encompassing and generalizing the examples given in the previous

section, assume that the observed  $Y$  is given by:

$$Y = Y^* + X^{*l}S \quad (2)$$

for some non-negative integer  $l$ , where  $E[S|X^*] = 0$ . By this construction, the measurement error  $X^{*l}S$  is mean zero, but has higher moments that can depend on  $X^*$ . Note that  $l = 0$  corresponds to the case of classical measurement error in  $Y^*$ , while the generalization to  $l > 0$  is useful for dealing with models such as Engel curves, where the variance in measurement errors increases with  $X^*$ .

The regressor  $X^*$  is also measured with error, with  $X$  satisfying:

$$X = X^* + W, \quad \text{with } E[W] = 0, \quad (3)$$

thereby allowing for additive measurement errors in  $X$ , while retaining the property that  $E[X] = E[X^*]$ . We will later generalize our identification result to the case of multiplicative measurement error, but to ease exposition we focus for now on the specification of measurement error given by equation (3).

In order to simplify notation let  $\mu^k(x^*) = E[Y^k | X^* = x^*]$  be the  $k$ -th conditional moment of the observed random variable  $Y$  given  $X^*$ . We now show identification of  $\mu^k(x^*)$  for  $k = 1, \dots, K$ , given knowledge of the observable triple  $(Y, X, Q)$ . Then, in Section 4, we will provide conditions under which identification of moments of the form  $E[Y^{*k} | X^*]$  can be achieved. The following Assumption will be maintained throughout:

**Assumption 1.** *The random variables  $Q, U, V, W$  and  $S$  are jointly i.i.d. and*

- (i)  $E[W^k | Q, V, U] = E[W^k]$  for  $k = 1, \dots, K$ ,
- (ii)  $E[S^k | Q, V, U] = E[S^k]$  for  $k = 1, \dots, K$ ,
- (iii)  $V$  is independent of  $Q$ ,
- (iv)  $E[WS | Q] = E[WS]$ .
- (v)  $U$  is independent of  $Q$  conditional on  $X^*$ .



The mean independence Assumptions (i) and (ii) with  $K = 1$  are a little weaker than assuming that measurement errors are classical. We assume these for values of  $k$  greater than one because we consider identification of these higher moments, not just the  $k = 1$  conditional moment. Assumption (iii), which is also made by Schennach (2007), is a standard control function assumption commonly used for identification and estimation of nonlinear models using instruments.<sup>3</sup>

Assumption (iv) is less restrictive than standard measurement error models, because standard models assume no correlations between measurement errors, and thereby trivially satisfy this assumption. Assumption (iv) would also follow from, and is strictly weaker than, the standard classical assumption that measurement errors be independent of correctly measured covariates. Finally Assumption (v) is a minimal instrument validity restriction.

Without loss of generality, define  $m(Q)$  by  $m(Q) \equiv E[X | Q]$ , so  $V$  has mean zero and  $m(Q)$  is nonparametrically identified. Defining  $Z = m(Q)$  and  $\tilde{V} = -V$ , we may conveniently rewrite equation (1) as:

$$X^* = Z - \tilde{V}, \quad (4)$$

which we will do hereafter. Following Newey (2001) and Schennach (2007) we will show that, under Assumption 1, knowledge of the conditional moments  $E[Y^k | Z]$ , for  $k = 1, \dots, K$ , and  $E[XY | Z]$  is enough to identify  $\mu^k(x^*)$  for  $k = 1, \dots, K$ .

To this end, under Assumption 1 we can rewrite the observed conditional expectations of  $Y^k$  and of the product  $XY$ , conditional on  $Z$ , as follows:

$$E[Y^k | Z] = E[\mu^k(x^*) | Z], \quad (5)$$

$$E[XY | Z] = E[x^* \mu^1(x^*) | Z] + E[x^{*l} | Z]E[WS]. \quad (6)$$

Detailed derivation of these equations is provided in the Appendix.

The proof of identification of  $\mu^k(x^*)$  is obtained by exploiting properties of the Fourier transform of these conditional expectations. The following assumption guarantees that these transforms and related objects are well defined.

---

<sup>3</sup>As pointed out by Schennach (2008), this assumption has testable implications. Independence between the estimated residuals of the feasible regression of  $X$  on  $Q$  and the instruments  $Q$  provides a testable sufficient condition for (iii) to be valid. This is in fact more restrictive than Assumption (iii), since the estimated residuals from this feasible regression are also functions of the measurement error  $W$ .

**Assumption 2.**  $\mu^k(x^*)$ ,  $E[Y^k | Z]$  and  $E[XY | Z]$  are scalar functions in  $\mathcal{S}$ , where  $\mathcal{S}$  is the space of functions  $f(t)$  for  $t \in \mathbb{R}$  such that each satisfies

$$\int (1+t^2)^{-r} |f(t)| dt < \infty, \text{ for some } r \geq 0.$$

Assumption 2 restricts the conditional expectations of interest to be members of a subclass of locally integrable functions, and also excludes specifications that rapidly approach infinity like the exponential function.

Under Assumption 1 we may write:

$$\begin{aligned} E[Y^k | Z] &= \int \mu^k(z-v) dF(v) \\ E[XY | Z] &= \int (z-v) \mu^1(z-v) dF(v) + \lambda \int (z-v)^l dF(v), \end{aligned}$$

where  $\lambda = E[WS]$ . These are convolution equations. We apply Fourier transform to the functions in these equations.

**Lemma 1.** *Equations (5) and (6) are equivalent to:*

$$\varepsilon_{y^k}(\zeta) = \gamma_k(\zeta) \phi(\zeta) \tag{7}$$

$$\mathbf{i} \varepsilon_{xy}(\zeta) = \dot{\gamma}_1(\zeta) \phi(\zeta) + \lambda \mathbf{i} \psi(\zeta) \phi(\zeta) \tag{8}$$

where  $\mathbf{i} = \sqrt{-1}$ , overdots denote derivatives with respect to  $z$ , and

$$\begin{aligned} \varepsilon_{y^k}(\zeta) &= \int E[Y^k | Z = z] e^{\mathbf{i}\zeta z} dz, & \gamma_k(\zeta) &= \int \mu^k(x^*) e^{\mathbf{i}\zeta x^*} dx^* \\ \varepsilon_{xy}(\zeta) &= \int E[XY | Z = z] e^{\mathbf{i}\zeta z} dz, & \phi(\zeta) &= \int e^{\mathbf{i}\zeta v} dF(v) \end{aligned}$$

where  $F(v)$  is the cdf of  $\tilde{V}$ ,  $\lambda = E[WS]$  and  $\psi(\zeta) = \int x^{*l} e^{\mathbf{i}\zeta x^*} dx^*$ .

Note that the right hand side integrals may diverge, however, in the space of generalized functions  $S^*$  (see Zinde-Walsh 2014) these objects are well defined and represent generalized functions. Examples are the well-known Dirac delta function and its derivatives, but the generalized functions defined above could be more complicated elements of  $S^*$ . These generalized functions are not

defined point-wise, but have a well-defined support and we keep the familiar (even if somewhat misleading) function notation to indicate their domain of definition.

Lemma 1 is a generalization of Lemma 1 in Schennach (2007), who considers the special case where  $l = 0$  and  $k = 1$ . Note that  $\phi(\zeta)$ , being the characteristic function of  $\tilde{V}$ , is an ordinary function. Products of generalized functions are not necessarily well defined, so we cannot just freely take the ratio of equations (7) and (8) like ordinary functions to cancel out the characteristic function  $\phi(\zeta)$ . Also note that the unknown quantities here are  $\gamma_k(\zeta)$  and  $\phi(\zeta)$ , while  $\psi(\zeta)$  is the Fourier transform of a power function, and hence is known and equal to the  $l$ -th generalized derivative of a Dirac delta function. For a more detailed treatment of generalized functions see Lighthill (1962) or the supplementary material in Schennach (2007).

**Assumption 3.** *The characteristic function of  $\tilde{V}$ ,  $\phi(\zeta)$ , is continuously differentiable and  $\phi(\zeta) \neq 0$  for all  $\zeta \in \mathbb{R}$ .*

**Assumption 4.** *For each  $k = 1, \dots, K$  there exists a finite or infinite constant  $\bar{\zeta}_k$  such that  $\text{supp}(\gamma_k) = \Omega^k = [-\bar{\zeta}_k, \bar{\zeta}_k]$ .*

Assumptions 3 and 4 are equivalent to Assumptions 2 and 3 in Schennach (2007) and are standard in the deconvolution literature. Since we are seeking nonparametric identification of  $\gamma_k(\zeta)$ , the characteristic function of  $\tilde{V}$  needs to be non-vanishing, thus excluding uniform or triangular like distributions, while  $\gamma_k(\zeta)$  needs to be either non-vanishing or must vanish on an infinite interval. This is required for  $\gamma_k(\zeta)$  to be fully nonparametrically identified. Assumption 4 would, for instance, rule out sinusoidal specifications for  $\mu(x^*)$ , which are not very common in economic applications.<sup>4</sup>

The following theorem states our main identification result.

**Theorem 1.** *Under Assumptions 1-4, if  $\Omega^k \subseteq \Omega^1$  then  $\mu^k(x^*)$  for  $k = 1, \dots, K$  are nonparametrically identified. Also, if  $\bar{\zeta}_1 > 0$  in Assumption 4 then*

$$\mu^k(x^*) = (2\pi)^{-1} \int \gamma_k(\zeta) e^{-i\zeta x^*} d\zeta$$

where

$$\gamma_k(\zeta) = \varepsilon_{y^k}(\zeta) / \phi(\zeta) \text{ for } \zeta \in (-\bar{\zeta}_1, \bar{\zeta}_1), \quad (9)$$

---

<sup>4</sup>If  $H(X^*, U)$  were parametrically specified, then Assumption 4 could be relaxed, because in that case information obtained from a finite number of points of  $\gamma_k(\zeta)$  would generally suffice for identification.

$\phi(\zeta)$  is the characteristic function of  $\tilde{V} \equiv -V$  given, for  $|\zeta| < \bar{\zeta}_1$ , by

$$\phi(\zeta) = \exp \int_0^\zeta \varkappa(t) dt, \quad (10)$$

with  $\varkappa(\zeta)$  being the uniquely defined solution in  $(-\bar{\zeta}_1, \bar{\zeta}_1)$ , to

$$\mathbf{i}\varepsilon_{(z-x)y}(\zeta) = \varepsilon_{y^1}(\zeta)\varkappa(\zeta),$$

where  $\varepsilon_{y^1}(\zeta) = \int E[Y|Z = z]e^{\mathbf{i}\zeta z} dz$  and  $\varepsilon_{(z-x)y}(\zeta) = \int E[(Z - X)Y|Z = z]e^{\mathbf{i}\zeta z} dz$  respectively.

Theorem 1 immediately implies identification of  $E(Y^* | X^* = x^*) = E(Y | X^* = x^*) = \mu^1(x^*)$ . Identification of higher moments of  $Y^*|X^*$  based on  $\mu^k(x^*)$  is discussed later and in the next section.

