# ESTIMATING PRICE ELASTICITIES WITH NONLINEAR ERRORS IN VARIABLES

Aprajit Mahajan*

*Abstract*—This paper estimates a price elasticity using a flexible demand specification on survey data where prices are observed with errors and are correlated with household characteristics. The demand function is modeled as a polynomial/trigonometric in the unobserved true prices, and the form of the dependency between the observed prices and household characteristics is modeled parametrically. I identify and estimate the model by adapting the approach of Hausmann et al. (1991) and Schennach (2004). The flexible specifications allow us to observe that price elasticities vary across the price distribution, something missed in previous work using linear demand specifications.

## I. Introduction

THIS paper contributes to the literature on the estimation of demand responses using spatial variation in prices derived from household survey data. The measure of prices obtained from such data is typically noisy and obtained in a manner so that the measurement error in prices is usually correlated with measurement error in quantities purchased. The object of interest in these exercises and in this paper is a price elasticity of demand to be estimated flexibly from a single cross-section of household data when prices are measured with error. Deaton (1997) estimates elasticities accounting for measurement error by imposing a linear structure on the outcome equation, while Deaton and Ng (1998) estimate elasticities completely nonparametrically but do not deal with the measurement error problem.

This paper serves as a midway point between the previous papers by estimating the elasticities using flexible (although still parametric) functional forms while also accounting for the mismeasured prices. The approach is an extrapolation of the ideas contained in Hausmann et al. (1991; henceforth HNIP) and the extensions in Schennach (2004) but incorporating the clustered feature of the data and the nonclassical nature of the measurement error.

Dividing expenditure by quantity provides a measure of price that is referred to in the literature as the unit value and is available on a household basis for households that make positive purchases. The basic idea is to use these unit values to estimate demand responses. However, unit values are not prices since they reflect, in part at least, the quality choices made by households (so that, for instance, richer households report higher unit values for the same aggregate commodity than do poorer households) and also measurement errors in expenditures or quantities. Deaton (1987, 1988) develops a framework for dealing with both problems in which the regression function for demand is modeled linearly. His results suggest that the quality issue is of smaller magnitude than the measurement error problem. Deaton and Ng (1998) calculate the price elasticity nonparametrically using the method of Hardle and Stoker (1989) but do not control for the measurement error problem.

This paper offers a more general specification for the regression function for quantity while accounting for measurement error in a manner essentially analogous to HNIP and Schennach (2004), but at the cost of ignoring the quality issue. I model the regression function of interest as a polynomial in the (unobserved) price as well as a polynomial in expenditure to account for possible nonlinearities in income effects. In addition, I experiment with Fourier flexible functional forms for the demand function for whose estimation I need a suitable extension to the HNIP approach.

The paper is organized as follows: Section II describes the data at hand and outlines the various demand specifications. Section III discusses the identification of these specifications. Section IV uses the identification results from the previous section to outline an estimation strategy and derives the large sample properties of the proposed estimators. I then discuss the performance of the estimators in the context of Monte Carlo experiments and implement them on the data set at hand in section VI. The conclusion follows.

## II. Data and Model

Before defining the model, it is instructive to describe the data since their key features will motivate the model. I use data from India's National Sample Survey (50th round, 1993–1994) for rural Maharashtra, which has information on 4,440 households from 445 villages. Households are sampled using a two-stage stratified sample design. In the first stage, $C$ clusters (villages) are chosen, and in the second, $H$ households are sampled from the selected clusters. The data record household characteristics (expenditure and age-sex-caste demographics, which I denote by $x_{ch}$), as well as total expenditure on wheat and total quantity purchased ($q_{ch}$) by each household. The commodity of interest is wheat,[1] a diet staple in the region, which is purchased by most households (about 82% in the survey), and its average budget share in total expenditure is about 3%.

The sampling design of the survey provides observations on several households within one cluster. I assume that the price of wheat is fixed within a cluster but varies exogenously across clusters, so if prices were observed, I could identify the demand parameters, which are assumed constant within and between clusters, from the between-cluster

Received for publication July 27, 2004. Revision accepted for publication March 21, 2008.

* Stanford University.

I thank Han Hong, Bo Honoré, and Elie Tamer for their tireless encouragement and advice. I have also benefited enormously from discussions with Angus Deaton, Ernst Schaumburg, and members of the Princeton Microeconometrics Reading Group.

