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

Michele De Nadai  Arthur Lewbel
University of Padova  Boston College

First version July 2011, revised July 2013

## 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 non-separable model errors. We also provide a nonparametric sieve 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 feature of most consumption data sets has been ignored in almost all previous demand applications. We find accounting for this feature casts doubt on Hildenbrand's (1994) "increasing dispersion" assumption.

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

---

# 1 Introduction

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

Our identification procedure allows us to distinguish measurement errors from other sources of error that are due to unobserved structural or behavioural 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.

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 form, to reduce the resultin error heteroskedasticity. However, 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 becomes considerably weaker, suggesting that this well documented feature of Engel curve estimates may be at least 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 $Y$, but an additional extension is to 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 be multiplicative instead of additive.

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. These Fourier transforms have both an ordinary and a singular component, and identification is based just on the ordinary components. Our main insight is that, if the 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 singular component of the Fourier transform, so identification based on the ordinary components can proceed as before. 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).[1]

Examples of Engel curve estimation with measurement errors include the following. Hausman, Newey, Ichimura, and Powell (1991) and Hausman, Newey, and Powell (1995) provide estimators for polynomial Engel curves with classically mismeasured $X$, Newey (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

---

[1]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 $Y$.

in $X$ and $Y$ as we do, but does so in the context of a parametric model of $Y$ given $X$.[2]

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 respectively. 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

With $i$ indexing observations, suppose that scalar random variables $Y_i^*$ and $X_i^*$ are measured with error, so we only observe $Y_i$ and $X_i$ where:

$$
\begin{aligned}
Y_i &= Y_i^* + S_i, \\
X_i &= X_i^* + W_i,
\end{aligned}
$$

with $S_i$ and $W_i$ being unobserved measurement errors that may be correlated with each other (more general models of measurement errors are considered below). Without loss of generality we can

---

[2]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.

write

$$Y_i^* = H(X_i^*, U_i),$$

where $Y_i^*$ equals an unknown function $H(\cdot, \cdot)$ of a scalar random regressor $X_i^*$, and a random scalar or vector of nonseparable unobservables $U_i$, 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. Our primary goal is identification (and later estimation) of the nonparametric regression function $E\left[H(X_i^*, U_i) \mid X_i^*\right]$, but we more generally consider identification of conditional moments $E\left[H(X_i^*, U_i)^k \mid X_i^*\right]$ 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.[3]

In the statistics literature, a measurement error $W_i$ in $X_i$ is called "differential" if it affects the observed outcome $Y_i$, after conditioning on the true $X_i^*$, that is, if $Y_i \mid X_i^*, W$ does not equal $Y_i \mid X_i^*$ (see, e.g., Carroll, Ruppert, Stefanski, and Crainiceanu (2006)). In our setup $W_i$ is in general differential since $W$ is correlated with the measurement error $S_i$ in $Y_i$, while it would be nondifferential if the true $Y_i^*$ were observed. An alternative application of our identification results would be for a model in which $Y_i$ is not mismeasured, and the additive error $S_i$ instead represents the effect of differential measurement error $W_i$ on the true outcome $Y_i$. To ease exposition this case will no longer be considered, and we will focus on the interpretation in which both $X_i$ and $Y_i$ are mismeasured.

To aid identification, assume we observe instruments $Q_i$ satisfying

$$X_i^* = m(Q_i) + V_i, \tag{1}$$

where the function $m$ is unknown and defined by $m(Q_i) = E(X_i^* \mid Q_i)$, and $V_i \perp Q_i$. This independence assumption, which is also made by Schennach (2007), is common in the control

---

[3]If the moment generating function of $Y_i^*$ given $X_i^*$ exists in an open interval around zero, then $E\left[Y_i^{*k} \mid X_i^*\right]$ is finite for all integers $k$ and our results can be used to identify these moments, which in turn can be used to identify the conditional distribution of $Y_i^*$ given $X_i^*$ via the moment generating function. Then results such as those in Matzkin (2007) can be applied to identify $H(X_i^*, U_i)$. If the conditional distribution of $Y_i^*$ given $X_i^*$ is continuous, then one such result defines $U_i$ to equal the conditional distribution function of $Y_i^*$ given $X_i^*$, and constructs $H$ as the inverse function of this distribution. Alternative identifying assumptions would similarly follow from full knowledge of the conditional distribution function.

function literature when dealing with nonlinear models and will be maintained throughout the paper.

Assume for now that the measurement errors have the classical properties that they are mean zero with $S_i, W_i \perp Y_i^*, X_i^*, Q_i$. This can be equivalently stated as $S_i, W_i \perp U_i, V_i, Q_i$. Our formal results will substantially relax these independence assumptions (specifically, they are replaced by Assumption 1 below). This independence is just made here now to ease exposition.

The two main complications in our set up compared to a classical measurement error problem are that the true regression model $H$ is an unknown function containing nonseparable model errors $U$, and that measurement errors in $Y_i^*$ and $X_i^*$ may be correlated.

Assuming $S_i$ is identically zero so $Y_i^* \equiv Y_i$, Schennach (2007) shows identification of the conditional mean function $E(Y_i^* \mid X_i^*)$, based on features of the Fourier transforms of $E[Y_i \mid m(Q_i)]$ and of $E[Y_i X_i \mid m(Q_i)]$. These transforms are in general not proper functions, but generalized functions, being the Fourier transforms of not-absolutely integrable functions. We follow Schennach (2007) in assuming that the generalized functions that arise in our context may be decomposed into the sum of an ordinary component that is a well behaved function, and a purely singular component, which may be seen as a linear combination of generalized derivatives of the Dirac delta function (see Lighthill 1962 and the supplementary material in Schennach 2007). The conditional mean $E(Y_i^* \mid X_i^*)$ is then shown by Schennach (2007) to be identified by knowledge of just the ordinary components of these Fourier transforms.

The intuition behind our extension builds on the fact that measurement errors $S_i$ and $W_i$ do not affect $E[Y_i \mid m(Q_i)]$, and they affect $E[Y_i X_i \mid m(Q_i)]$ only by adding the term $E[S_i W_i \mid m(Q_i)]$. Under the identifying assumption that the covariance between $S_i$ and $W_i$ does not depend on the instruments $Q_i$ (which is an implication of classical measurement error), this additional term $E(S_i W_i)$ is a constant.

Since the Fourier Transform of a constant is a purely singular generalized function, the ordinary component of the Fourier Transform of $E[Y_i X_i \mid m(Q_i)]$ is unaffected by this additional term, and so the conditional mean function of interest $E(Y_i^* \mid X_i^*)$ remains identified as before. We generalize this argument to also identify higher order conditional moments.