Theorem 1 is a generalization of Theorem 1 in Schennach (2007) and of Theorem 3(B) in Zinde-Walsh (2014). The proof is in the appendix, but essentially works as follows. The nonzero covariance between measurement errors  $W$  and  $S$ , coming from  $\lambda \neq 0$ , introduces the additional term  $\lambda \mathbf{i}\psi(\zeta)\phi(\zeta)$  in (8). The function  $\psi(\zeta)$  is the  $l$ -th generalized derivative of the Dirac delta function, whose support is the set  $\{0\}$ . This modifies the equation of the Fourier transform of  $E[XY|Z = z]$ ,  $\varepsilon_{xy}(\zeta)$ . Still, it can be shown that identification of the characteristic function of  $\tilde{V}$ , which is non-zero and continuous by Assumption (3), remains unaffected, and is obtained as in the case of  $\lambda = 0$ . This in turn allows identification of the function of interest  $\gamma_k(\zeta)$  as in equation (9) by means of equation (7) in Lemma 1. Finally, this function's inverse Fourier transform gives  $\mu^k(x^*)$ . This proof of Theorem 1 only needs to consider the case where  $\bar{\zeta}_k > 0$ , since the case where  $\bar{\zeta}_k = 0$  only occurs when  $\mu^k(x^*)$  is a polynomial in  $X^*$ , and that specification that has already been shown to be identified by Hausman, Newey, Ichimura, and Powell (1991).

Now consider the multiplicative measurement error structure, where the observed random variable  $X$  is such that

$$X = X^*W \quad \text{with } E[W] = 1, \quad (11)$$

where  $E[W] = 1$  ensures that  $E[X] = E[X^*]$ . In this case equation (5) still holds, while (6) becomes

$$E[XY | Z] = E[x^* \mu^1(x^*) | Z] + E[x^{*l+1} | Z]E[WS]. \quad (12)$$

A detailed derivation of this equation is provided in the Appendix. This implies that a slightly modified version of Lemma 1 holds, that is

**Lemma 2.** *Under Assumption 2, equations (5) and (12) are equivalent to*

$$\begin{aligned}\varepsilon_{y^k}(\zeta) &= \gamma_k(\zeta)\phi(\zeta) \\ \mathbf{i}\varepsilon_{xy}(\zeta) &= \dot{\gamma}_1(\zeta)\phi(\zeta) + \lambda\mathbf{i}\tilde{\psi}(\zeta)\phi(\zeta)\end{aligned}$$

where again  $\mathbf{i} = \sqrt{-1}$ , overdots denote derivatives with respect to  $z$ , and now

$$\begin{aligned}\varepsilon_{y^k}(\zeta) &= \int E[Y^k|Z=z]e^{\mathbf{i}\zeta z} dz, & \gamma_k(\zeta) &= \int \mu^k(x^*)e^{\mathbf{i}\zeta x^*} dx^* \\ \varepsilon_{xy}(\zeta) &= \int E[XY|Z=z]e^{\mathbf{i}\zeta z} dz, & \phi(\zeta) &= \int e^{\mathbf{i}\zeta v} dF(v)\end{aligned}$$

where  $F(v)$  is the cdf of  $\tilde{V}$ ,  $\lambda = E[WS]$  and  $\tilde{\psi}(\zeta) = \int x^{*l+1} e^{\mathbf{i}\zeta x^*} dx^*$ .

Comparing Lemmas 1 and 2, the only effect of considering a multiplicative rather than additive specification for measurement error in  $W$  is that of replacing the generalized function  $\psi(\zeta)$  in Lemma 1 with  $\tilde{\psi}(\zeta)$  in Lemma 2. We show that Theorem 1 still holds under this alternative specification (details are given in the proofs in the Appendix). Thus the moments  $\mu^k(x^*)$  are also identified under the multiplicative measurement error structure defined in equation (11).

An interesting side result of Theorem 1, useful for some applications, is that the distribution of the unobserved random variable  $X^*$  is also nonparametrically identified.

**Corollary 1.** *Let  $\phi_{x^*}(\zeta)$  be the characteristic function of the unobserved  $X^*$ . Under the Assumptions of Theorem 1,  $\phi_{x^*}(\zeta)$  is identified for  $|\zeta| < \bar{\zeta}_1$  and is given by:*

$$\phi_{x^*}(\zeta) = \frac{\phi_z(\zeta)}{\phi(\zeta)}, \tag{13}$$

where  $\phi_z(\zeta)$  is the characteristic function of the observed random variable  $Z$  and  $\phi(\zeta)$  is the characteristic function of the unobserved random variable  $\tilde{V}$  whose expression is given in equation (10).

Moreover, if  $\bar{\zeta}_1 = \infty$ , the density of the unobserved  $X^*$  is also identified as

$$f(x^*) = (2\pi)^{-1} \int \phi_{x^*}(\zeta) e^{-\mathbf{i}\zeta x^*} d\zeta.$$

Equation (13) is just a standard property of the characteristic function of the difference between two independent random variables.

Theorem 1 establishes a set of assumptions under which  $\mu^k(x^*) = E[Y^k|X^* = x^*]$  is identified for integers  $k$ . However, the policy relevant objects are usually the conditional moments of the true, unobserved  $Y^*$ , that is  $\omega^k(x^*) = E[Y^{*k}|X^* = x^*]$ , or more generally the conditional distribution of  $Y^*$  given  $X^*$ . For  $k = 1$  it is easy to see that  $E[Y|X^* = x^*] = E[Y^*|X^* = x^*]$ , that is  $\omega^1(x^*) = \mu^1(x^*)$ , implying that the first moment, which is usually the primary concern in empirical applications, is directly identified by Theorem 1. This was also the only estimand considered by Schennach (2007).

We now consider identification of higher order moments, that is, identification of  $\omega^k(x^*)$  for  $k > 1$ .

**Corollary 2.** *Let Assumptions of Theorem 1 hold. Then*

$$\omega^k(x^*) = \mu^k(x^*) - \sum_{j=0}^{k-1} \binom{k}{j} \omega^j(x^*) x^{*l(k-j)} E[S^{k-j}], \quad (14)$$

and hence  $\omega^k(x^*)$  is identified up to knowledge of  $E[S^j]$  for  $j = 2, \dots, k$ .

Corollary 2, which is based on the binomial expansion of equation (2), shows that in order to identify  $\omega^k(x^*)$  for  $k > 1$ , knowledge of moments of the unconditional distribution of  $S$  is needed. Identification of moments of the distribution of  $S$  requires additional information which may be provided by a combination of additional data, restrictions imposed on the dependence structure between  $W$  and  $S$ , or additional information regarding features of  $H(X^*, U)$ . The availability of such additional information is not a general feature of our model and therefore depends on context. To illustrate, we now provide one set of additional assumptions that suffice to identify  $\omega^k(x^*)$  for any integer  $k$ . Then, in the next section we discuss alternative identifying assumptions that are particularly appropriate for our Engel curve application.

Suppose that we let the assumptions of Theorem 1 hold with  $l = 0$ , so  $Y = Y^* + S$  and  $X = X^* + W$ . Assume in addition the boundary condition that  $H(0, U) \equiv 0$  for all values of  $U$  (in the next section we show how this boundary condition would hold by construction in the Engel

curve example). Using  $l = 0$  we can rearrange equation (14) as

$$E[S^k] = \mu^k(x^*) - \omega^k(x^*) - \sum_{j=1}^{k-1} \omega^{k-j}(x^*)E[S^j].$$

The restriction that  $H(0, U) \equiv 0$  in turn implies that  $E[Y^{*k} | X^* = 0] = \omega^k(0) = 0$  for all  $k \geq 1$ , so  $E[S^k] = \mu^k(0) - \sum_{j=1}^{k-1} \omega^{k-j}(0)E[S^j] = \mu^k(0)$ . Combining this result with Corollary 2 then gives  $\omega^k(x^*) = \mu^k(x^*) - \sum_{j=0}^{k-1} \binom{k}{j} \omega^j(x^*) x^{*l(k-j)} \mu^{s-k}(0)$ , which shows identification of  $\omega^k(x^*)$  for any integer  $k$ , given that  $\mu^1(x^*), \mu^2(x^*), \dots, \mu^k(x^*)$  is identified using Theorem 1.

The above argument provides one set of conditions for the identification of  $E[Y^{*k}|X^*]$  for any positive integer  $k$  (others are in the next section). More generally we might be interested in estimating the conditional distribution of  $Y^*$  given  $X^*$ , or in estimating  $H(X_i^*, U_i)$ . We do not consider formal estimation of these objects, but outline here how they could be identified and estimated.

Conditions under which a distribution is identified from its moments are well known. See, e.g., Assumption 7 of Fox, Kim, Ryan, and Bajari (2012). One such condition is that the moment generating function of  $Y_i^*$  given  $X_i^*$  exists in an open interval around zero. This ensures both that  $\omega^k(x^*)$  is finite for any positive integer  $k$ , and that the distribution of  $Y^*$  given  $X^*$  is identified from these moments. In terms of estimation, given this well behaved moment generating function and assuming that the conditional density of  $Y^*$  given  $X^*$  is sufficiently smooth, Gallant and Nychka (1987) show that we could approximate any member of that class of density functions to an arbitrary degree of precision by means of a standard normal density times a polynomial of a sufficiently high even power. Call this power  $K$ , so the larger is  $K$ , the better is this approximation. Knowledge of the conditional moments  $E[Y^{*k}|X^*]$  for  $k = 1, \dots, K$  would then suffice to recover the polynomial coefficients of this approximation, and thereby estimate the density of  $Y^*$  given  $X^*$ , using, e.g., method of moments estimation.

Given identification of the distribution of  $Y_i^*$  conditional on  $X_i^*$ , results such as those in Matzkin (2007) could then be applied to identify  $H(X_i^*, U_i)$ . One such result is that, if the conditional distribution of  $Y_i^*$  given  $X_i^*$  is continuous, and if  $H$  is monotonic in a scalar  $U_i$ , then we could define  $U_i$  as equaling the conditional distribution function of  $Y_i^*$  given  $X_i^*$ , and construct the function  $H$  as the inverse of this distribution function.

## 4 Identification of Engel curves

As discussed in the previous section,  $E(Y^* | X^* = x^*)$  equals  $\mu^1(x^*)$  and so is identified without additional assumptions by Theorem 1 and its corollaries. However, identification of higher moments  $\omega^k(x^*) = E(Y^{*k} | X^* = x^*)$  for  $k > 1$  based on  $\mu^k(x^*)$  requires additional information. In this section we consider such information in the form of identifying assumptions suitable for the particular context of Engel curve estimation. Let  $Y_i^*$  be unobserved expenditures on a particular good (or group of goods)  $i$ , for  $i = 1, \dots, I$  and let  $X^* = \sum_{i=1}^I Y_i^*$  be total expenditures on all goods. Let  $Y^* = Y_1^*$  denote the particular good of interest. Then  $Y^* = H(X^*, U)$  is the Engel curve for the good of interest, where  $U$  is now a vector of individual consumer specific utility (preference) related parameters. The goal is then identification of moments of the conditional distribution of  $Y^*$  given  $X^*$ .

We now describe two different possible models of measurement errors, each of which yields sufficient information to identify all moments of the conditional distribution of demands  $Y^*$  given total expenditure  $X^*$  using Theorem 1 and Corollary 2. In each case it is assumed, as is generally true empirically, that observed total expenditures  $X$  are obtained by summing the observed expenditures  $Y_i$  on each good  $i$ , so  $X = \sum_{i=1}^I Y_i$ .

For the first of these two models, assume  $Y_i = Y_i^* + S_i$  for each good  $i$ , corresponding to measurement error in the form of equation (2) with  $l = 0$ . Assume also that one cannot purchase negative amounts of any good, so  $Y_i^* \geq 0$  since each  $Y_i^*$  is defined as a level of expenditures. These two conditions then suffice to yield identification as described in the previous section. To see this, observe first that  $X = X^* + W$  follows by construction from  $X = \sum_{i=1}^I Y_i$  where the measurement error  $W = \sum_{i=1}^I S_i$ . In addition, the boundary condition  $H(0, U) \equiv 0$  holds for all values of  $U$ , because when  $X^* = \sum_{i=1}^I Y_i^* = 0$  having every  $Y_i^* \geq 0$  implies that every  $Y_i^* = 0$ , and so in particular  $Y_1^* = H(0, U) = 0$ .