[1] The figures are based on the aggregate commodity, which includes market wheat, suji, and bread.

*The Review of Economics and Statistics,* November 2009, 91(4): 793–805

variation in prices. I am implicitly assuming that each village constitutes a different market, that prices vary exogenously across markets, and that this variation is independent of other village characteristics that influence household demand. The exogenous variation in prices can be based on the geographical dispersion of the sampled clusters across space, coupled with high transportation costs. These assumptions, while extremely stringent, are common to all exercises in estimating demand responses from survey data where prices are derived in the fashion described above, and their plausibility needs to be weighed against the advantage of having a rich data set with substantial price variation, usually hard to come by in a time series context in developing countries (see Deaton, 1987, 1988).

In addition, I assume that within clusters, household quantities are a small enough fraction of aggregate demand so that I can ignore simultaneity between prices and quantities in the regression specification (note that this disaggregation argument is not unproblematic; see Kennan, 1989). Finally, I abstract from further complex survey design issues and assume that across clusters, the observations are independent and identically distributed (i.i.d.).

The model is specified by two equations for each household: one for quantities and one for unit values (the latter is referred to as the measurement error equation). Following Deaton (1988), I parameterize the error equation as

$$u_{ch} = z_c + x'_{ch}\gamma_2 + \eta_{ch}, \tag{1}$$

where $u_{ch}$ is the log of the unit value for household $h$ in cluster $c$; $z_c$ is the (log of the) unobserved cluster price; $x_{ch}$ is a vector of household characteristics, including expenditure, which is allowed to enter nonlinearly; and $\eta_{ch}$ is the household-specific measurement error. The logarithmic transformation is adopted for convenience and, if $\gamma_2 = 0$, could be interpreted as implying that the error term $\eta_{ch}$ is expressed (approximately) as a fraction deviation from the true price. In subsequent work, I plan to experiment with untransformed unit values so that the errors will be treated as being in levels.

There is no theoretical justification for this particular parameterization, and it differs from the typical (also usually atheoretical) statement of classical measurement error because of the presence of the household characteristics. Their linearity in (1) is purely pragmatic, but clearly their presence itself is important because I need to account for household heterogeneity in some fashion (since richer households on average report higher unit values).

From (1) I define the purged unit value $v_{ch}$ as

$$\begin{aligned} v_{ch} &\equiv u_{ch} - x'_{ch}\gamma_2 \\ &= z_c + \eta_{ch}, \end{aligned} \tag{2}$$

which states that the residual from the best linear prediction of the unit value $u_{ch}$ given household characteristics $x_{ch}$ can be expressed as the sum of two random variables: a village-specific term ($z_c$) that I interpret as the true (unobserved) price and a household-specific term ($\eta_{ch}$) that is interpreted as measurement error.

I specify three possible demand equations. The first (D1) expresses quantities purchased as a polynomial function of the unobserved price

$$\ln q_{ch} = \sum_{j=0}^{K} \beta_j (z_c)^j + x'_{ch}\gamma_1 + f_c + \varepsilon_{ch}. \tag{D1}$$

$q_{ch}$ is the quantity demanded by household $h$ in cluster $c$, and $f_c$ is a cluster-level effect independent of both the prices and the demographics $x_{ch}$. $K$ is the order of the polynomial in the unobserved price variable,[2] and for $K > 1$ the model is not linear in the measurement error $\eta_{ch}$ so that standard instrumental variable approaches are no longer feasible. The independence of the cluster-level effect and prices is quite strong but necessary for identification. A possible scenario is that prices vary across villages solely due to transportation costs and that these variations are independent of any other village-level characteristics that affect household demand.

The second specification (D2) follows Deaton and Muellbauer (1980),

$$w_{ch} = \sum_{j=0}^{K} \beta_j (z_c)^j + x'_{ch}\gamma_1 + f_c + \varepsilon_{ch}, \tag{D2}$$

where $w_{ch}$ is the budget share of wheat for the household.