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

(see also Lewbel 1996) is:

$$Y_i = Y_i^* + X_i^* S_i,$$

$$X_i = X_i^* W_i.$$

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

In this Engel curve model we identify $E\left(Y_i^{*k}|X_i^*\right) = E[H(X_i^*, U_i)^k|X_i^*]$ for an arbitrary integer $k$, thereby separating the effects on $Y_i$ of observed and unobserved heterogeneity ($X_i$ and $U_i$) from the measurement errors $S_i$ and $W_i$. For the Engel curve model we provide two different identification strategies, depending on different assumptions regarding the measurement errors. The first approach builds on the fact that for utility derived Engel curves $H(0, U_i) \equiv 0$, while the second approach makes use of the specific dependence structure between $W_i$ and $S_i$ implied by the definition of $Y_i$ and the construction of $X_i$ in the Engel curve framework. We show 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 $Y_i^*$ and $X_i^*$ as $Y_i^* = H(X_i^*, U_i)$ and $X_i^* = m(Q_i) + V_i$, where $Q_i$ is a vector of instruments, $V_i$ is a scalar unobserved random variable independent of $Q_i$, $U_i$ is a vector of unobserved disturbances, and the function $H(\cdot, \cdot)$ is unknown. The scalar random variable $Y_i^*$ is unobserved, but, encompassing and generalizing the examples given in the previous section, assume that the observed $Y_i$ is given by:

$$Y_i = Y_i^* + X_i^{*l} S_i \tag{2}$$

for some non-negative integer $l$, where $E[S_i|X_i^*] = 0$, so that the measurement error $X_i^{*l} S_i$ is mean zero, but has higher moments that can depend on $X_i^*$. 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 $X_i^*$.

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

$$X_i = X_i^* + W_i \qquad \text{with } E[W_i] = 0 \tag{3}$$

or

$$X_i = X_i^* W_i \qquad \text{with } E[W_i] = 1. \tag{4}$$

thereby allowing for either additive or multiplicative measurement errors while retaining the property that $E[X_i] = E[X_i^*]$. Let $\mu^k(x_i^*) = E[Y_i^k | X_i^*]$ be the k-th conditional moment of the observed random variable $Y_i$ given $X_i^*$. The goal of this Section is first to provide identification of $\mu^k(x_i^*)$ for $k = 1, \ldots, K$, given only knowledge of $(Y_i, X_i, Q_i)$ where $X_i$ is defined by either (3) or (4). We then consider identification of moments $E[Y_i^{*k} | X_i^*]$

We assume the following, which will be maintained throughout:

**Assumption 1.** *The random variables $Q_i$, $U_i$, $V_i$, $W_i$ and $S_i$ are jointly i.i.d. and*

   *1.*

  *(i) $E[W_i^k | Q_i, V_i, U_i] = E[W_i^k]$ for $k = 1, \ldots, K$,*

  *(ii) $E[S_i^k | Q_i, V_i, U_i] = E[S_i^k]$ for $k = 1, \ldots, K$,*

 *(iii) $V_i$ is independent of $Q_i$,*

 *(iv) $E[W_i S_i | Q_i] = E[W_i S_i]$.*

The mean independence Assumptions (i) and (ii) with $K = 1$ are standard for errors in variables models and are basically stating that measurement errors are classical. We assume these for higher $K$ because we consider identification of these higher moments, not just the $k = 1$ conditional moment as in standard models. 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.[4]

---

[4]As pointed out by Schennach (2008), this assumption is testable, by looking at the dependence between the estimated residuals of the feasible first stage and the instruments, given the maintained assumptions (i) and (ii).

Assumption (iv) is in fact less restrictive than standard measurement error assumptions. In particular, standard models that do not allow for correlations between measurement errors will trivially satisfy this assumption with $E[W_i S_i | Q_i] = E[W_i S_i] = 0$. Assumption (iv) would also follow from, and is strictly weaker than, the standard classical assumption that measurement errors be independent of correctly measured covariates.

Without loss of generality, assume that $V_i$ has mean zero, so $m(Q_i) \equiv E[X_i | Q_i]$ is nonparametrically identified. Defining $Z_i = m(Q_i)$ and $\tilde{V}_i = -V_i$, we may conveniently rewrite equation (1) as:

$$X_i^* = Z_i - \tilde{V}_i, \tag{5}$$

which we will do hereafter. Following Newey (2001) and Schennach (2007) we will show that, under Assumption 1, knowledge of the conditional moments $E[Y_i^k | Z_i]$, for $k = 1, \ldots, K$, and $E[X_i Y_i | Z_i]$ is enough to identify $\mu^k(x_i^*)$ for $k = 1, \ldots, K$. In the remainder of the paper, for ease of notation, we will drop the subscript $i$ when not needed.

It follows from Assumption 1 that

$$
\begin{aligned}
E[Y^k | Z] &= E\left[ E\left[ Y^k | X^*, Z \right] | Z \right], \\
&= E[\mu^k(x^*) | Z], \tag{6}
\end{aligned}
$$

and, if measurement error in $X^*$ is additive (i.e. if (3) holds), then

$$
\begin{aligned}
E[XY | Z] &= E\left[ (X^* + W) \left[ H(X^*, U) + X^{*l} S \right] | Z \right], \\
&= E[X^* H(X^*, U) | Z] + E[X^{*l} | Z] E[WS | Z], \\
&= E[x^* \mu^1(x^*) | Z] + E[X^{*l} | Z] E[WS]. \tag{7}
\end{aligned}
$$

With a similar argument it may be shown that, if measurement error in $X^*$ is multiplicative (i.e. if (4) holds), then

$$E[XY | Z] = E[x^* \mu^1(x^*) | Z] + E[X^{*l+1} | Z] E[WS]. \tag{8}$$

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.

**Assumption 2.** $|\mu^k(x^*)|$, $|E[Y^k|Z]|$ and $|E[XY|Z]|$ are defined and bounded by polynomials for $x^*$ and $z \in \mathbb{R}$ and for any $k = 1, \ldots, K$.

Assumption 2 essentially excludes specifications for $\mu^k(x^*)$ which rapidly approach infinity like the exponential function and suffices for the following Lemma to hold:

**Lemma 1.** Under Assumption 2, equations (6), (7) and (8) are equivalent to

$$\varepsilon_{y^k}(\zeta) = \gamma_k(\zeta)\phi(\zeta) \tag{9}$$

$$\mathbf{i}\varepsilon_{xy}(\zeta) = \dot{\gamma}_1(\zeta)\phi(\zeta) + \lambda \mathbf{i}\psi(\zeta)\phi(\zeta) \tag{10}$$

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

$$\varepsilon_{y^k}(\zeta) = \int E[Y^k|Z=z]e^{\mathbf{i}\zeta z}dz, \qquad \gamma_k(\zeta) = \int \mu^k(x^*)e^{\mathbf{i}\zeta x^*}dx^*$$

$$\varepsilon_{xy}(\zeta) = \int E[XY|Z=z]e^{\mathbf{i}\zeta z}dz, \qquad \phi(\zeta) = \int e^{\mathbf{i}\zeta v}dF(v)$$

where $F(v)$ is the cdf of $\tilde{V}$, $\lambda = E[WS]$ and $\psi(\zeta) = \int x^{*l}e^{\mathbf{i}\zeta x^*}dx^*$ if (3) holds or $\psi(\zeta) = \int x^{*l+1}e^{\mathbf{i}\zeta x^*}dx^*$ if (4) holds.