This first model of measurement error illustrates our identification methodology, but it imposes the restrictive (for Engel curves) assumption that the independent measurement error is additive even though consumption must be non-negative. Moreover, it is likely that measurement errors in expenditures increase with the level of expenditures. Therefore, for estimation later we will focus on a nonparametric generalization of an alternative specification proposed by Lewbel (1996)



in the context of a parametric Engel curve model. This model assumes  $Y_i = Y_i^* + X^*S_i$  for each good  $i$ , corresponding to measurement error in the form of equation (2) with  $l = 1$  for  $S = S_1$ . Let  $\tilde{S} = \sum_{i=2}^I S_i$ . Assume that  $S$  and  $\tilde{S}$  (corresponding to measurement errors for different goods) have mean zero and are independent of each other. Then, by equation (2), summing up expenditures on different goods we obtain  $X = \sum_{i=1}^I Y_i = X^* \left(1 + S + \tilde{S}\right)$ , corresponding to multiplicative measurement error in  $X$  given by equation (11) with

$$W = 1 + S + \tilde{S} \quad (15)$$

and  $E(W) = 1$  as required by Theorem 1. In the following we argue that such a measurement error structure imposes restrictions on the joint distribution of the couple  $(S, W)$  which allows us to identify moments of the marginal distribution of  $S$  from knowledge of moments of the form  $E[W^k]$  and  $E[W^k S]$ , which we show to be identified from the data in the multiplicative measurement error setup. The following Theorem establishes formal identification for these quantities.

**Theorem 2.** *Let Assumptions 1-4 and equations (1) and (11) hold. Let the first  $K$  moments of  $X$  be finite with  $E[X^{*k} | Z]$  being a strictly positive functions of  $z$  for every  $k = 1, \dots, K$ . Then the first  $K$  moments of  $W$  are identified and*

$$E[W^k] = \frac{E[X^k]}{\sum_{j=0}^k \binom{k}{j} \mathbf{i}^{k-j} E[Z^j] \phi^{(k-j)}(0)}. \quad (16)$$

Furthermore, if  $\bar{\zeta}_1 = \infty$  in Assumption 4, then:

$$g_k(z) = E[W^k S] h_k(z), \quad (17)$$

with  $g_k(z)$  and  $h_k(z)$  being functions of  $z$  given by the following expressions:

$$\begin{aligned} g_k(z) &:= E[X^k Y | Z = z] - (2\pi)^{-1} E[W^k] \int (-\mathbf{i})^k \gamma_1^{(k)}(\zeta) \phi(\zeta) e^{-\mathbf{i}\zeta z} d\zeta, \\ h_k(z) &:= \sum_{j=0}^{k+l} \binom{k+l}{j} z^j (-\mathbf{i})^{k+l-j} \phi^{(k+l-j)}(0), \end{aligned}$$

where  $\gamma_1^{(k)}(\zeta)$  is the  $k$ -th derivative of  $\gamma_1(\zeta)$  as defined in equation (9), while  $\phi(\zeta)$  is defined as in

(10). Moments  $E[W^k S]$  for  $k = 1, \dots, K - l$  are then identified by evaluating the ratio of these two functions at an arbitrary  $z$ .

The proof of Theorem 2 is given in the Appendix. Intuitively identification of  $E[W^k]$  for  $k = 1, \dots, K$  follows by noting that, from equation (11) and by Assumption 1,  $E[X^k] = E[X^{*k}]E[W^k]$ , and since the unobserved distribution  $X^*$  is identified by Corollary 1 every moment of  $W$  is also identified. Furthermore from equation (11) we have

$$E[X^k Y | Z] = E[X^* \mu^1(x^*) | Z] E[W^k] + E[X^{*k+l} | Z] E[W^k S].$$

This equation depends only on  $E[W^k S]$  and on identified moments, so solving this expression for  $E[W^k S]$  shows that  $E[W^k S]$  is identified.

Both  $Y^*$  and  $X^*$  are non-negative random variables, so the requirement that the first  $K$  marginal and conditional moments of  $X^*$  be non-zero is satisfied as long as  $X^*$  is non-degenerate. This is because we are considering raw moments and not central ones, hence we are not for example ruling out symmetric distributions, for which the third central moment would be zero. Furthermore, the assumption of  $\bar{\zeta}_1 = \infty$ , covers empirically relevant frameworks, as was discussed in Section 3 (see also Lewbel 1996).

Knowledge of moments of the form  $E[W^k]$  and  $E[W^k S]$  then allows us to recover moments of  $S$  and  $\tilde{S}$  as follows. By rearranging equation (15) we get

$$E[S^k] = E[W^{k-1} S] - \sum_{j=1}^{k-1} \binom{k-1}{j-1} E[S^j] E[(1 + \tilde{S})^{k-j}], \quad (18)$$

where  $E[(1 + \tilde{S})^k]$  is given by

$$E[(1 + \tilde{S})^k] = E[W^k] - \sum_{j=0}^{k-1} \binom{k}{j} E[S^{k-j}] E[(1 + \tilde{S})^j], \quad (19)$$

thereby showing that moments of  $S$  and  $\tilde{S}$  can be recursively obtained from the identified moments in Theorem 2. Details on the derivation of equations (18) and (19) are provided in the Appendix.

It is worth noting that, from equations (18) and (19) with  $k = 2$ , we obtain  $E[S^2] = E[WS]$  and  $E[(1 + \tilde{S})^2] = E[W^2] - E[S^2] = E[W^2] - E[WS]$ , meaning that the second order moments of

the measurement error distributions are directly identified from the quantities derived in Theorem 2.

More generally, under the assumptions of Theorems 1 and 2, any conditional moment of the distribution of the unobserved  $Y^*$  on  $X^*$  is identified. As noted at the end of the previous section, assuming  $Y^*|X^*$  has a sufficiently regular moment generating function then implies that the distribution function of  $Y^*$  given  $X^*$  is identified.

## 5 Estimation

In this Section we propose a sieve based nonparametric estimator for the conditional moments of the distribution of  $Y^*$  given  $X^*$ , which we will apply to the estimation of Engel curves. Many studies have documented a variety of nonlinearities and substantial unobserved heterogeneity in Engel curve shapes, see, e.g., Blundell, Browning, and Crawford (2003) and Lewbel and Pendakur (2009), or Lewbel (2008) for a survey. It is therefore useful to provide an estimator that allows for the presence of measurement error of the specific kind implied by expenditure data, while not imposing functional form restrictions. Also as noted earlier, unlike previous studies, we are able to disentangle the variance components due to measurement error from those due to preference heterogeneity.

The estimator we propose is essentially an application of the sieve GMM estimator of Ai and Chen (2003) under the conditional moment restrictions outlined in Theorem 2. For ease of exposition we describe estimation of the first two conditional moments  $\omega^k(x^*) = E(Y^{*k} | X^* = x^*)$  for  $k = 1, 2$ , but given our identification results the corresponding extension to estimation of higher moments is purely mechanical. We focus on the model

$$Y = Y^* + X^*S, \quad X = X^*W, \quad \text{and} \quad X^* = m(Q) + V.$$

The data consist of an i.i.d. sample of size  $N$  from the triple  $(Y, X, Q)$  and the goal is consistent estimation of the functions  $\omega^1(\cdot)$  and  $\omega^2(\cdot)$ . Equations (5) and (12) imply the following specifications

for  $Y$ ,  $Y^2$  and  $XY$ :

$$Y = \omega^1(X^*) + \epsilon_1, \quad (20)$$

$$Y^2 = \omega^2(X^*) + \lambda X^{*2} + \epsilon_2, \quad (21)$$

$$XY = X^* \omega^1(X^*) + \lambda X^{*2} + \epsilon_3, \quad (22)$$

with  $E[\epsilon_1 | X^*] = E[\epsilon_2 | X^*] = E[\epsilon_3 | X^*] = 0$  and where from equation (18)  $E[S^2] = E[WS]$ ; we denote this by  $\lambda$ . Starting from the identification results provided in Theorems 1 and 2, and then integrating out the unobserved distribution of  $V$  from equations (20) to (22), yields

$$\rho_0(\mathbf{W}; \boldsymbol{\theta}) = Y - \int \omega^1(m(Q) - \sigma v) f(v) dv, \quad (23)$$

$$\rho_1(\mathbf{W}; \boldsymbol{\theta}) = Y^2 - \int [\omega^2(m(Q) - \sigma v) + \lambda(m(Q) - \sigma v)^2] f(v) dv, \quad (24)$$

$$\rho_2(\mathbf{W}; \boldsymbol{\theta}) = XY - \int [(m(Q) - \sigma v) \omega^1(m(Q) - \sigma v) + \lambda(m(Q) - \sigma v)^2] f(v) dv, \quad (25)$$

where  $\mathbf{W} = (Y, X, Q)$  and  $\boldsymbol{\theta} = (\lambda, \sigma, m(\cdot), \omega^1(\cdot), \omega^2(\cdot), f(\cdot))$ , with  $f(\cdot)$  being the probability density function of the first stage error term  $V$ , normalized to zero mean and unit variance, and  $\sigma$  being the corresponding standard deviation. As in equation (4),  $m(\cdot)$  is the function describing the first stage regression of the observed  $X$  on  $Q$ , so

$$\psi(\mathbf{W}; \boldsymbol{\theta}) = X - m(Q).$$

Note that the quantities  $\rho_j(\mathbf{W}; \boldsymbol{\theta})$  correspond to the residuals  $\epsilon_j$  in equations (20) to (22), while  $\psi(\mathbf{W}; \boldsymbol{\theta})$  corresponds to the first stage error term  $V$ . Under Assumption 1 we have  $E[V | Q] = 0$  and  $E[\epsilon_j | Q] = 0$ , for  $j = 1, 2, 3$ , so by defining  $\boldsymbol{\rho}(\mathbf{W}; \boldsymbol{\theta}) = (\rho_0(\mathbf{W}; \boldsymbol{\theta}), \rho_1(\mathbf{W}; \boldsymbol{\theta}), \rho_2(\mathbf{W}; \boldsymbol{\theta}), \psi(\mathbf{W}; \boldsymbol{\theta}))'$  the following vector valued moment condition holds:

$$E[c_j(Q) \boldsymbol{\rho}(\mathbf{W}; \boldsymbol{\theta})] = 0,$$

where  $c_j(Q)$  is the  $j$ -th element of a vector of known (chosen by the econometrician) basis functions  $\mathbf{c}(Q) = (c_1(Q), \dots, c_{K^c}(Q))$ .

Given our identification results, a consistent estimator for  $\theta$  is obtained by minimizing the sample analogue of the above conditional expectation, that is:

$$\frac{1}{N} \sum_{i=1}^N [\mathbf{c}(q_i) \otimes \boldsymbol{\rho}(\mathbf{w}_i; \theta)]' A^{-1} [\mathbf{c}(q_i) \otimes \boldsymbol{\rho}(\mathbf{w}_i; \theta)],$$

for some positive definite square matrix  $A$ , where  $\mathbf{w}_i = (y_i, x_i, q_i)$  denotes observations and  $\otimes$  is the Kroenecker product. The computation of  $\boldsymbol{\rho}(\mathbf{w}_i; \theta)$  is complicated by two main factors. First, the parameter vector  $\theta$  is infinite-dimensional due to the presence of the unknown functions  $\omega^1(\cdot)$ ,  $\omega^2(\cdot)$ ,  $f(\cdot)$  and  $m(\cdot)$ . Second, even if these functions were finitely parameterized, the computation of  $\boldsymbol{\rho}(\mathbf{w}_i; \theta)$  would involve integrals (23)-(25) which do not have a closed form solution.