The final specification (D3) follows the suggestion of Gallant (1981) and Eubank and Speckman (1990), and the regression function is modeled using a Fourier expansion by adding trigonometric functions to the quadratic terms. The simple Fourier flexible form estimated in this paper is

$$\begin{aligned} w_{ch} = \sum_{j=0}^{K} \beta_j (z_c)^j + x'_{ch}\gamma_1 + \sum_{m=1}^{M} \beta_{K+m} \sin(mz_c) \\ + \beta'_{K+m} \cos(mz_c) + f_c + \varepsilon_{ch}. \end{aligned} \tag{D3}$$

The justification for this specification arises from both the observation that exclusively polynomial expansions are often sensitive to outliers and the general arguments that a combination of polynomial and trigonometric terms often has desirable approximation properties (see, e.g., Gallant, 1981, for an exposition on the advantages of such a specification).

In all three specifications, the error term $\varepsilon_{ch}$ is independent of the random vector $\{z_c, x_{ch}\}$ and is independent across households. However, it is allowed to be correlated

[2] In principle, one could formulate the problem still more flexibly by letting $K$ in all the specifications depend on the sample size. This generalization and its implications for the empirical work are left for future research.

with $\eta_{ch}$ for a given household, and this is most likely the case since measurement errors in quantities translate quite directly into measurement errors in unit values as the latter is obtained by dividing expenditure by the former. We assume that $(\varepsilon_{ch}, \eta_{ch})$ have mean 0 and are i.i.d. across households and that all required higher moments exist. These assumptions are all stated formally below.

All three demand specifications attempt to account for the effect of household characteristics on unit values. In all of them, the difference between unit value and its expectation conditional on household variables $x$ is assumed to be representable as a sum of two independent random variables:

$$v_h - \mathbb{E}(v_h|x_h) = z + \eta_h.$$

In principle, one could obtain $E(v|x)$ nonparametrically if the dimension of $x$ is not too large. For the data we use, graphical displays suggest that this conditional expectation is approximately linear when $x$ is (the log of) expenditure and household size, so we may not lose too much by directly modeling the conditional expectation as linear in $x$.

## III. Identification

Throughout, we assume (for expositional ease) that the number of households in each village is fixed and that the econometrician has available an i.i.d. sample from the distribution of the random vector $\{q_h, u_h, x_h\}_{h=1}^H$, where $h$ indexes households. We will also discuss identification assuming that the values of the parameters $(\gamma_1, \gamma_2)$ are known so that the purged unit value $v_{ch}$ is identified. This is without loss of generality, since under the assumptions stated now, these objects are consistently estimable by a variety of standard techniques. For instance, under assumption 1, the parameter vector $(\gamma_1, \gamma_2)$ is identified:

**Assumption 1.** The error terms $(\varepsilon_{ch}, \eta_{ch})$ have 0 mean and are uncorrelated with the household demographics $x_{ch'}$ for all households $h' \in \{1, \ldots H\}$, and the matrix $\sum_{h=1}^H \mathbb{E}\{\ddot{x}_{ch}\ddot{x}'_{ch}\}$ has full rank.[3]

### A. *Identification of Specification D1 and D2*

The basic idea for identification in the first two specifications is straightforward, and we outline the basic idea before stating the result. Note that if prices were observed for each village (so that the econometrician observed an i.i.d. sample from the random vector $\{\{q_h, u_h, x_h\}_{h=1}^H, z\}$) and were independent of the village effect, the parameter vector $\beta$ in each of (D1) and (D2) would be identified (e.g., as the probability limit of the pooled OLS estimates) since

$$\beta = \mathbb{E}(\tilde{z}_{cK}\tilde{z}'_{cK})^{-1}\mathbb{E}\left(\frac{1}{H}\sum_{h=1}^{H} \tilde{z}_{cK}y_{ch}\right), \tag{3}$$

where $\tilde{z}_K = (1, z, z^2, \ldots, z^K)'$ where $y_h = \ln q_h - x'_h\gamma_1$ for the first specification and $y_h = w_h - x'_h\gamma_1$ for the second specification. In the case where $z$ is not observed, $\beta$ would still be identified if we could identify the outer product matrix $\mathbb{E}(\tilde{z}_K\tilde{z}'_K)$ and the vector $\mathbb{E}(\frac{1}{H}\sum_{h=1}^H \tilde{z}_K y_h)$. The basic idea behind identification is that these objects are identified as long as enough structure is placed on the error terms in the measurement error and demand specifications. This is the strategy first outlined in Hausmann et al. (1991), and we modify the argument to account for the clustered nature of the data and the nonclassical nature of the measurement error.