Lemma 1, whose proof is given in the Appendix, is a generalization of Lemma 1 in Schennach (2007), who considers the special case where $l = 0$ and $k = 1$. It is important to note that while $\phi(\zeta)$, being the characteristic function of $\tilde{V}_i$, is a proper function, $\varepsilon_{xy}(\zeta)$, $\varepsilon_{y^k}(\zeta)$, $\gamma_k(\zeta)$ and $\psi(\zeta)$ are more abstract generalized functions. Products of generalized functions are not necessarily well defined, so that equations (9) and (10) cannot be manipulated as usual functions to get rid of 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 the suitable 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.** $E[|\tilde{V}|] < \infty$ and $\phi(\zeta) \neq 0$ for all $\zeta \in \mathbb{R}$.

**Assumption 4.** For each $k = 1, \ldots, K$ there exists a finite or infinite constant $\bar{\zeta}_k$ such that $\gamma_k(\zeta) \neq 0$ almost everywhere in $[-\bar{\zeta}_k, \bar{\zeta}_k]$ and $\gamma_k(\zeta) = 0$ for all $|\zeta| > \bar{\zeta}_k$,

**Assumption 5.** *The functions $\gamma_k(\zeta)$ are such that the following decomposition holds:*

$$\gamma_k(\zeta) = \gamma_{k;o}(\zeta) + \gamma_{k;s}(\zeta),$$

*where $\gamma_{k;o}$ is an ordinary function, while $\gamma_{k;s}$ is a purely singular component, which may be seen as a linear combination of generalized derivatives of the Dirac delta function.*

Assumptions 3 and 4 are the equivalent of 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 $\tilde{V}_i$ 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 essentially rules out only specifications for $\gamma_k(\zeta)$ which exhibit sinusoidal behaviours, which is very unlikely in most economic applications.[5] Assumption 5 only excludes pathological (and generally empirically irrelevant) cases in which such decomposition may not hold on a set of measure zero.

The following theorem states the main identification result.

**Theorem 1.** *Under Assumptions 1-5, $\mu^k(x^*)$ for $k = 1, \ldots, K$ are nonparametrically identified. Also, if $\bar{\zeta}_1 > 0$ in Assumption 4 then*

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

*where*

$$\gamma_k(\zeta) = \begin{cases} 0 & \text{if } \varepsilon_{y^k}(\zeta) = 0 \\ \varepsilon_{y^k}(\zeta)/\phi(\zeta) & \text{otherwise,} \end{cases} \tag{11}$$

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

$$\phi(\zeta) = \exp\left(\int_0^\zeta \frac{\mathbf{i}\varepsilon_{(z-x)y,o}(\zeta)}{\varepsilon_{y,o}(\zeta)} d\zeta\right) \tag{12}$$

*and where $\varepsilon_{y,o}(\zeta)$ and $\varepsilon_{(z-x)y,o}(\zeta)$ denote the ordinary function components of $\varepsilon_y(\zeta) = \int E[Y|Z = z] e^{\mathbf{i}\zeta z} dz$ and $\varepsilon_{(z-x)y}(\zeta) = \int E[(Z - X)Y|Z = z] e^{\mathbf{i}\zeta z} dz$ respectively.*

---

[5] 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.

Our Theorem 1 is a generalization of Theorem 1 in Schennach (2007). The proof is in the appendix, but essentially works as follows. By Assumption 5, $\gamma_k(\zeta)$ can be written as the sum of an ordinary function and a linear combination of generalized derivatives of the Dirac delta function, which correspond to the purely singular component. Hence by Lemma 1 a similar decomposition also holds for $\varepsilon_{y^k}(\zeta)$ and $\varepsilon_{xy}(\zeta)$. The last term in (10) is a purely singular generalized function and so only affects the singular component of $\varepsilon_{xy}(\zeta)$. Since Schennach (2007) proved that $\phi(\zeta)$ is identified by knowledge of just the ordinary components of $\varepsilon_y(\zeta)$ and $\varepsilon_{xy}(\zeta)$, allowing for $W$ and $S$ to be correlated does not alter the identification of $\phi(\zeta)$. The function of interest $\gamma_k(\zeta)$ is then obtained from equation (9) as in (11), and its inverse Fourier transform gives $\mu^k(x^*)$.

Theorem 1 provides nonparametric identification of the first $K$ conditional moments of $Y$ given $X^*$. If $Y$ given $X^*$ has a well-defined moment generating function, then applying Theorem 1 with $K$ infinite provides nonparametric identification of the entire conditional distribution of $Y$ given $X^*$, as follows:

**Corollary 1.** *Under the Assumptions of Theorem 1 if $\bar{\zeta}_1 = \infty$ then the characteristic function of the unobserved $X^*$, $\phi_{X^*}(\zeta)$, is identified 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 $Z$.*

This result immediately follows from equation (5) and by noting that equation (12) gives the expression of the characteristic function of $\tilde{V}$. The assumption of $\bar{\zeta}_1 = \infty$ encompasses most of the empirically relevant specifications for the conditional mean, and so is not unduly restrictive. Note that in the proof of Theorem 1 we only considered $\bar{\zeta}_k > 0$ since the case where $\bar{\zeta}_k = 0$ only occurs if $\mu^k(x^*)$ is a polynomial in $X^*$, and that specification has already been shown to be identified by Hausman, Newey, Ichimura, and Powell (1991).

Theorem 1 establishes a set of assumptions under which $\mu^k(x^*) = E[Y^k|X^*]$ is identified for integers $k$. However, the objects of policy relevance are usually the conditional moments of the true, unobserved $Y^*$, that is $\omega^k(x^*) = E[Y^{*k}|X^*]$, or more generally the conditional distribution of $Y^*$ given $X^*$. For $k = 1$ we have $E[Y|X^*] = E[Y^*|X^*]$, so $\omega^1(x^*) = \mu^1(x^*)$ and the primary moment of interest is already identified by Theorem 1. Now consider identification of higher moments, that

is $\omega^k(x^*)$ for $k > 1$.

**Corollary 2.** *Let Assumptions of Theorem 1 hold, then:*

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

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

Corollary 2 shows that what is needed to identify $\omega^k(x^*)$ for $k > 1$ is knowledge of corresponding moments of the unconditional distribution of $S$. Identifying 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 $H(X^*, U)$. Such information needs to be considered on a case by case basis.

In the next section we consider the Engel curve setting where the nature of the variables involved implies the existence of a specific dependence structure between $W$ and $S$, or a boundary restriction on $H(X^*, U)$, either of which will provide sufficient additional information to identify moments of $S$ and hence $\omega^k(x^*)$ using Corollary 2.