We address both of these issues by adopting a minimum distance sieve estimator as in Ai and Chen (2003), replacing the space  $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_f \times \mathcal{H}_m$  with a finite-dimensional sieve space  $\mathcal{H}_n = \mathcal{H}_{1n} \times \mathcal{H}_{2n} \times \mathcal{H}_{fn} \times \mathcal{H}_{mn}$  which becomes dense in the original space  $\mathcal{H}$  as  $n$  increases as in Grenander (1981). The spaces  $\mathcal{H}_i$ , for  $i = 1, 2$  and  $\mathcal{H}_m$  are the spaces of locally integrable functions, bounded by polynomials, while  $\mathcal{H}_f$  is the space of continuous distributions with mean zero and well-behaved characteristic function which satisfies Assumption 3.

Computations involving integrals are simplified by a convenient choice of the sieve space  $\mathcal{H}_n$ . We employ cosine polynomial and Hermite polynomial sieve spaces to approximate the conditional moments ( $\omega^1(x^*)$  and  $\omega^2(x^*)$ ), and the density function  $f(v)$  respectively, that is,

$$\omega^t(x^*) \approx \sum_{j=0}^{N_t} \beta_{jt} b_{jt}(x^*), \quad t = 1, 2, \quad (26)$$

$$f(v) \approx \sum_{j=0}^{N_f} \alpha_j h_j(x^*), \quad (27)$$

for some  $N_1, N_2, N_f$ , and where the basis functions  $\{b_{jt}(x^*), j = 0, 1, \dots\}$  and  $\{h_j(x^*), j = 0, 1, \dots\}$  are given by:

$$b_{jt}(x^*) = \cos\left(\frac{i\pi(x^* - a_t)}{b_t - a_t}\right), \quad h_j(x^*) = H_j(v)\phi(v)$$

for some  $a_t, b_t$ ,  $t = 1, 2$ . The function  $\phi(\cdot)$  is the standard normal density function, while  $H_j(\cdot)$  is the  $j$ -th order Hermite polynomial. Cosine polynomial sieves are chosen to approximate conditional moments since they are known for well approximating aperiodic functions on an interval (see Chen

2007 and Newey and Powell 2003). Hermite polynomial sieves, on the other hand, are well suited for approximating the density function  $f(v)$  for two reasons. First, standard restrictions for the approximating density to integrate to one and to be mean zero with unit variance are trivially imposed by setting  $\alpha_0 = 1$  and  $\alpha_1 = \alpha_2 = 0$ . Second, the fact that a Hermite polynomial is multiplied by the standard normal density allows us to easily compute the integrals in equations (23)-(25) along the lines of Newey (2001) and Wang and Hsiao (2011). Finally, we employ a polynomial sieve space for the first stage regression function  $m(\cdot)$ , that is, we let

$$m(q_i; \boldsymbol{\delta}) \approx \sum_{j=0}^{N_m} \delta_j q_i^j, \quad (28)$$

for some integer  $N_m$  and vector of parameters  $\boldsymbol{\delta} = (\delta_0, \dots, \delta_{N_m})$ .

By substituting (28), (26) and (27) into the sample analogue of (23), (24) and (25) we obtain:

$$\begin{aligned} \rho_0(\mathbf{w}_i; \boldsymbol{\eta}) &= y_i - \sum_{k=0}^{N_1} \sum_{l=0}^{N_f} \beta_{k1} \alpha_l \int b_{k1} (m(q_i; \boldsymbol{\delta}) - \sigma v) H_l(v) \phi(v) dv, \\ \rho_1(\mathbf{w}_i; \boldsymbol{\eta}) &= y_i^2 - \sum_{k=0}^{N_2} \sum_{l=0}^{N_f} \beta_{k2} \alpha_l \int \left[ b_{k2} (m(q_i; \boldsymbol{\delta}) - \sigma v) + \lambda (m(q_i; \boldsymbol{\delta}) - \sigma v)^2 \right] H_l(v) \phi(v) dv, \\ \rho_2(\mathbf{w}_i; \boldsymbol{\eta}) &= x_i y_i - \sum_{k=0}^{N_1} \sum_{l=0}^{N_f} \beta_{k1} \alpha_l \int \left[ (m(q_i; \boldsymbol{\delta}) - \sigma v) b_{k1} (m(q_i; \boldsymbol{\delta}) - \sigma v) + \lambda (m(q_i; \boldsymbol{\delta}) - \sigma v)^2 \right] H_l(v) \phi(v) dv, \end{aligned}$$

where  $\boldsymbol{\eta} = (\lambda, \sigma, \alpha_0, \dots, \alpha_{N_f}, \beta_{10}, \dots, \beta_{1N_1}, \beta_{20}, \dots, \beta_{2N_2}, \delta_1, \dots, \delta_{N_m})$ . Thus the integrals involved in  $\rho(\mathbf{w}_i; \boldsymbol{\eta})$  can be computed with an arbitrary degree of precision by averaging the value of the integrand function over randomly drawn observations from a standard normal density. For instance, let  $v_j$  for  $j = 1, \dots, J$  denote random draws from a standard normal distribution,  $\rho_0(\mathbf{w}_i; \boldsymbol{\eta})$  is computed as

$$\rho_0(\mathbf{w}_i; \boldsymbol{\eta}) = y_i - J^{-1} \sum_{k=0}^{N_1} \sum_{l=0}^{N_f} \beta_{k1} \alpha_l \sum_{j=1}^J b_{k1} (m(q_i; \boldsymbol{\delta}) - \sigma v_j) H_l(v_j).$$

Similar expressions hold for  $\rho_1(\mathbf{w}_i; \boldsymbol{\eta})$  and  $\rho_2(\mathbf{w}_i; \boldsymbol{\eta})$ .<sup>5</sup>

It then follows from Theorems 1 and 2 and from Lemma 3.1 in Ai and Chen (2003) that a

---

<sup>5</sup>While  $N_1, N_2, N_f$  and  $N_m$  need to increase with sample size and play the role of smoothing parameters,  $J$  only affects the degree of precision with which the integrals are evaluated and is set as large as is computationally practical, analogous to the choice of the fineness of the grid in ordinary numerical integration.

consistent estimator for  $\boldsymbol{\eta}$  is given by

$$\arg \min_{\boldsymbol{\eta}} \frac{1}{N} \sum_{i=1}^N [\mathbf{c}(q_i) \otimes \boldsymbol{\rho}(\mathbf{w}_i, \boldsymbol{\eta})]' [\boldsymbol{\Sigma}(\mathbf{w}_i)]^{-1} [\mathbf{c}(q_i) \otimes \boldsymbol{\rho}(\mathbf{w}_i, \boldsymbol{\eta})],$$

for a positive definite matrix  $\boldsymbol{\Sigma}(\mathbf{w}_i)$ . We implement this result by applying a standard two-step GMM procedure:

- (1) Obtain an initial estimate  $\hat{\boldsymbol{\eta}}$  from the consistent estimator:

$$\arg \min_{\boldsymbol{\eta}} \frac{1}{N} \sum_{i=1}^N [\mathbf{c}(q_i) \otimes \boldsymbol{\rho}(\mathbf{w}_i, \boldsymbol{\eta})]' [\mathbf{c}(q_i) \otimes \boldsymbol{\rho}(\mathbf{w}_i, \boldsymbol{\eta})].$$

- (2) Improve the efficiency of the estimator by applying the minimization:

$$\arg \min_{\boldsymbol{\eta}} \frac{1}{N} \sum_{i=1}^N [\mathbf{c}(q_i) \otimes \boldsymbol{\rho}(\mathbf{w}_i, \boldsymbol{\eta})]' [\hat{\boldsymbol{\Sigma}}(\mathbf{w}_i)]^{-1} [\mathbf{c}(q_i) \otimes \boldsymbol{\rho}(\mathbf{w}_i, \boldsymbol{\eta})],$$

where  $\hat{\boldsymbol{\Sigma}}(\mathbf{w}_i)$  is obtained from the first step estimator  $\hat{\boldsymbol{\eta}}$  as:

$$\hat{\boldsymbol{\Sigma}}(\mathbf{w}_i) = \frac{1}{N} \sum_{i=1}^N [\mathbf{c}(q_i) \otimes \boldsymbol{\rho}(\mathbf{w}_i, \hat{\boldsymbol{\eta}})]' [\mathbf{c}(q_i) \otimes \boldsymbol{\rho}(\mathbf{w}_i, \hat{\boldsymbol{\eta}})].$$

Ai and Chen (2003) (see also Newey and Powell (2003)) show that this is a consistent estimator for  $\boldsymbol{\eta}$ , and derive the asymptotically normal limiting distribution for the parametric components of  $\boldsymbol{\theta}$ . Our primary interest is estimation of the functions  $\omega^1(\cdot)$  and  $\omega^2(\cdot)$ . Ai and Chen (2003) show that, under suitable assumptions, the rate of convergence of infinite dimensional components of  $\boldsymbol{\theta}$  like these is faster than  $n^{1/4}$ .

## 6 Simulation Study

A simulation study is employed to assess the finite sample performance of the estimator derived in Section 5. For simplicity we focus on estimation of the conditional mean of  $Y^*$ . The simulation

design is

$$\begin{aligned} Y^* &= g(X^*) + U, & U &\sim N(0, \sigma_U^2), \\ X^* &= 1 + 0.4Q - V, & Q &\sim N(5, 1.5), \quad V \sim N(0, 0.3^2) \end{aligned}$$

where  $\sigma_U^2$  is set so that the  $R^2$  of the regression of  $Y^*$  on  $X^*$  is roughly 0.75. Three different specifications for the conditional mean function  $g(\cdot)$  are considered. The first is the standard Working (1943) and Leser (1963) parametric specification of Engel curves, corresponding to budget shares linear in the logarithm of  $X^*$ . The others are a third order polynomial Engel curve and a Fourier function Engel curve. The choice of the parameters for each of the three specifications makes

$$\begin{aligned} g_1(X^*) &= X^* - 0.5X^* \log(X^*) \\ g_2(X^*) &= 0.8X^* + 0.02X^{*2} - 0.03X^{*3} \\ g_3(X^*) &= 4 - 2 \sin(2\pi(X^* - 0)/4) + 0.5 \cos(2\pi(X^* - 0)/4) \end{aligned}$$

Data  $(Y^*, X^*)$  are assumed to be contaminated by measurement errors, so what is observed is the couple  $(Y, X)$  given by

$$\begin{aligned} Y &= Y^* + X^*S, \\ X &= X^*W, \end{aligned}$$

with  $W = S + \tilde{S} + 1$ ,  $E[S] = E[\tilde{S}] = 0$ . The variances of measurement errors  $S$  and  $\tilde{S}$  are chosen such that  $Var[\log X^*] \approx Var[\log W]$ , so half of the variation in the observed  $\log X^*$  is measurement error. This is a substantial amount of measurement error, though it is roughly comparable to what we later find empirically.

We draw 1000 samples of 500, 1000 and 2000 observations from these three data generating processes corresponding to  $g_1(\cdot)$ ,  $g_2(\cdot)$  and  $g_3(\cdot)$ . For each of these samples the conditional mean function of  $Y^*$  given  $X^*$  is estimated by the sieve estimator proposed in Section 5.

To substantially reduce the computational burden we do not estimate and employ correlations



between  $\psi(\mathbf{w}_i; \boldsymbol{\theta})$  and the remaining components of the vector  $\boldsymbol{\rho}(\mathbf{w}_i; \boldsymbol{\theta})$ , by setting the corresponding elements of the GMM weighting matrix  $\Sigma(\mathbf{w}_i)$  to zero. This preserves consistency and asymptotic normality while significantly reducing the complexity of the estimator. The only downside of this restriction on the weighting matrix is that it potentially reduces asymptotic efficiency, which if anything is likely to make our estimator look worse rather than better, compared to alternatives.