*Identification of* $\mathbb{E}(\tilde{z}_K\tilde{z}'_K)$. To illustrate, as long as the measurement error term $\eta_h$ has zero expectation, $\mathbb{E}(v_{ch}) = \mathbb{E}(z_c)$ so that the mean of the price vector is identified. Next, as long as the error terms are independent across households and independent of prices, the second moment $\mathbb{E}(z_c^2)$ will be identified since for $h \neq h' \mathbb{E}(v_{ch}v_{ch'}) = \mathbb{E}(z_c^2)$. In fact, as we show below, for $j < H$, the $j$th uncentered moment of $z$ is equal to $\mathbb{E}(v_{ch_1}, \ldots, v_{ch_j})$ for $h_s \neq h_k$ for all $(s, k) \in \{1, \ldots, j\}$. We first record this result.

**Assumption 2.** $\mathbb{E}(\eta_{ch}) = 0$ for all $h$ and $(\eta_{ch}, \eta_{ch'})$ are independent of each other for all $h \neq h'$ and $\eta_{ch}$ is independent of $z_c$.

**Lemma 1.** *Suppose that* $\mathbb{E}(z_c^j)$ *exists and that* $j \leq H$ *and that* $\gamma_2$ *is known. Then, under assumption 2,* $\mathbb{E}(z_c^j)$ *is identified. In particular,* $\mathbb{E}(v_{ch_1}, \ldots, v_{ch_j}) = \mathbb{E}(z_c^j)$. *In fact, we can consider all* $\binom{H}{j}$ *distinct subsets of size* $j$ *from the set* $\{v_1, \ldots, v_H\}$ *so that*

$$\mathbb{E}(z_c^j) = \mathbb{E}\left[\frac{1}{\binom{H}{j}}\sum_{h_1=1}^{H-j+1}\sum_{h_2=h_1+1}^{H-j+2}\cdots\sum_{h_j=h_{j-1}}^{H} v_{ch_1}v_{ch_2}\cdots v_{ch_j}\right].$$

Since the data set for this paper typically contains about 10 households per village and the maximum order of polynomials considered is quintic, the condition $j \leq H$ is satisfied for the empirical application, and so identification for the second-moment matrix $\mathbb{E}(\tilde{z}^K\tilde{z}^{K\prime})$ for the application will follow from lemma 1.

If, however, the computation of the second-moment matrix involves the calculation of $\zeta_j \equiv \mathbb{E}(z^j)$ for $j > H$, then we need to strengthen assumption 2 and require that the higher moments of the measurement error $\eta_{ch}$ exist. The $\zeta_j$ are then identified recursively using the formulas below for $q = 0, 1, 2, \ldots$

[3] The double dots above a variable $\ddot{x}_{ch}$ denote that each element in it is expressed as a deviation from the corresponding cluster mean.

$$\mathbb{E}(v_1^{q+1}v_2v_3\cdots v_H)=\sum_{l=0}^{q}\binom{q}{l}\zeta_{H+l}\mathbb{E}(\eta^{q-l})+\sum_{l=0}^{q}\binom{q}{l}\mathbb{E}(\eta^{q-l+1})\zeta_{H+l-1} \tag{4}$$

$$E(v_1^{q+1})=\sum_{l=0}^{q+1}\binom{q+1}{l}\zeta_l\mathbb{E}(\eta^{q+1-l}).$$

**Assumption 3.** $\mathbb{E}(\eta_{ch})=0$ for all $h$ and all higher moments of the error $\mathbb{E}(\eta_{ch}^l)$ and of the prices, $\zeta_j\equiv\mathbb{E}(z^l)$ in the formulas above exist. In addition, $(\eta_{ch}, \eta_{ch'})$ are independent of each other for all $h\neq h'$, and $\eta_{ch}$ is independent of $z_c$.