## 4 Identification of Engel curves

Let $Y_\ell^*$ be unobserved expenditures on a particular good (or group of goods) $\ell$, for $\ell = 1, ..., L$ and let $X^* = \sum_{\ell=1}^{L} Y_\ell^*$ 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 a vector of individual consumer specific utility (preference) related parameters. The goal is then identification of the conditional distribution of $Y^*$ given $X^*$.

We now describe two different possible models of measurement errors, each of which yields information that may be used to identify 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 $Y_\ell$ on each good $\ell$, so $X = \sum_{\ell=1}^{L} Y_\ell$.

For the first of these two models, assume $Y_\ell = Y_\ell^* + S_\ell$ for each good $\ell$, corresponding to measurement error in the form of equation (2) with $l = 0$. It then follows that the observed

$X = X^* + W$ where the measurement error $W = \sum_{\ell=1}^{L} S_\ell$ is of the form given in equation (3). We may then apply Theorem 1 to identify $\mu^k(x^*)$. To next apply Corollary 2 assume first that one cannot purchase negative amounts of any good, so $Y_\ell^* \geq 0$ since each $Y_\ell^*$ is defined as a level of expenditures. It then follows that the boundary condition $H(0, U) \equiv 0$ holds for all values of $U$, because if $X^* = \sum_{\ell=1}^{L} Y_\ell^* = 0$ and every $Y_\ell^* \geq 0$, then every $Y_\ell^* = 0$. The restriction that $H(0, U) \equiv 0$ in turn implies that $\omega^k(0) = 0$ for all $k \geq 1$, and by equation (14) we obtain

$$E[S^k] = \mu^k(0) - \sum_{j=1}^{k} \omega^j(0)E[S^{k-j}] = \mu^k(0) \tag{15}$$

which, with $\mu^k(x^*)$ identified by Theorem 1, identifies $E[S^k]$, and this in turn permits application of Corollary 2 to identify the moments of interest $\omega^k(x^*)$ for all $k$, and by extension provides identification of the entire conditional distribution of $Y^*$ given $X^*$, assuming the existence of a well defined moment generating function for this distribution.

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's 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 $Y_\ell = Y_\ell^* + X^* S_\ell$ for each good $\ell$, corresponding to measurement error in the form of equation (2) with $l = 1$ for $S = S_1$. Let $\tilde{S} = \sum_{\ell=2}^{L} S_\ell$. Assume that $S$ and $\tilde{S}$ (corresponding to measurement errors for different goods) have mean zero and are independent of each other. Then $X = \sum_{\ell=1}^{L} Y_\ell = X^* \left(1 + S + \tilde{S}\right)$, corresponding to multiplicative measurement error in $X$ given by equation (4) with

$$W = 1 + S + \tilde{S} \tag{16}$$

and $E(W) = 1$ as required by Theorem 1. Given this structure, moments of $S$ are identified given

knowledge of $E[W^k]$ and $E[W^k S]$ for $k = 1, \ldots,$, as follows. Using equation (16) we obtain:

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

which implies:

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

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

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

The following Theorem establishes formal identification for these quantities.

**Theorem 2.** *Let Assumptions 1-4 and equations* (1) *and* (4) *hold. Let the first $K$ moments of $X$ exist finite with $E[X^{*k}] \neq 0$ and $E[X^{*k}|Z] \neq 0$ for every $k = 1, \ldots, K$, then the first $K$ moments of $W$ are identified and*

$$
E[W^k] = \frac{E[X^k]}{\sum_{j=0}^{k} \binom{k}{j} (-\mathbf{i})^{k-j} E[Z^j] \phi^{(k-j)}(0)}. \tag{17}
$$

*Furthermore if $\bar{\zeta}_1 = \infty$ in Assumption 4 then the moments $E[W^k S]$ for $k = 1, \ldots, K - l$ are also identified and*

$$
E[W^k S] = \frac{E[X^k Y|Z = z] - (2\pi)^{-1} E[W^k] \int (-\mathbf{i})^k \gamma_1^{(k)}(\zeta) \phi(\zeta) e^{-\mathbf{i}\zeta z} d\zeta}{\sum_{j=0}^{k+l} \binom{k+l}{j} z^j (-\mathbf{i})^{k+l-j} \phi^{(k+l-j)}(0)}. \tag{18}
$$

*where $\gamma_1^{(k)}(\zeta)$ is the $k$-th derivative of $\gamma_1(\zeta)$ as defined in equation* (11), *while $\phi(\zeta)$ is as in* (12).

The proof of Theorem 2 is given in the Appendix. Intuitively identification of $E[W^k]$ for $k = 1, \ldots, K$ follows by noting that, from equation (4) 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 (4) we have

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

which only involves identified moments apart from $E[W^k S]$, hence proving identification of $E[W^k S]$.

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 instance ruling out symmetric distributions, for which the third central moment would be zero. Furthermore, the assumption of $\bar{\zeta}_1 = \infty$, generally covers empirically relevant frameworks as was discussed in Section 3.

Under the assumptions of Theorems 1 and 2 any conditional moment of the distribution of the unobserved $Y^*$ on $X^*$ is identified. It then follows that the entire conditional distribution of $Y^*$ given $X^*$ is identified, assuming that this conditional distribution has a well defined moment generating function.

# 5  Estimation

In this Section we propose a sieve based nonparametric estimator for the conditional moments of the distribution of $Y^*$ given $X^*$ in the case of Engel curves. Many studies have documented a variety of nonlinearities and substantial unobserved heterogeneity in shapes, see Blundell, Browning, and Crawford (2003) and Lewbel and Pendakur (2009) for instance, or Lewbel (2010) 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 above, 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 based on applying the sieve GMM estimator of Ai and Chen (2003) to the conditional moments used for Theorem 2. For ease of exposition we will just consider estimation of the first two conditional moments, but given our identification results the extension

to higher moments is purely mechanical. The model is

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

The goal is consistent estimation of the functions $\omega^1(\cdot)$ and $\omega^2(\cdot)$, where by construction:

$$Y^* = \omega^1(X^*) + \epsilon_1 \quad \text{and} \quad Y^{*2} = \omega^2(X^*) + \epsilon_2$$

with $E[\epsilon_1|X^*] = E[\epsilon_2|X^*] = 0$ by definition of $\omega^1(\cdot)$ and $\omega^2(\cdot)$. The data consist of an i.i.d. sample of size $N$ from the triple $(Y, X, Q)$.