Results are compared to three other estimators: the one proposed by Lewbel (1996), which assumes the parametric Working (1943) and Leser (1963) linear in logarithms budget share functional form for  $g_1(X^*)$ ; a nonparametric sieve estimator, which ignores the presence of measurement error; and the infeasible nonparametric sieve estimator computed on the unobserved data  $Y^*$  and  $X^*$ . The latter is considered in order to compare our results with the ideal oracle alternative scenario in which measurement error is not an issue.

We set  $N_m = 2$  and  $N_f = 3$ , with  $\alpha_0 = 1$  and  $\alpha_1 = \alpha_2 = 0$  so that the resulting density is suitably normalized to have zero mean and unit variance. We select  $J = 100$  in order to lower the computational burden of the algorithm, while the constants  $a_1$  and  $b_1$  are chosen so that the corresponding interval contains all of the observations for  $X$ , resulting in  $a_1 = 0$  and  $b_1 = 6$ . The values of  $N_1$  considered are 2, 3 and 4, while the set of instruments is given by a constant,  $Z$  and  $\log(Z)$ . These instruments, which correspond to the  $c_j(\cdot)$  functions in the previous section, are chosen to match standard functional forms commonly used in empirical Engel curve analyses, including linear and Working-Leser models.

To compare estimators we calculate a measure of the distance between each median estimated curve with the true one. The measure considered is the Integrated Mean Squared Error (IMSE) also considered by Ai and Chen (2003), defined as:

$$IMSE = \frac{(v_I - v_0)}{I} \sum_{i=1}^I (\hat{\omega}(v_i) - g(v_i))^2,$$

where  $(v_0, \dots, v_I)$  is a sufficiently fine equally spaced grid of points over which the comparison is made and  $\hat{\omega}(x)$  is the median over all the estimated curves in  $x$ .

Results are summarized in Tables 1-3, where the IMSE is calculated for all the combinations of  $N$  and  $N_1$  for each of the specifications considered above.

The infeasible estimator based on data that is not mismeasured far outperforms the feasible

estimators, showing that the cost of measurement error on estimation accuracy is substantial. Not knowing the correct functional form of the Engel curve is also quite costly, as can be seen by comparing our proposed estimator to the parametric estimator proposed by Lewbel (1996) in Table 1. Our proposed estimator performs significantly better than the feasible estimators that ignore measurement error, and significantly better than the parametric estimator when that estimator misspecifies the Engel curve.

## 7 Empirical Application

We provide an application of the estimator derived in Section 5 using data from the US *Consumer Expenditure Survey*. The sample considered is the same as in Battistin, Blundell, and Lewbel (2009), using data for the range of years 2001 to 2003. We restrict our attention to couples, composed of husband and wife, in which the male is aged between 35 and 65.<sup>6</sup> The final sample consists of 1149 households.

We focus on the estimation of Engel curves for food and clothing, using real income as an instrument. We implement the estimator derived in Section 5 to estimate the conditional mean and variance of food and clothing expenditures given true total expenditures  $X^*$ , where  $N_1$  and  $N_2$  are set equal to 2.<sup>7</sup> To limit the potential effect of simulation errors resulting from the standard normal random draws used for approximating the integrals, we select  $J = 1000$ .<sup>8</sup>

The estimated conditional mean functions along with analogous curves obtained by several alternative estimators are reported in Figures 1 and 2. One alternative is a parametric model, assuming the linear in logarithms Working-Leser budget share specification. The parametric model is estimated both with and without allowing for the presence of measurement errors (the former being obtained by applying the estimator discussed in Lewbel 1996). We also consider nonparametric estimates based on sieves, which either ignore the presence of measurement errors, or allow for the presence of measurement errors only in total expenditures  $X$  (hence relying on the moment conditions discussed in Schennach (2007) which is equal to imposing  $\lambda = 0$  in our model). These

<sup>6</sup>Reported expenditures are deflated by the annual US *Consumer Price Index*, and are deflated by the number of household members.

<sup>7</sup>We also considered the case of  $N_1 = N_2 = 3$ , but that appeared to overfit the data, resulting in highly variable estimates of the coefficients of interest, thus suggesting the choice for the degree of smoothing of  $N_1 = N_2 = 2$ .

<sup>8</sup>We further trim data by 0.5% on both  $Y$  and  $X$  in an attempt to mitigate effects of the well known sensitivity of sieve estimators to outliers.

are all compared to our proposed estimator which allows for both nonparametric functional forms of the Engel curves and allows for correlated measurement errors in both  $X$  and  $Y$ .

A check on the adequacy of our model is to compare the estimates of the variance of  $W$ , i.e., the measurement error in total expenditures  $X$ , which should be asymptotically the same in the food and clothing equations. The estimated variances are in fact quite similar, equalling 0.098 in the food equation and 0.086 in the clothing equation. This corresponds to a noise to signal ratio for observed levels of total expenditures of 0.353 and 0.315 respectively <sup>9</sup>, meaning that roughly one third of the variance of observed total expenditures  $X$  may be attributed to measurement error, which is a rather large estimate.

It is well documented that the Working-Leser log linear budget share Engel curve model fits food demand reasonably well, but not clothing (see, e.g., the survey Lewbel (2008) and references therein). We similarly find evidence that food but not clothing is close to Working-Leser. The estimated variance of  $W$  in the Working-Leser food equation is 0.138, not too far from our nonparametric estimate. In contrast, the estimated variance of  $W$  in the Working-Leser clothing equation is negative (  $-0.588$ ), providing strong evidence that clothing is not Working-Leser.

The estimated correlation coefficients between  $W$  and measurement errors in food and clothing are 0.068 and 0.024 respectively. This implies roughly that seven and three percent of the standard deviation of measurement errors in total expenditures  $X$  are accounted for by measurement errors in food and clothing respectively. As Figures 1 and 2 show, accounting for these measurement errors visibly alters the estimated Engel curves.

Figures 3 and 4 show the estimated conditional variance functions, defined as  $Var(Y^*|X^*)$ , for food and clothing. Reported are i) nonparametric estimates obtained by ignoring measurement errors in both  $Y$  and  $X$ , ii) accounting for measurement error in  $X$  alone (hence applying the Schennach 2007 estimator to the second conditional moment) and iii) accounting for measurement errors in both  $Y$  and  $X$  by implementing our proposed second moment estimator as described in section 5. These results suggest that measurement error is responsible for a large portion of the observed conditional variance of the observed  $Y$ .

It has long been known that for most goods,  $Var(Y|X)$  increases with  $X$ . For example, this

---

<sup>9</sup>Following equation (11) we have  $X = X^* + X^*(W - 1)$ , hence the estimated noise to signal ratio is given by  $Var(X^*(W - 1))/Var(X^*)$

features prominently in Hildenbrand (1994) (see chapter 3 on increasing dispersion), and is the reason why Engel curves are often estimated in budget share form (since regressing  $Y/X$  on  $X$  reduces the heteroskedasticity of the error term relative to regressing  $Y$  on  $X$ ). Figures 3 and 4 clearly show this feature in the uncorrected estimates. However, the estimates of variance after correcting for measurement error are not increasing, which suggests that this well documented feature of empirical Engel curves may be at least in part an artifact of correlated measurement errors in  $X$  and  $Y$  rather than a significant feature of underlying behavior.

## 8 Conclusions

We have considered identification and estimation of conditional moments of  $Y^*$  given  $X^*$  when both are mismeasured (so we instead observe  $Y$  and  $X$ ) and the measurement errors in  $Y$  and  $X$  are correlated with each other. We showed nonparametric identification of  $E(Y^*|X^*)$  under general conditions, and we showed identification of higher moments  $E(Y^{*k}|X^*)$  for  $k > 1$  given some additional structural assumptions that, in the case of Engel curves at least, follow from the definitions and construction of  $Y$  and  $X$ .

Given identification, we then proposed an Ai and Chen (2003) type nonparametric sieve based GMM estimator of these conditional moments. Our identification and estimation do not require strong a priori functional form restrictions. We verified with a simulation study that in finite samples our estimator greatly reduces mean squared error relative to alternative available estimators.

An empirical application was also provided to the estimation of food and clothing Engel curves. The results indicate the presence of relatively substantial measurement errors in recorded total expenditures, and the presence of measurement errors in both food and clothing expenditures that correlate with the measurement error in total expenditures. Accounting for these jointly correlated measurement errors produces moderate changes in the shape and location of the estimated Engel curves, and generates more pronounced changes in the estimates of  $Var(Y^*|X^*)$ . These latter estimates suggest that the well documented increasing dispersion property of Engel curves is likely due at least in part to correlated measurement errors.

## References

- AI, C., AND X. CHEN (2003): “Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions,” *Econometrica*, 71(6), 1795–1843.
- BATTISTIN, E., R. BLUNDELL, AND A. LEWBEL (2009): “Why Is Consumption More Log Normal than Income? Gibrat’s Law Revisited,” *Journal of Political Economy*, 117(6), 1140–1154.
- BERAN, R., A. FEUERVERGER, AND P. HALL (1996): “On nonparametric estimation of intercept and slope distributions in random coefficient regression,” *The Annals of Statistics*, 24(6), 2569–2592.
- BLUNDELL, R., M. BROWNING, AND I. A. CRAWFORD (2003): “Nonparametric Engel Curves and Revealed Preference,” *Econometrica*, 71(1), 205 – 240.
- BLUNDELL, R., X. CHEN, AND D. KRISTENSEN (2007): “Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves,” *Econometrica*, 75(6), 1613 – 1669.
- CARROLL, R., X. CHEN, AND Y. HU (2010): “Identification and estimation of nonlinear models using two samples with nonclassical measurement errors,” *Journal of Nonparametric Statistics*, 22(4), 379–399.
- CARROLL, R., D. RUPPERT, L. STEFANSKI, AND C. CRAINICEANU (2006): *Measurement Error in Nonlinear Models*. Chapman & Hall, second edition edn.
- CARROLL, R. J., A. DELAIGLE, AND P. HALL (2011): “Testing and Estimating Shape-Constrained Nonparametric Density and Regression in the Presence of Measurement Error,” *Journal of the American Statistical Association*, 106(493), 191–202.
- CHEN, X. (2007): “Large Sample Sieve Estimation of Semi-Nonparametric Models,” in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, chap. 76. Elsevier.
- CHEN, X., H. HONG, AND D. NEKIPELOV (2011): “Nonlinear Models of Measurement Errors,” *Journal of Economic Literature*, 49(4), 901–937.
- CHESHER, A. (2003): “Identification in nonseparable models,” *Econometrica*, 71(5), 1405–1441.

- DELAIGLE, A., J. FAN, AND R. J. CARROLL (2009): “A Design-Adaptive Local Polynomial Estimator for the Errors-in-Variables Problem,” *Journal of the American Statistical Association*, 104(485), 348–359.
- FOX, J., K. KIM, S. RYAN, AND P. BAJARI (2012): “The random coefficients logit model is identified,” *Journal of Econometrics*, 166(2), 204–212.
- GALLANT, A. R., AND D. D. W. NYCHKA (1987): “Semi-Nonparametric Maximum Likelihood Estimation,” *Econometrica*, 55(2), 363 – 390.
- GRENANDER, U. (1981): *Abstract Inference*. Wiley (New York).
- HAUSMAN, J., W. K. NEWEY, H. ICHIMURA, AND J. L. POWELL (1991): “Identification and estimation of polynomial errors-in-variables models,” *Journal of Econometrics*, 50(3), 273–295.
- HAUSMAN, J., W. K. NEWEY, AND J. L. POWELL (1995): “Nonlinear errors in variables Estimation of some Engel curves,” *Journal of Econometrics*, 65(1), 205–233.
- HILDENBRAND, W. (1994): *Market Demand: Theory and Empirical Evidence*. Princeton University Press.
- HODERLEIN, S., AND E. MAMMEN (2007): “Identification of marginal effects in nonseparable models without monotonicity,” *Econometrica*, 75(5), 1513–1518.
- HODERLEIN, S., L. NESHEIM, AND A. SIMONI (2011): “Semiparametric estimation of random coefficients in structural economic models,” *Unpublished Manuscript*.
- IMBENS, G. W., AND W. K. NEWEY (2009): “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity,” *Econometrica*, 77(5), 1481–1512.
- LESER, C. E. V. (1963): “Forms of Engel Functions,” *Econometrica*, 31(4), 694–703.
- LEWBEL, A. (1996): “Demand Estimation with Expenditure Measurement Errors on the Left and Right Hand Side,” *The Review of Economics and Statistics*, 78(4), 718.
- LEWBEL, A. (2008): “Engel Curves,” entry for *The New Palgrave Dictionary of Economics*, 2nd Edition.