Under assumption 3, we can identify $\mathbb{E}(z^j)$ for arbitrary $j\in\mathbb{N}$ and $j>H$ recursively. At each step $q$, the second equation in (4) uses the previous information to identify $\mathbb{E}(\eta^{q+1})$, and this moment is used in the first equation in (4) to identify the moment $\zeta_{H+q}$. There are potentially several different ways to identify the $\zeta_j$, and we make no claims for the optimality of the procedure outlined above. An alternative strategy that places somewhat fewer assumptions on the higher moments of the measurement error is outlined in the appendix.

*Identification of* $\mathbb{E}(\tilde{z}_Ky)$. Identification of the inner product $\mathbb{E}(z^sy)$ also follows using the same arguments as in the previous section. To illustrate, consider the directly identified object $\mathbb{E}(v_{h_1}y_{h_2})$ for $h_1\neq h_2$. As long as the error terms for household $h_1$ are independent of the error terms for the other households and also independent of the cluster-level variables $(f_c, z_c)$,[4] then simple calculations show that $\mathbb{E}(v_{ch_1}y_{ch_2})=\mathbb{E}(z_cy_{ch_2})$ for households $h_1\neq h_2$. This result is recorded here for future reference.

**Assumption 4.** Suppose that the error terms $(\varepsilon_{ch}, \eta_{ch})$ have mean zero, are independent across households, and are independent of $(z_c, f_c)$.

**Lemma 2.** *Under assumption 4 and for* $j<H$, $\xi_j\equiv\mathbb{E}(z^jy_{ch})$ *is identified and is equal to* $\mathbb{E}(v_{ch_1},\ldots,v_{ch_j}y_{ch_{j+1}})$ *where* $h_s\neq h_k$ *for* $(s, k)\in\{1,\ldots,j+1\}$

For the data set at hand, this is an adequate identification strategy since the highest cross-moment needed is $\mathbb{E}(y_{ch}z_c^5)$, and there are about ten households per cluster.

*Identification of* $\beta$. If the prices are independent of the village effect in the demand equation, the parameter vector $\beta$ is equal to the expression (3), each of whose two component terms are identified by the arguments above. We record the result here for the empirically relevant case $2K\leq H$. The identification result for the case where $2K>H$ will require further assumptions on the second and higher moments of the error terms (as in assumption 3).

[4] In fact, the independence between the village-level effect $f_c$ and $\varepsilon_{ch}$ is not necessary, but is imposed for ease of exposition and to maintain symmetry between the assumptions on the different error terms.

**Lemma 3.** *Suppose that* $z_c\perp f_c$, *the moments on the right-hand side of (3) exist, and assumptions 1 and 4 hold and* $2K\leq H$. *Then the parameter vector* $\beta$ *is identified.*

## B. *Identification of Specification (D3)*

Following the lines of the previous argument (and assuming for simplicity that $K=2$ and $M=1$), under assumption 1 and if $z_c\perp f_c$, the parameter of interest,

$$\beta=(\mathbb{E}(f(z_c)f(z_c)'))^{-1}\mathbb{E}(f(z_c)y_{ch}),$$

where

$$f(z)=(\tilde{z}_2, \sin(z), \cos(z))'. \tag{5}$$

The only new objects in these moment matrices are $\mathbb{E}(f(z)(\sin(z), \cos(z)))$, and $\mathbb{E}((\sin z, \cos z)'y)$. In a more general setting, Schennach (2004) identifies these moments by identifying the characteristic function of the unobserved variable $z$ and then applying Parseval's identity. This approach can be adapted to the panel data set context in much the same way as the Hausmann et al. (1991) approach: by exploiting the independence of the error terms across households within a cluster.

For expositional ease, assume that there are only two households per cluster. First, since we assume $\gamma$ is identified, then the unit values $\{v_1, v_2\}$ are also identified, and therefore their joint characteristic function $(\mathbb{E}(\exp(i(t_1v_1+t_2v_2))))$ is identified. Substituting into the characteristic function using (2) and using the fact that $\eta_h\perp z$, we obtain

$$\mathbb{E}(\exp(i(t_1v_1+t_2v_2)))=\mathbb{E}(\exp(iz(t_1+t_2)))\mathbb{E}(\exp(it_1\eta_{11}))\mathbb{E}(\exp(it_2\eta_2)).$$