Based on Theorems 1 and 2 we exploit the moment restrictions implied by equations (6) and (8) and define:

$$\rho_0(\boldsymbol{q}_i; \boldsymbol{\theta}) = y_i - \int \omega^1(\hat{z}_i - \sigma v) f(v) dv, \tag{19}$$

$$\rho_1(\boldsymbol{q}_i; \boldsymbol{\theta}) = y_i^2 - \int \left[ \omega^2(\hat{z}_i - \sigma v) + \lambda(\hat{z}_i - \sigma v)^2 \right] f(v) dv, \tag{20}$$

$$\rho_2(\boldsymbol{q}_i; \boldsymbol{\theta}) = x_i y_i - \int \left[ (\hat{z}_i - \sigma v)\omega^1(\hat{z}_i - \sigma v) + \lambda(\hat{z}_i - \sigma v)^2 \right] f(v) dv, \tag{21}$$

where $\boldsymbol{q}_i = (y_i, x_i, \hat{z}_i)$, $i = 1, \ldots, N$, denotes observations and $\boldsymbol{\theta} = (\lambda, \sigma, \omega^1(\cdot), \omega^2(\cdot), f(\cdot))$, with $f(\cdot)$ being the probability density function of the first stage error term $V$. According to equation (5) $\hat{z}_i$ are the first stage fitted values of a nonparametric regression of the observed $X$ on $Q$. Defining $\boldsymbol{\rho}(\boldsymbol{q}_i; \boldsymbol{\theta}) = (\rho_0(\boldsymbol{q}_i; \boldsymbol{\theta}), \rho_1(\boldsymbol{q}_i; \boldsymbol{\theta}), \rho_2(\boldsymbol{q}_i; \boldsymbol{\theta}))'$, a consistent estimator is then obtained from:

$$E[\boldsymbol{c}(\hat{z}_i)' \boldsymbol{\rho}(\boldsymbol{q}_i; \boldsymbol{\theta})] = 0 \tag{22}$$

for a suitable vector of instruments $\boldsymbol{c}(\hat{z}_i)$.

The computation of (22) is complicated by several factors. First, the parameter vector $\boldsymbol{\theta}$ is infinite-dimensional due to the presence of the unknown functions $\omega^1(\cdot)$, $\omega^2(\cdot)$ and $f(\cdot)$. Second, even if these functions were finitely parametrized, the computation of $\boldsymbol{\rho}_i(\boldsymbol{q}_i; \boldsymbol{\theta})$ would involve integrals (19)-(21) which do not have a closed form solution.

We address both of these issues by adopting a minimum distance sieve estimator along the lines of Ai and Chen (2003), replacing the space $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_f$ with a finite-dimensional sieve

17

space $\mathcal{H}_n = \mathcal{H}_{1n} \times \mathcal{H}_{2n} \times \mathcal{H}_{fn}$ which becomes dense in the original space $\mathcal{H}$ as $n$ increases as in Grenander (1981). Computations involving integrals are simplified by a convenient choice of the sieve space $\mathcal{H}_{fn}$.

We consider cosine polynomial and Hermite polynomial sieve spaces to approximate conditional moments ($\omega^1(x^*)$ and $\omega^2(x^*)$) and the density function $f(v)$ respectively, that is:

$$\omega^j(x^*) \approx \sum_{i=0}^{N_j} \beta_{ij} b_{ij}(x^*), \qquad j = 1, 2, \tag{23}$$

$$f(v) \approx \sum_{i=0}^{N_f} \alpha_i h_i(x^*), \tag{24}$$

for some $N_1$, $N_2$, $N_f$, and where the basis functions $\{b_{ij}(x^*), i = 0, 1, \ldots\}$ and $\{h_i(x^*), i = 0, 1, \ldots\}$ are given by:

$$b_{ij}(x^*) = \cos\left(\frac{i\pi(x^* - a_j)}{b_j - a_j}\right), \qquad h_i(x^*) = H_i(v)\phi(v)$$

for some $a_j, b_j$, $j = 1, 2$. The function $\phi(\cdot)$ is the standard normal density function, while $H_i(\cdot)$ is the i-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 (19)-(21) along the lines of Newey (2001) and Wang and Hsiao (2011). Indeed, by substituting (23) and (24) in (19), (20) and (21) we obtain:

$$\hat{\rho}_0(\boldsymbol{q}_i; \boldsymbol{\eta}) = y_i - \sum_{i=0}^{N_1} \sum_{l=0}^{N_f} \beta_{i1} \alpha_l \int b_{i1}(\hat{z}_i - \sigma v) H_l(v) \phi(v) dv,$$

$$\hat{\rho}_1(\boldsymbol{q}_i; \boldsymbol{\eta}) = y_i^2 - \sum_{i=0}^{N_2} \sum_{l=0}^{N_f} \beta_{i2} \alpha_l \int \left[b_{i2}(\hat{z}_i - \sigma v) + \lambda(\hat{z}_i - \sigma v)^2\right] H_l(v) \phi(v) dv,$$

$$\hat{\rho}_2(\boldsymbol{q}_i; \boldsymbol{\eta}) = x_i y_i - \sum_{i=0}^{N_1} \sum_{l=0}^{N_f} \beta_{i1} \alpha_l \int \left[(\hat{z}_i - \sigma v) b_{i1}(\hat{z}_i - \sigma v) + \lambda(\hat{z}_i - \sigma v)^2\right] H_l(v) \phi(v) dv,$$

where $\boldsymbol{\eta} = (\lambda, \sigma, \alpha_0, \ldots, \alpha_{N_f}, \beta_{10}, \ldots, \beta_{1N_1}, \beta_{20}, \ldots, \beta_{2N_2})$. Thus the integrals involved in $\hat{\boldsymbol{\rho}}(\boldsymbol{q}_i; \boldsymbol{\eta})$ 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. Letting $v_j$ for $j = 1, \ldots, J$ be J random draws from a standard normal distribution, $\hat{\rho}_0(\boldsymbol{q}_i; \boldsymbol{\eta})$ is in this way computed as

$$\hat{\rho}_0(\boldsymbol{q}_i; \boldsymbol{\eta}) = y_i - J^{-1} \sum_{i=0}^{N_1} \sum_{l=0}^{N_f} \beta_{i1} \alpha_l \sum_{j=1}^{J} b_{i1}(\hat{z}_i - \sigma v_j) H_l(v_j).$$

Similar expressions hold for $\hat{\rho}_1(\boldsymbol{q}_i; \boldsymbol{\eta})$ and $\hat{\rho}_2(\boldsymbol{q}_i; \boldsymbol{\eta})$.[6]

It then follows from Theorems 1 and 2 and from Lemma 3.1 in Ai and Chen (2003) that a consistent estimator for $\boldsymbol{\eta}$ is given by

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

with a positive definite matrix $\Sigma(\boldsymbol{q}_i)$. We apply the two-step GMM procedure:

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

$$\arg\min_{\boldsymbol{\eta}} \frac{1}{N} \sum_{i=1}^{N} \hat{\boldsymbol{\rho}}(\boldsymbol{q_i}, \boldsymbol{\eta})' \boldsymbol{c}(z_i) \boldsymbol{c}(z_i)' \hat{\boldsymbol{\rho}}(\boldsymbol{q_i}, \boldsymbol{\eta}).$$