- LEWBEL, A., AND K. PENDAKUR (2009): “Tricks with Hicks: The EASI Demand System,” *The American Economic Review*, 99(3), 827–863.
- (2011): “Generalized Random Coefficients With Equivalence Scale Applications,” *Unpublished Manuscript, Boston College*.
- LI, T. (2002): “Robust and consistent estimation of nonlinear errors-in-variables models,” *Journal of Econometrics*, 110(1), 1–26.
- LIGHTHILL, M. J. (1962): *Introduction to Fourier Analysis and Generalized Functions*. Cambridge University Press, London.
- MATZKIN, R. (2003): “Nonparametric estimation of nonadditive random functions,” *Econometrica*, 71(5), 1339–1375.
- (2007): “Nonparametric Identification,” in *Handbook of Econometrics*, chap. 73, pp. 5307–5368.
- MEISTER, A. (2011): “Rate-optimal nonparametric estimation in classical and Berkson errors-in-variables problems,” *Journal of Statistical Planning and Inference*, 141(1), 102–114.
- NEWKEY, W. K. (2001): “Flexible Simulated Moment Estimation of Nonlinear Errors-in-Variables Models,” *Review of Economics and Statistics*, 83(4), 616–627.
- NEWKEY, W. K., AND J. L. POWELL (2003): “Instrumental Variables Estimation of Nonparametric Models,” *Econometrica*, 71(5), 1565–1578.
- RUMMEL, D., T. AUGUSTIN, AND H. KÜCHENHOF (2010): “Correction for Covariate Measurement Error in Nonparametric Longitudinal Regression,” *Biometrics*, 66(4), 1209–1219.
- SCHENNACH, S. M. (2004): “Estimation of Nonlinear Models with Measurement Error,” *Econometrica*, 72(1), 33–75.
- (2007): “Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models,” *Econometrica*, 75(1), 201 – 239.
- SCHENNACH, S. M. (2008): “Quantile Regression with Mismeasured Covariates,” *Econometric Theory*, 24, 1010–1043.

- WANG, L., AND C. HSIAO (2011): “Method of moments estimation and identifiability of semiparametric nonlinear errors-in-variables models,” *Journal of Econometrics*, 165(1), 30–44.
- WORKING, H. (1943): “Statistical Laws of Family Expenditure,” *Journal of the American Statistical Association*, 38(221), 43–56.
- ZINDE-WALSH, V. (2014): “Measurement error and deconvolution in spaces of generalized functions,” *Econometric Theory*, 30, 1207–1246.



# Tables

Table 1: Integrated Mean Squared Error – Working-Leser Specification

		$N_1 = 2$	$N_1 = 3$	$N_1 = 4$
<b>N = 500</b>	<b>Proposed Nonparametric</b>	0.0278	0.0250	0.0374
	<b>Lewbel Working-Leser</b>	0.0050	0.0044	0.0043
	<b>Sieve OLS</b>	0.1264	0.1157	0.1367
	<b>Infeasible Sieve OLS</b>	0.0029	0.0016	0.0002
<b>N = 1000</b>	<b>Proposed Nonparametric</b>	0.0194	0.0145	0.0077
	<b>Lewbel Working-Leser</b>	0.0029	0.0020	0.0017
	<b>Sieve OLS</b>	0.1338	0.1216	0.1272
	<b>Infeasible Sieve OLS</b>	0.0025	0.0013	0.0001
<b>N = 2000</b>	<b>Proposed Nonparametric</b>	0.0075	0.0061	0.0044
	<b>Lewbel Working-Leser</b>	0.0013	0.0014	0.0012
	<b>Sieve OLS</b>	0.1249	0.1249	0.1238
	<b>Infeasible Sieve OLS</b>	0.0023	0.0015	0.0001

Table 2: Integrated Mean Squared Error – Polynomial Specification

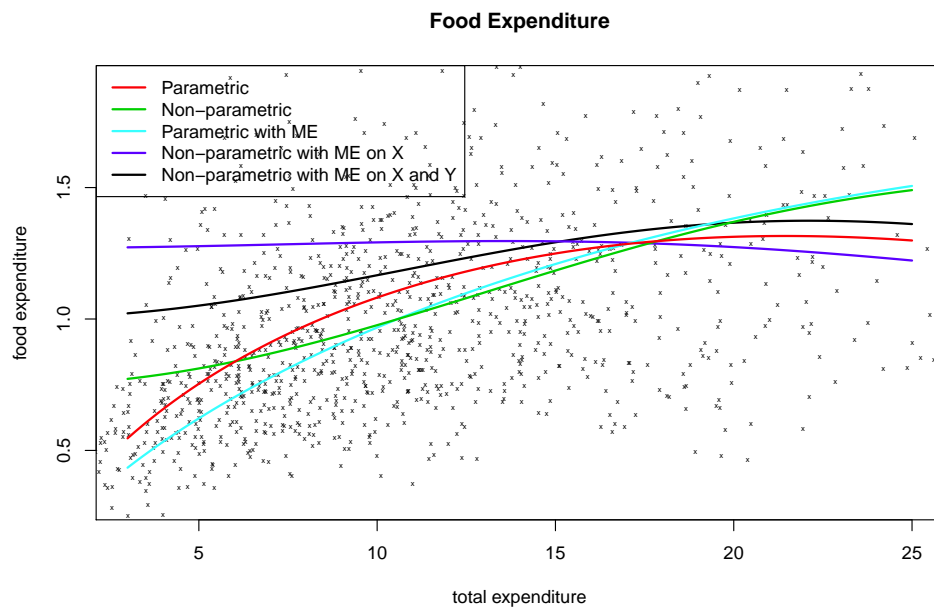
		$N_1 = 2$	$N_1 = 3$	$N_1 = 4$
<b>N = 500</b>	<b>Proposed Nonparametric</b>	0.0412	0.0316	0.0419
	<b>Lewbel Working-Leser</b>	0.4450	0.3887	0.4236
	<b>Sieve OLS</b>	0.4352	0.3664	0.4120
	<b>Infeasible Sieve OLS</b>	0.0164	0.0070	0.0008
<b>N = 1000</b>	<b>Proposed Nonparametric</b>	0.0317	0.0252	0.0201
	<b>Lewbel Working-Leser</b>	0.4439	0.4409	0.4397
	<b>Sieve OLS</b>	0.4186	0.4027	0.4061
	<b>Infeasible Sieve OLS</b>	0.0146	0.0078	0.0007
<b>N = 2000</b>	<b>Proposed Nonparametric</b>	0.0222	0.0134	0.0086
	<b>Lewbel Working-Leser</b>	0.4449	0.4411	0.4472
	<b>Sieve OLS</b>	0.4204	0.3927	0.4154
	<b>Infeasible Sieve OLS</b>	0.0158	0.0075	0.0006

Table 3: Integrated Mean Squared Error – Fourier Specification

		$N_1 = 2$	$N_1 = 3$	$N_1 = 4$
<b>N = 500</b>	<b>Proposed Nonparametric</b>	1.2341	0.7136	0.2643
	<b>Lewbel Working-Leser</b>	17.4530	17.2250	17.4841
	<b>Sieve OLS</b>	6.9265	6.6896	7.2122
	<b>Infeasible Sieve OLS</b>	0.9254	0.5535	0.0321
<b>N = 1000</b>	<b>Proposed Nonparametric</b>	0.6948	0.5832	0.1526
	<b>Lewbel Working-Leser</b>	17.6906	18.1249	17.5878
	<b>Sieve OLS</b>	7.0111	6.7868	7.0255
	<b>Infeasible Sieve OLS</b>	0.8377	0.5154	0.0173
<b>N = 2000</b>	<b>Proposed Nonparametric</b>	0.6220	0.4513	0.0600
	<b>Lewbel Working-Leser</b>	17.8650	18.0427	17.8537
	<b>Sieve OLS</b>	7.1276	6.7988	6.9826
	<b>Infeasible Sieve OLS</b>	0.8377	0.4767	0.0113

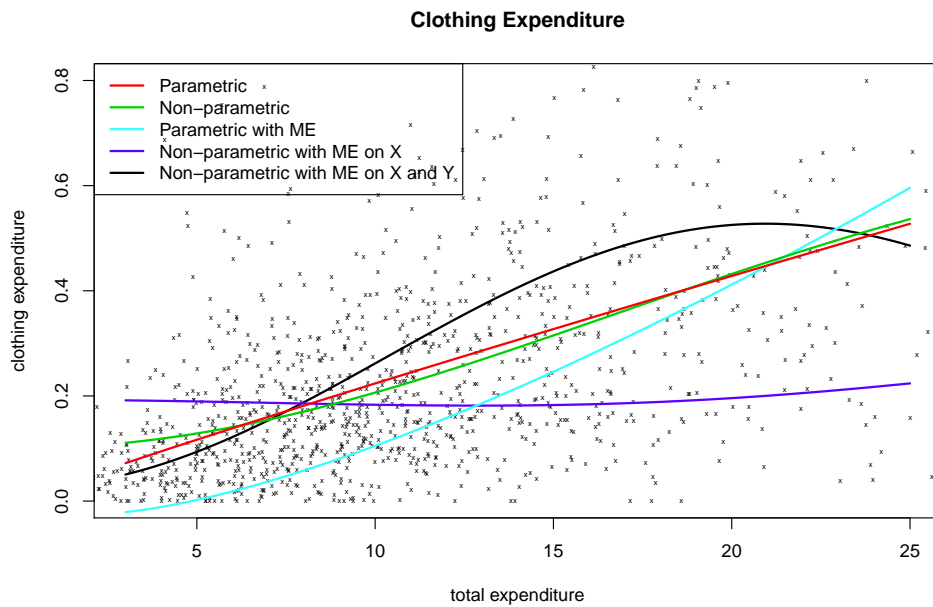
# Figures

Figure 1: Food expenditure Engel curves



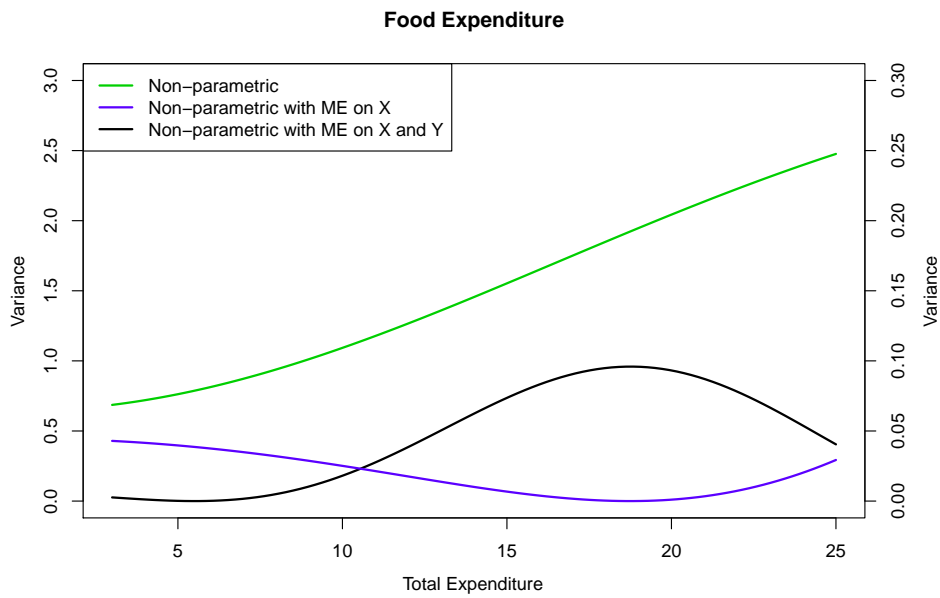
**Notes.** Estimated conditional mean functions for  $Y^*$  given  $X^*$ . Parametric estimate ignoring measurement error (red line), parametric estimate accounting for measurement error in both  $Y$  and  $X$  (light-blue line), non-parametric estimate ignoring measurement error (green line), non-parametric estimate accounting for measurement error in  $X$  (blue line) and non-parametric estimate accounting for measurement error in both  $Y$  and  $X$  (black line).