Taking logs on both sides, we obtain

$$\Pi_{v_1,v_2}(t_1, t_2)=\Pi_z(t_1+t_2)+\Pi_{\eta_1}(t_1)+\Pi_{\eta_2}(t_2),$$

where $\Pi_x(s)$ denotes the log of the characteristic function of the random variable $x$ evaluated at the point $s$. Taking the derivative on both sides of the equation above with respect to $t_1$ and evaluating the derivative function at $t_1=0$ and using the fact that the mean of $\eta_2$ is zero, we obtain

$$\frac{d}{dt_1}\Pi_{v_1,v_2}(t_1, t_2)\bigg|_{t_1=0}=\frac{d}{dt_1}\Pi_Z(t_1+t_2)\bigg|_{t_1=0}. \tag{6}$$

The last expression is the derivative of the log of the characteristic function of $z$ evaluated at $t_2$. Define the left-hand side (which is identified) to be $\phi(t_2)$. We can then use the fundamental theorem of calculus and obtain the characteristic function of $z$, $\psi(s)$ as

$$\psi(s) \equiv \exp(\Pi_z(s)) = \exp\left(\int_0^s \phi(t_2)dt_2\right)$$

$$= \exp\left(\int_0^s \frac{\mathbb{E}[iv_1 \exp(it_2v_2)]}{\mathbb{E}[\exp(it_2v_2)]} dt_2\right),$$

assuming enough smoothness to interchange the differentiation and integration operations. Once the characteristic function $\psi$ is known, $\mathbb{E}(q(z))$ for any (measurable) function $q(\cdot)$ can be calculated most cleanly using Parseval's identity,

$$\mathbb{E}(q(z)) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Xi(q(\cdot), -s)\psi(s)ds, \tag{7}$$

where $\Xi(q(\cdot), s) = \int q(t)\exp(its)dt$ is the generalized Fourier transform (GFT) of the function $q$. The GFTs of polynomial and trigonometric functions (and their products) are relatively easy to calculate and are given by the Dirac delta function (denoted by $\delta(s)$) and its derivatives evaluated at a finite set of points.[5] To illustrate,

$$\mathbb{E}(\sin(z)) = \frac{1}{2\pi}\int \Xi(\sin(\cdot), -s)\psi(s)ds$$

$$= \frac{i}{2\pi}(\psi(1) - \psi(-1)).$$

A complete list of the relevant GFTs is available on request. For all the functions $q(\cdot)$ considered here, the integral in (7) reduces to the evaluation of the characteristic function $\psi$ and its derivatives at a finite set of points. Parallel to the argument above, the expectations $\mathbb{E}(yf_j(z))$ for an element $f_j(z)$ of the vector $f(z)$ are similarly identified, and the formula is

$$\mathbb{E}(yf_j(z)) = \frac{-i}{2\pi}\int_{-\infty}^{\infty} \Xi(f(\cdot), -s)\psi(s)\frac{\mathbb{E}(i\exp(isv_2)y_1)}{\mathbb{E}(\exp(isv_2))}ds. \tag{8}$$

As in the preceding argument, this calculation simplifies because of the relatively simple form of the GFTs of the functions being considered.

[5] See, e.g., Lighthill (1996) for a relatively formal treatment of these objects (also see Schennach, 2004).

## IV. Estimation and Large Sample Distribution

The estimation strategy follows the identification argument closely. We first construct the sample versions of the population parameters, which for convenience we define as

$$\theta \equiv (\{\zeta_j\}_{j=1}^{2K}, \{\xi_j\}_{j=1}^{K}). \tag{9}$$

We first discuss the estimation of the $\zeta_j$ parameters (the discussion for estimation of $\xi_j$ is similar). First, if we knew the value of $\gamma_2$, an unbiased and consistent estimator of $\mathbb{E}(z_c^j)$ would be the statistic $\tilde{\zeta}_j$, given by

$$\tilde{\zeta}_j = \frac{1}{C}\sum_{c=1}^{C}\left[\frac{1}{\binom{H}{j}}\sum_{h_1=1}^{H-j+1}\sum_{h_2=h_1+1}^{H-j+2}\cdots\sum_{h_j=h_{j-1}}^{H} v_{ch_1}v_{ch_2}\cdots v_{ch_j}\right].$$