(2) Improve the efficiency of the estimator $\tilde{\boldsymbol{\eta}}$ by applying the minimization:

$$\arg\min_{\boldsymbol{\eta}} \frac{1}{N} \sum_{i=1}^{N} \hat{\boldsymbol{\rho}}(\boldsymbol{q_i}, \boldsymbol{\eta})' \boldsymbol{c}(z_i) \left[ \hat{\Sigma}(\boldsymbol{q}_i) \right]^{-1} \boldsymbol{c}(z_i)' \hat{\boldsymbol{\rho}}(\boldsymbol{q_i}, \boldsymbol{\eta}).$$

where $\hat{\Sigma}(\boldsymbol{q}_i) = \{\hat{\sigma}_{jl}(\boldsymbol{q}_i)\}$ is obtained from $\hat{\boldsymbol{\eta}}$ as:

$$\hat{\sigma}_{jl}(\boldsymbol{q}_i) = \frac{1}{N} \sum_{i=1}^{N} c_j(z_i) c_j(z_i) \hat{\rho}_j(\boldsymbol{q}_i, \hat{\boldsymbol{\eta}}) \hat{\rho}_l(\boldsymbol{q}_i, \hat{\boldsymbol{\eta}}).$$

Ai and Chen (2003) and Newey and Powell (2003) show that this is a consistent estimator for $\hat{\boldsymbol{\eta}}$, and derive the asymptotically normal limiting distribution for the parametric components of $\boldsymbol{\theta}$.

---

[6]While $N_1$, $N_2$ and $N_f$ 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 should be set as large as is computationally practical, analogous to the choice of the fineness of the grid in ordinary numerical integration.

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 the estimation of the conditional mean of $Y^*$. The simulation design is:

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

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) 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 is:

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

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

20

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. Results are compared with 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 alternative scenario in which measurement errors is not an issue.

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

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

Table 2: Integrated Mean Squared Error – Polynomial Specification

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

We set $N_f = 3$, 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

Table 3: Integrated Mean Squared Error – Fourier Specification

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

and 4, while the set of instruments is given by a constant, $Z$ and $\log(Z)$.

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 $(v_0, \ldots, v_I)$ 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.[7] The final sample is composed 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.[8] To limit the dependency of the result from the standard normal random draws used for approximating the integrals we select $J = 1000$. [9]

The estimated conditional mean functions along with analogous curves obtained by several alternative estimators are reported in Figures 1 and 2. Parametric estimates are obtained from the linear in logarithms Working-Leser budget share specification, with and without allowing for the presence of measurement errors (the former being obtained by applying the estimator discussed in Lewbel 1996). Nonparametric estimates are based on sieves which either completely 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), and our full model which 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 $X$, which should be asymptotically the same in the two models. The estimated variances are 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 [10], meaning that roughly one third of the variance of 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 (2010) and refer-

---

[7]Reported expenditures are deflated by the annual US *Consumer Price Index*, and are deflated by the number of household members.

[8]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 $N_1 = N_2 = 2$.

[9]We further trim data by 0.5% on both $Y$ and $X$ in an attempt to avoid the well known sensitivity of sieve estimators on outliers.

[10]Following equation (4) it is: $X = X^* + X^*(W - 1)$, hence the estimated noise to signal ratio is given by $Var(X^*(W - 1))/Var(X^*)$

Figure 1: Food expenditure Engel curves
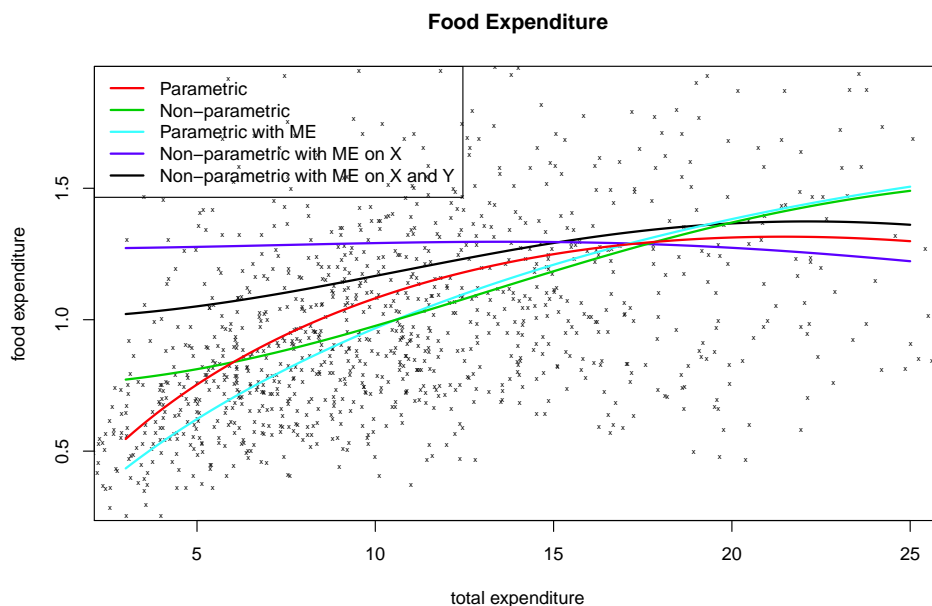


**Food Expenditure**

**Notes.** Estimated conditional mean functions for $Y^*$ given $X^*$. Parametric estimate ignoring measurement error (red line), parametric estimate accounting for mearurement 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).

ences 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 can be seen from figures, accounting for these measurement errors does visibly alter the estimated Engel curves.

Figures 3 and 4 show the estimated conditional variance functions, defined as $Var(Y^*|X^*)$. Reported are non-parametric estimates obtained by ignoring measurement errors in both $Y$ and $X$, acounting for measurement error in $X$ alone (hence applying Schennach 2007 estimator to the second conditional moment) and accounting for measurement error in both $Y$ and $X$ as a result of the implementation of the estimator proposed in section 5. These suggest that measurement error

Figure 2: Clothing expenditure Engel curves
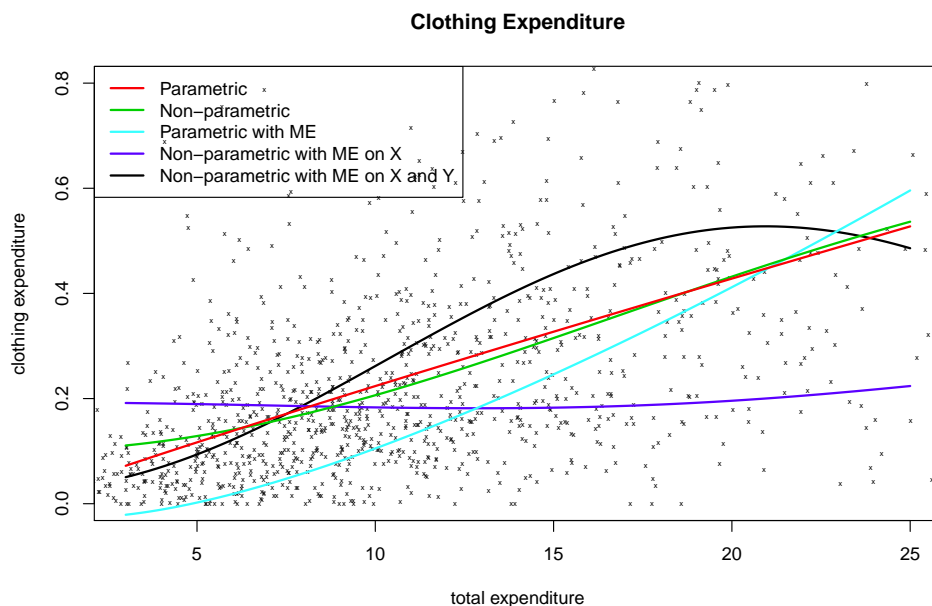
**Clothing Expenditure**



**Notes.** Estimated conditional mean functions for $Y^*$ given $X^*$. Parametric estimate ignoring measurement error (red line), parametric estimate accounting for mearurement 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).

is responsible for a large part 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 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 Engel curve estimates may be at least in part an artifact of measurement errors rather than a feature of behavior.