Figure 2: Clothing expenditure Engel curves



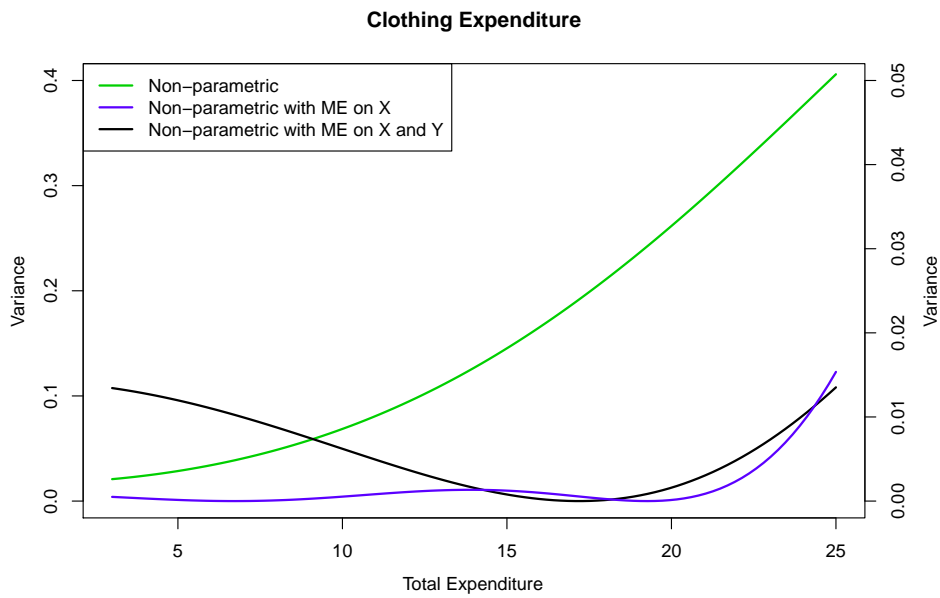
**Notes.** Estimated conditional mean functions for  $Y^*$  given  $X^*$ . Parametric estimate ignoring measurement error (red line), parametric estimate accounting for measurement error in both  $Y$  and  $X$  (light-blue line), non-parametric estimate ignoring measurement error (green line), non-parametric estimate accounting for measurement error in  $X$  (blue line) and non-parametric estimate accounting for measurement error in both  $Y$  and  $X$  (black line).

Figure 3: Conditional variance function for food



**Notes.** Estimated conditional variance functions for  $Y^*$  given  $X^*$ . Non-parametric estimate ignoring measurement error (green line - left hand side axis), non-parametric estimate accounting for measurement error in  $X$  (blue line - right hand side axis) and non-parametric estimate accounting for measurement error in both  $Y$  and  $X$  (black line - right hand side axis).

Figure 4: Conditional variance function for clothing



**Notes.** Estimated conditional variance functions for  $Y^*$  given  $X^*$ . Non-parametric estimate ignoring measurement error (green line - left hand side axis), non-parametric estimate accounting for measurement error in  $X$  (blue line - right hand side axis) and non-parametric estimate accounting for measurement error in both  $Y$  and  $X$  (black line - right hand side axis).

## Appendix

### Derivation of equations (5) and (6).

First consider the conditional expectation of  $Y^k$  given  $Z$ :

$$\begin{aligned} E[Y^k | Z] &= E \left[ E \left[ Y^k | X^*, Z \right] | Z \right], \\ &= E \left[ E \left[ Y^k | X^* \right] | Z \right], \\ &= E \left[ \mu^k(x^*) | Z \right], \end{aligned}$$

where the second equality follows from Assumption 1 (ii). Next consider the conditional expectation of  $XY$  given  $Z$ :

$$\begin{aligned} E[XY | Z] &= E \left[ E[XY | X^*, Z] | Z \right], \\ &= E \left[ (X^* + W) \left[ H(X^*, U) + X^{*l} S \right] | Z \right], \\ &= E \left[ X^* H(X^*, U) | Z \right] + E \left[ X^{*l+1} S | Z \right] + E \left[ WH(X^*, U) | Z \right] + E \left[ X^{*l} WS | Z \right], \end{aligned}$$

and by Assumption 1 (i) and (ii) we get:

$$E \left[ X^* H(X^*, U) | Z \right] + E \left[ X^{*l} WS | Z \right],$$

which under Assumption 1 (iv) yields:

$$E \left[ X^* H(X^*, U) | Z \right] + E \left[ X^{*l} | Z \right] E[WS],$$

and by applying iterated expectations as above, under Assumption 1 (ii) we obtain:

$$\begin{aligned} &E \left[ E \left[ X^* H(X^*, U) | X^*, Z \right] | Z \right] + E \left[ E \left[ X^{*l} | X^*, Z \right] | Z \right] E[WS] = \\ &E \left[ x^* \mu^1(x^*) | Z \right] + E \left[ x^{*l} | Z \right] E[WS]. \end{aligned}$$



Finally consider the multiplicative measurement error structure implied by (11). The conditional expectation of  $Y^k$  given  $Z$  is still obtained as above, while the conditional expectation of  $XY$  given  $Z$  becomes:

$$\begin{aligned} E[XY | Z] &= E[E[XY | X^*, Z] | Z], \\ &= E\left[(X^*W) \left[H(X^*, U) + X^{*l}S\right] | Z\right], \\ &= E[X^*H(X^*, U)W | Z] + E\left[X^{*l+1}SW | Z\right], \end{aligned}$$

and by Assumption 1 (i) and (ii) we get:

$$E[X^*H(X^*, U) | Z] + E\left[X^{*l+1}WS | Z\right],$$

which under Assumption 1 (iv) yields:

$$E[X^*H(X^*, U) | Z] + E\left[X^{*l+1} | Z\right] E[WS],$$

and by applying iterated expectations as above, under Assumption 1 (ii) we obtain:

$$\begin{aligned} E[E[X^*H(X^*, U) | X^*, Z] | Z] + E\left[E\left[X^{*l+1} | X^*, Z\right] | Z\right] E[WS] &= \\ E\left[x^* \mu^1(x^*) | Z\right] + E\left[x^{*l+1} | Z\right] E[WS]. \end{aligned}$$

## Proof of Lemma 1

Taking Fourier transform on both sides of equations (5) and (6) we obtain:

$$\begin{aligned} \varepsilon_{y^k}(\zeta) &= \int \int \mu^k(z - v) dF(v) e^{i\zeta z} dz, \\ &= \int \int \mu^k(x^*) e^{i\zeta(x^* + v)} dx^* dF(v), \\ &= \int \int \mu^k(x^*) e^{i\zeta x^*} dx^* e^{i\zeta v} dF(v), \\ &= \int \mu^k(x^*) e^{i\zeta x^*} dx^* \int e^{i\zeta v} dF(v), \\ &= \gamma_k(\zeta) \phi(\zeta), \end{aligned}$$

and

$$\begin{aligned}
\varepsilon_{xy}(\zeta) &= \int \int (z-v)\mu^1(z-v)dF(v)e^{i\zeta z}dz + \int \int \lambda(z-v)^l dF(v)e^{i\zeta z}dz, \\
&= \int \int x^*\mu^1(x^*)e^{i\zeta(x^*+v)}dx^*dF(v) + \lambda \int \int x^{*l}e^{i\zeta(x^*+v)}dx^*dF(v), \\
&= \int x^*\mu^1(x^*)e^{i\zeta x^*}dx^* \int e^{i\zeta v}dF(v) + \lambda \int x^{*l}e^{i\zeta x^*}dx^* \int e^{i\zeta v}dF(v), \\
&= \left(-\mathbf{i}\frac{\partial}{\partial\zeta} \int \mu^1(x^*)e^{i\zeta x^*}dx^*\right)\phi(\zeta) + \lambda\psi(\zeta)\phi(\zeta), \\
&= -\mathbf{i}\dot{\gamma}_1(\zeta)\phi(\zeta) + \lambda\psi(\zeta),
\end{aligned}$$

hence  $\mathbf{i}\varepsilon_{xy}(\zeta) = \dot{\gamma}_1(\zeta)\phi(\zeta) + \lambda\psi(\zeta)\phi(\zeta)$ . Theorem 1 in Zinde-Walsh (2014) ensures that these are products of Fourier transforms in the space of generalized functions  $S^*$ .

A similar expression holds true under multiplicative measurement error as in equation (11), by replacing  $\psi(\zeta)$  with  $\tilde{\psi}(\zeta)$ .

Q.E.D.

## Proof of Theorem 1

By manipulating (7) and (8) we obtain

$$\varepsilon_{y^k}(\zeta) = \gamma_k(\zeta)\phi(\zeta), \quad (29)$$

$$\mathbf{i}\varepsilon_{(z-x)y}(\zeta) = \gamma_1(\zeta)\dot{\phi}(\zeta) - \lambda\mathbf{i}\psi(\zeta)\phi(\zeta), \quad (30)$$

where  $\mathbf{i}\varepsilon_{(z-x)y}(\zeta) = \mathbf{i}\varepsilon_{zy}(\zeta) - \mathbf{i}\varepsilon_{xy}(\zeta)$  with  $\mathbf{i}\varepsilon_{zy}(\zeta) \equiv \dot{\varepsilon}_y(\zeta) = \dot{\gamma}_1(\zeta)\phi(\zeta) + \gamma_1(\zeta)\dot{\phi}(\zeta)$ . The main point to keep in mind is that  $\varepsilon_y(\zeta)$ ,  $\varepsilon_{(z-x)y}(\zeta)$ ,  $\gamma_k(\zeta)$  and  $\psi(\zeta)$  are generalized functions (see Lighthill 1962), so that some algebraic operations, like the product, between two of them is not defined. On the other hand  $\phi(\zeta)$  is a continuous well-defined ordinary function by Assumption 3.