# 8    Conclusions

We have considered identification and estimation of conditional moments of $Y$ given $X$ when both are mismeasured and measurement errors are correlated with each other. We showed nonparametric
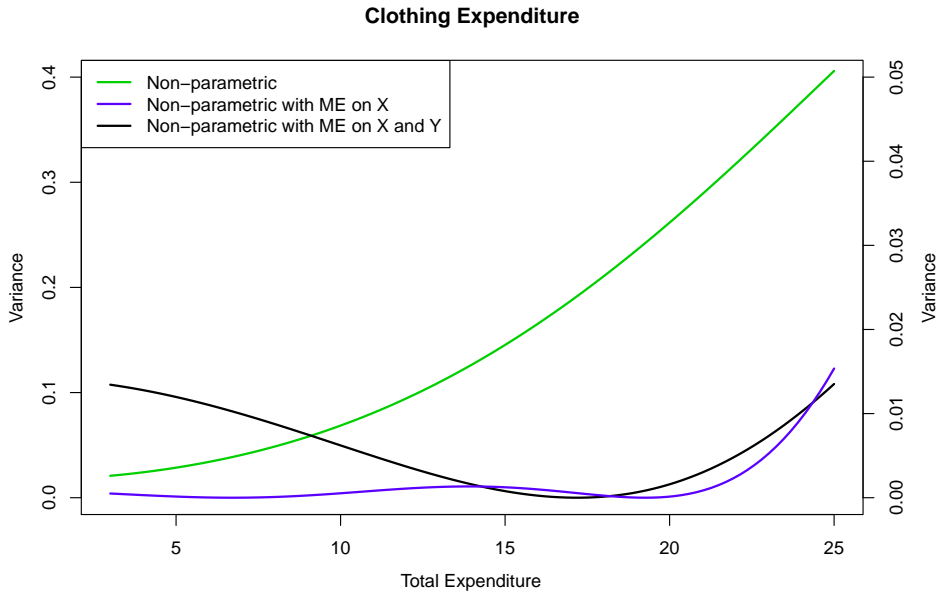
Figure 3: Conditional variance function for food
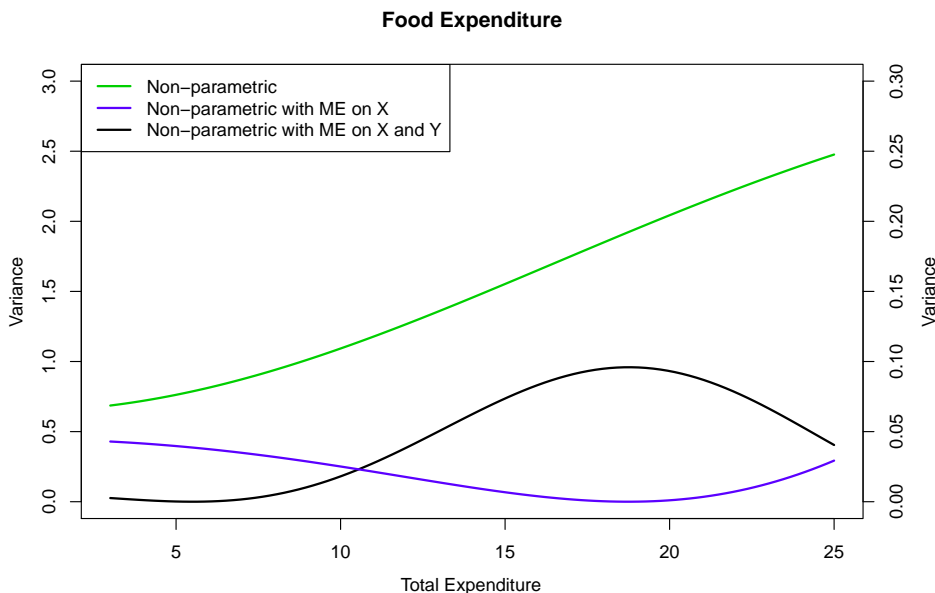


**Food Expenditure**

**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).

identification of conditional moments for a general class of measurement error models. Identification of higher moments requires some structural assumptions that, in the case of Engel curves, follow from the definitions and construction of $Y$ and $X$.

Given identification, we proposed a nonparametric estimator based on a sieve approximation of the conditional moments which takes the form of a conditional GMM estimator. The 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 error 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 measurement errors produces moderate changes in the shape and location of these 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

Figure 4: Conditional variance function for clothing

**Clothing Expenditure**



Non-parametric
Non-parametric with ME on X
Non-parametric with ME on X and Y

**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).

likely to be due at least in part to measurement errors rather than consumer behavior.

# 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.

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.

———— (2010): "Shape-Invariant Demand Functions," *The Review of Economics and Statistics*, 92(3), 549–556.

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.

NEWEY, W. K. (2001): "Flexible Simulated Moment Estimation of Nonlinear Errors-in-Variables Models," *Review of Economics and Statistics*, 83(4), 616–627.

NEWEY, 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.

# Appendix

## Proof of Lemma 1

Under Assumption 1, if equation (3) holds, 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 $\lambda = E[WS]$. Now taking the Fourier transform on both sides of the equation we obtain:

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

and

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

hence $\mathbf{i}\varepsilon_{xy}(\zeta) = \dot{\gamma}_1(\zeta)\phi(\zeta) + \mathbf{i}\lambda\psi(\zeta)\phi(\zeta)$. A similar expression holds true under multiplicative measurement error as in equation (4).