Solving for  $\gamma_k(\zeta)$  in equation (29) gives

$$\gamma_k(\zeta) = \varepsilon_{y^k}(\zeta)\phi^{-1}(\zeta), \quad (31)$$

where the generalized function  $\varepsilon_{y^k}(\zeta)\phi^{-1}(\zeta)$  is always well defined under Assumptions 2 and 3 (see

Lemma 1 in Zinde-Walsh 2014). If  $\phi(\zeta)$  were known, equation (31) would provide an expression for  $\gamma_k(\zeta)$  in its support, which by Assumption 4 is  $\Omega^k = [-\bar{\zeta}_k, \bar{\zeta}_k]$ . Taking the inverse Fourier transform of  $\gamma_k(\zeta)$  would then produce

$$\mu^k(x^*) = (2\pi)^{-1} \int \gamma_k(\zeta) e^{-i\zeta x^*} d\zeta.$$

In order to prove identification of  $\phi(\zeta)$  we focus on (31) with  $k = 1$ . Substitution into equation (30) yields

$$\mathbf{i}\varepsilon_{(z-x)y}(\zeta) = \varepsilon_{y^1}(\zeta)\phi^{-1}(\zeta)\dot{\phi}(\zeta) - \lambda\mathbf{i}\psi(\zeta)\phi(\zeta), \quad (32)$$

which is a differential equation involving generalized functions. We should keep in mind that the only unknown function in equation (32) is  $\phi(\zeta)$ , since  $\varepsilon_{(z-x)y}(\zeta)$  and  $\varepsilon_{y^1}(\zeta)$  are Fourier transforms of observable quantities and  $\psi(\zeta)$  is a generalized derivative of the Dirac's delta function.<sup>10</sup> In the following we prove uniqueness of the continuous function  $\phi(\zeta)$ , which satisfies the differential equation defined in (32).

First note that the generalized function  $\lambda\mathbf{i}\psi(\zeta)\phi(\zeta)$  has support  $\{0\}$ , being the product of an ordinary continuous function with a generalized derivative of a Dirac's delta function, whose support is  $\{0\}$ . Moreover the support of the generalized function

$$\varepsilon_{y^1}(\zeta)\phi^{-1}(\zeta)\dot{\phi}(\zeta) - \mathbf{i}\varepsilon_{(z-x)y}(\zeta) - \lambda\mathbf{i}\psi(\zeta)\phi(\zeta), \quad (33)$$

is equal to  $\Omega^1 = \{-\bar{\zeta}_1, \bar{\zeta}_1\}$ , by Assumption 4. Consider a generalized function that is a restriction of function (33) to  $\tilde{\Delta} = \Delta^1 \setminus 0$ :

$$\varepsilon_{y^1}(\zeta)\phi^{-1}(\zeta)\dot{\phi}(\zeta) - \mathbf{i}\varepsilon_{(z-x)y}(\zeta), \quad (34)$$

Theorem 3(b) in Zinde-Walsh (2014) shows that in  $\tilde{\Delta}$  there exists a unique function  $\varkappa$  that satisfies

$$\varepsilon_{y^1}(\zeta)\varkappa(\zeta) - \mathbf{i}\varepsilon_{(z-x)y}(\zeta) = 0. \quad (35)$$

---

<sup>10</sup>The actual order of the derivative depends on the structure of the measurement error in  $X$ , as shown in Lemma 1 and Lemma 2. When  $W$  enters additively  $\psi(\zeta)$  is the  $l$ -th generalized derivative of the Dirac's delta function, while if  $W$  enters multiplicatively  $\psi(\zeta)$  is the  $(l + 1)$ -th generalized derivative of the Dirac's delta function. The proof we derive applies to both cases, since the order of derivative in  $\psi(\zeta)$  does not affect identification.

By continuity  $\varkappa$  is defined in all of  $\Delta^1$ . The function  $\phi(\zeta)$  is then obtained as the unique solution to  $\phi(0) = 1$  and  $\phi^{-1}(\zeta)\dot{\phi}(\zeta) = \varkappa(\zeta)$  by

$$\phi(\zeta) = \exp \int_0^\zeta \varkappa(t) dt, \text{ for } \zeta \in \Omega^1. \quad (36)$$

Given continuity of the functions  $\phi(\zeta)$ ,  $\dot{\phi}(\zeta)$  and hence of  $\varkappa(\zeta)$  provided by Assumption 3, perturbation of the function (34) in zero does not affect uniqueness and identification of  $\varkappa(\zeta)$ , which is still obtained as the unique solution to (35), and the function of interest  $\phi(\zeta)$  is given by (36). Finally note that the above argument provides identification of  $\phi(\zeta)$  for  $\zeta \in \Omega^1$ , hence implying that  $\gamma_k(\zeta)$  is identified in  $\Omega^1 \cap \Omega^k$ . Therefore  $\mu^k(x^*)$  is identified under the assumption that  $\Omega^k \subseteq \Omega^1$ .

Q.E.D.

## Proof of Corollary 2

Exploiting the additive nature of measurement error in  $Y^*$ , let us rewrite the k-th conditional moment of the observed  $Y$  as:

$$\begin{aligned} \mu^k(x^*) &= E \left[ \left( H(X^*, U) + X^{*l} S \right)^k \right] \\ &= \sum_{j=0}^k \binom{k}{j} E[H^j(X^*, U) X^{*l(k-j)} S^{k-j} | X_i^*] \\ &= \sum_{j=0}^k \binom{k}{j} \omega^j(x^*) x^{*l(k-j)} E[S^{k-j}], \end{aligned}$$

where the second equality holds because of (ii) in Assumption 1. Noting that  $\mu^0(x^*) = \omega^0(x^*) = 1$  and solving for  $\omega^k(x^*)$  we obtain equation (14). Since by Theorem 1  $\mu_k(x^*)$  is identified, if  $E[S^j]$  for  $j = 2, \dots, k$  is known then  $\omega_k(x^*)$  is identified.

## Proof of Theorem 2

Using equation (11) we rewrite the k-th moment of the observed random variable  $X$  as  $E[X^k] = E[X^{*k}]E[W^k]$ , which implies

$$E[W^k] = \frac{E[X^k]}{E[X^{*k}]}, \quad (37)$$

which is well defined since by assumption  $E[X^{*k}] \neq 0$ .

The first  $K$  moments of  $X$  exist by assumption, so that  $E[Z^k]$  also exists for  $k = 1, \dots, K$ . Under Assumptions 1 to 4, from Theorem 1,  $\phi(\zeta)$  is identified in a neighborhood of the origin, hence  $\phi^{(k)}(0)$  is identified for  $k = 1, \dots, K$  and  $E[\tilde{V}^k] = \mathbf{i}^k \phi^{(k)}(0)$ . Exploiting equation (4), we can write

$$\begin{aligned}
E[X^{*k}] &= E[(Z - \tilde{V})^k], \\
&= E \left[ \sum_{j=0}^k \binom{k}{j} Z^j (-\tilde{V})^{k-j} \right], \\
&= \sum_{j=0}^k \binom{k}{j} E[Z^j] E[(-\tilde{V})^{k-j}], \\
&= \sum_{j=0}^k \binom{k}{j} E[Z^j] (-1)^{k-j} (-\mathbf{i}^{k-j}) \phi^{(k-j)}(0), \\
&= \sum_{j=0}^k \binom{k}{j} \mathbf{i}^{k-j} E[Z^j] \phi^{(k-j)}(0).
\end{aligned} \tag{38}$$

Substitution of equation (38) into (37) yields equation (16).

Similarly from equations (11) and (2), and by Assumptions 1 we have that

$$\begin{aligned}
E[X^k Y | Z] &= E[X^{*k} W^k H(X^*, U) | Z] + E[X^{*k} W^k X^{*l} S | Z], \\
&= E[x^{*k} \mu^1(x^*) | Z] E[W^k] + E[X^{*k+l} | Z] E[W^k S],
\end{aligned}$$

which gives

$$g_k(z) = E[W^k S] h_k(z), \tag{39}$$

with  $h_k(z) = E[X^{*k+l} | Z = z]$  and  $g_k(z) = E[X^k Y | Z = z] - E[x^{*k} \mu^1(x^*) | Z = z] E[W^k]$ . The functions  $g_k(z)$  and  $h_k(z)$  are functions of  $z$  that involve either observable quantities, like  $E[X^k Y | Z]$ , or quantities already shown to be identified. In particular  $h_k(z) = E[X^{*k+l} | Z = z]$

is obtained from knowledge of  $\phi(\zeta)$  as

$$\begin{aligned}
E[X^{*k+l} | Z = z] &= \int (z - v)^{k+l} dF(v), \\
&= \int \sum_{j=0}^{k+l} \binom{k+l}{j} z^j v^{k+l-j} dF(v), \\
&= \sum_{j=0}^{k+l} \binom{k+l}{j} z^j \int v^{k+l-j} dF(v), \\
&= \sum_{j=0}^{k+l} \binom{k+l}{j} z^j (-\mathbf{i})^{k+l-j} \phi^{(k+l-j)}(0). \tag{40}
\end{aligned}$$

Moreover, if  $\bar{\zeta}_1 = \infty$  in Assumption 4, it is

$$\begin{aligned}
\int E[x^{*k} \mu^1(x^*) | Z = z] e^{\mathbf{i}\zeta z} dz &= \int \int (z - v)^k \mu^1(z - v) dF(v) e^{\mathbf{i}\zeta z} dz, \\
&= (-\mathbf{i})^k \gamma_1^{(k)}(\zeta) \phi(\zeta),
\end{aligned}$$

where  $\gamma_1(\zeta)$  is defined as in equation (9) and  $\gamma_1^{(k)}(\zeta)$  is the  $k$ -th derivative of  $\gamma_1(\zeta)$ . Taking the inverse Fourier transform we finally get

$$E[x^{*k} \mu^1(x^*) | Z = z] = (2\pi)^{-1} \int (-\mathbf{i})^k \gamma_1^{(k)}(\zeta) \phi(\zeta) e^{-\mathbf{i}\zeta z} d\zeta. \tag{41}$$

Substitution of equation (41) into the definition of  $g_k(z)$  gives

$$g_k(z) = E[X^k Y | Z = z] - E[W^k] (2\pi)^{-1} \int (-\mathbf{i})^k \gamma_1^{(k)}(\zeta) \phi(\zeta) e^{-\mathbf{i}\zeta z} d\zeta.$$

Q.E.D.

**Derivation of equations (18) and (19).**

From equation (15) it is

$$\begin{aligned}
 E[W^k S] &= E[S[S + (1 + \tilde{S})]^k], \\
 &= E \left[ S \sum_{j=0}^k \binom{k}{j} S^j (1 + \tilde{S})^{k-j} \right], \\
 &= \sum_{j=0}^k \binom{k}{j} E[S^{j+1}] E[(1 + \tilde{S})^{k-j}],
 \end{aligned}$$

where the last equality follows from the independence between  $S$  and  $\tilde{S}$ . Rearranging terms and solving for  $E[S^{k+1}]$  we obtain

$$\begin{aligned}
 E[S^{k+1}] &= E[W^k S] - \sum_{j=0}^{k-1} \binom{k}{j} E[S^{j+1}] E[(1 + \tilde{S})^{k-j}], \\
 &= E[W^k S] - \sum_{j=1}^k \binom{k}{j-1} E[S^j] E[(1 + \tilde{S})^{k-j+1}],
 \end{aligned}$$

which is equivalent to

$$E[S^k] = E[W^{k-1} S] - \sum_{j=1}^{k-1} \binom{k-1}{j-1} E[S^j] E[(1 + \tilde{S})^{k-j}].$$

Similarly from  $E[W^k]$  it is

$$\begin{aligned}
 E[W^k] &= E \left[ [S + (1 + \tilde{S})]^k \right], \\
 &= E \left[ \sum_{j=0}^k \binom{k}{j} S^{k-j} (1 + \tilde{S})^j \right], \\
 &= \sum_{j=0}^k \binom{k}{j} E[S^{k-j}] E[(1 + \tilde{S})^j], \\
 &= E[(1 + \tilde{S})^k] + \sum_{j=0}^{k-1} \binom{k}{j} E[S^j] E[(1 + \tilde{S})^{k-j}],
 \end{aligned}$$

which yields

$$E[(1 + \tilde{S})^k] = E[W^k] - \sum_{j=0}^{k-1} \binom{k}{j} E[S^{k-j}] E[(1 + \tilde{S})^j].$$