<div align="right">Q.E.D.</div>

**Proof of Theorem 1**

By manipulating (9) and (10) we obtain

$$
\begin{aligned}
\varepsilon_{y^k}(\zeta) &= \gamma_k(\zeta)\phi(\zeta), \\
\mathbf{i}\varepsilon_{(z-x)y}(\zeta) &= \gamma_1(\zeta)\dot{\phi}(\zeta) - \lambda\mathbf{i}\psi(\zeta)\phi(\zeta),
\end{aligned}
\tag{25}
$$

where $\mathbf{i}\varepsilon_{(z-x)y}(\zeta) = \mathbf{i}\varepsilon_{zy}(\zeta) - \mathbf{i}\varepsilon_{xy}(\zeta)$ and $\mathbf{i}\varepsilon_{zy}(\zeta) \equiv \dot{\varepsilon}_y(\zeta) = \dot{\gamma}_1(\zeta)\phi(\zeta) + \gamma_1(\zeta)\dot{\phi}(\zeta)$. 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 Lighthill 1962, so that the product operation between two of them is not defined. However by Assumption 5 $\gamma_k(\zeta)$ may be decomposed into the sum of an ordinary function (denoted with subscript $o$) and a purely singular function (denoted with subscript $s$), so that

$$
\begin{aligned}
\varepsilon_{y,o}(\zeta) + \varepsilon_{y,s}(\zeta) &= \left[ \gamma_{1;o}(\zeta) + \gamma_{1;s}(\zeta) \right] \phi(\zeta), \\
\mathbf{i}\varepsilon_{(z-x)y,o}(\zeta) + \mathbf{i}\varepsilon_{(z-x)y,s}(\zeta) &= \left[ \gamma_{1;o}(\zeta) + \gamma_{1;s}(\zeta) \right] \dot{\phi}(\zeta) - \lambda\mathbf{i}\psi(\zeta)\phi(\zeta).
\end{aligned}
$$

Note that we do not put subscripts on $\phi(\zeta)$ or $\psi(\zeta)$, since the former is an ordinary function, being the Fourier transform of an absolutely integrable function, and the latter is a purely singular function as $\psi(\zeta) = 2\pi\delta(\zeta)$. Now applying Lemma 2 in Schennach (2007), which states that the product of an ordinary function with an ordinary function is an ordinary function, whereas the product of a purely singular component with an ordinary function is purely singular, and equating the ordinary components in the above equations we obtain equations (50) and (51) in Schennach (2007), namely

$$
\begin{aligned}
\varepsilon_{y,o}(\zeta) &= \gamma_{1;o}(\zeta)\phi(\zeta), \\
\mathbf{i}\varepsilon_{(z-x)y,o}(\zeta) &= \gamma_{1;o}(\zeta)\dot{\phi}(\zeta).
\end{aligned}
$$

These are now all ordinary functions and they may be manipulated as in Schennach (2007) to obtain identification of $\phi(\zeta)$ for $\zeta \in [-\bar{\zeta}_1, \bar{\zeta}_1]$ with

$$
\phi(\zeta) = \exp\left( \int_0^\zeta \frac{\mathbf{i}\varepsilon_{(z-x)y,o}(\zeta)}{\varepsilon_{y,o}(\zeta)} d\zeta \right).
\tag{26}
$$

We restrict our attention to the case in which $\bar{\zeta}_1 > 0$, however $\bar{\zeta}_1 = 0$ only when the conditional mean $E[Y^*|X^*]$ is a polynomial in $X^*$, a case which has already been shown to be identified by Hausman, Newey, Ichimura, and Powell (1991). Substitution of (26) in (25) leads to

$$\gamma_k(\zeta) = \begin{cases} 0 & \text{if } \varepsilon_{y^k}(\zeta) = 0 \\ \varepsilon_{y^k}(\zeta)/\phi(\zeta) & \text{otherwise,} \end{cases}$$

and hence by taking the inverse Fourier transform of $\gamma_k(\zeta)$:

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

proving the identification of $\mu^k(X^*)$.

<div align="right">Q.E.D.</div>

## 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 $j = 2, \ldots, k$ is known then $\omega_k(x^*)$ is identified.

## Proof of Theorem 2

Since the first $K$ moments of $X$ exist by assumption then $E[Z^k]$ also exists. Under Assumptions 1 to 4, by Theorem 1, $\phi(\zeta)$ is shown to be identified in a neighborhood of the origin, hence $\phi^{(k)}(0)$ is

<div align="center">33</div>

also identified for $k = 1, \ldots, K$ and, exploiting equation (5), we may write:

$$
\begin{aligned}
E[X^{*k}] &= \sum_{j=0}^{k} \binom{k}{j} E[Z^j] E[(-\tilde{V})^{k-j}], \\
&= \sum_{j=0}^{k} \binom{k}{j} E[Z^j] (-1)^{k-j} (-\mathbf{i}^{k-j}) \phi^{(k-j)}(0), \\
&= \sum_{j=0}^{k} \binom{k}{j} (-\mathbf{i})^{k-j} E[Z^j] \phi^{(k-j)}(0).
\end{aligned}
\tag{27}
$$

The k-th moment of the observed $X$, because of equation (4), may be written as $E[X^k] = E[X^{*k}]E[W^k]$, which implies:

$$
E[W^k] = \frac{E[X^k]}{E[X^{*k}]},
\tag{28}
$$

which is well defined since by assumption $E[X^{*k}] \neq 0$. Substitution of equation (27) into (28) yields equation (17).

Equation (18) follows by noting that from equation (4) 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

$$
E[W^k S] = \frac{E[X^k Y | Z = z] - E[X^{*k} \mu^1(x^*) | Z] E[W^k]}{E[X^{*k+l} | Z = z]}.
\tag{29}
$$

The right hand side of equation (29) involves functions that are either observable, like $E[X^k Y | Z]$, or already shown to be identified. In particular $E[X^{*k+l} | Z]$ is obtained from knowledge of $\phi(\zeta)$ as

$$
\begin{aligned}
E[X^{*k+l} | Z = z] &= \int (z - v)^{k+l} dF(v) = \int \sum_{j=0}^{k+l} \binom{k+l}{j} z^j v^{k+l-j} dF(v), \\
&= \sum_{j=0}^{k+l} \binom{k+l}{j} z^j \int v^{k+l-j} dF(v), \\
&= \sum_{j=0}^{k+l} \binom{k+l}{j} z^j (-\mathbf{i})^{k+l-j} \phi^{(k+l-j)}(0).
\end{aligned}
\tag{30}
$$

On the other hand, 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 (11) and by taking the inverse Fourier transform we get:

$$
E[X^{*k}\mu^1(x^*)|Z = z] = (2\pi)^{-1}\int (-\mathbf{i})^k \gamma_1^{(k)}(\zeta)\phi(\zeta)e^{-\mathbf{i}\zeta z}d\zeta. \tag{31}
$$

Substitution of equations (30) and (31) into (29) gives:

$$
E[W^k S] = \frac{E[X^k Y|Z = z] - (2\pi)^{-1}E[W^k]\int (-\mathbf{i})^k \gamma_1^{(k)}(\zeta)\phi(\zeta)e^{-\mathbf{i}\zeta z}d\zeta}{\sum_{j=0}^{k+l}\binom{k+l}{j}z^j(-\mathbf{i})^{k+l-j}\phi^{(k+l-j)}(0)}.
$$

Q.E.D.