However, since in practice $\gamma_2$ is estimated (see below for more details), we construct $\hat{v}_{ch} = u_{ch} - x_{ch'}\hat{\gamma}_2$, and the estimator is given by

$$\hat{\zeta}_j = \frac{1}{C}\sum_{c=1}^{C}\left[\frac{1}{\binom{H}{j}}\sum_{h_1=1}^{H-j+1}\sum_{h_2=h_1+1}^{H-j+2}\cdots\sum_{h_j=h_{j-1}}^{H} \hat{v}_{ch_1}\hat{v}_{ch_2}\cdots \hat{v}_{ch_j}\right]. \tag{10}$$

Similar arguments lead to

$$\hat{\xi}_j = \frac{1}{C}\sum_{c=1}^{C}\left[\frac{1}{\binom{H}{j}}\sum_{h_1=1}^{H-j+1}\sum_{h_2=h_1+1}^{H-j+2}\cdots\sum_{h_j=h_{j-1}}^{H} \hat{v}_{ch_1}\hat{v}_{ch_2}\cdots \hat{v}_{ch_{j-1}}\hat{y}_{ch_j}\right], \tag{11}$$

where $\hat{y}_{ch} = \ln q_{ch} - x'_{ch}\hat{\gamma}_2$ for the first specification and $\hat{y}_{ch} = w_{ch} - x'_{ch}\hat{\gamma}_2$ for the second specification. In an additional appendix, we outline the asymptotic properties of the estimator $\hat{\theta} = (\{\hat{\zeta}_j\}_{j=1}^{2K}, \{\hat{\xi}_j\}_{j=1}^{K})$ and show that under a standard set of conditions (e.g., as contained in section 7 of Newey & McFadden, 1994),

$$\sqrt{C}\,(\hat{\theta} - \theta) \Rightarrow \mathcal{N}(0, V_\theta).$$

For the first two specifications, the object of interest $\beta$ is a smooth function only of $\theta$[6] given by

$$q(\theta) = \begin{bmatrix} \zeta_0 & \zeta_1 & \cdot & \zeta_K \\ \zeta_1 & \cdot & \cdot & \zeta_{K+1} \\ \cdot & \cdot & \cdot & \cdot \\ \zeta_K & \cdot & \cdot & \zeta_{2K} \end{bmatrix}^{-1} \begin{bmatrix} \xi_0 \\ \cdot \\ \cdot \\ \xi_K \end{bmatrix}$$

so that we can estimate $\beta$ by

$$\hat{\beta} = q(\hat{\theta}).$$

[6] Where $\zeta_0 = 1$ and $\xi_0 = \mathbb{E}(y_{ch})$.

TABLE 1.—MONTE CARLO RESULTS
(500 SIMULATIONS, CLUSTERS = 300, HH = 3)

| Error Correlation | −.25 | | −.5 | | −.75 | |
|---|---|---|---|---|---|---|
| Method | Deaton | HNIP | Deaton | HNIP | Deaton | HNIP |
| Mean (true $\beta = 5$) | 4.976587 | 5.003137 | 4.977571 | 5.004216 | 4.974316 | 5.000928 |
| s.d. | .04987 | .05058 | .05777 | .05854 | .0524 | .052953 |

An application of the delta method therefore suffices to yield the asymptotic normality of the estimator $\hat{\beta}$ with asymptotic variance given by

$$V_\beta = q'(\theta)V_\theta(q'(\theta))^T.$$

For the third specification, in addition to θ, we also need to estimate the moments $\mathbb{E}(\tilde{z}_2(\sin z, \cos z))$, $\mathbb{E}((\sin z, \cos z)'(\sin z, \cos z))$, and $\mathbb{E}((\sin z, \cos z)'y)$. Following the identification argument, we estimate the characteristic function of the price variable $z_c$ and then use Parseval's identity to compute the required moments. Recall that the characteristic function was given by $\psi(t) = \exp(\int_0^t \phi(s)ds)$, where the function $\phi$ is defined in (6) and is estimated by

$$\hat{\phi}(s) = \frac{\sum_{c=1}^{C} (\hat{v}_{c1} \exp(is\hat{v}_{c2}))}{\sum_{c=1}^{C} [\exp(is\hat{v}_{c2})]},$$

where I have assumed for simplicity that each cluster has only two households. In practice, we can compute the numerator and the denominator by taking all possible $\binom{H}{2}$ distinct combinations of the households to form $\hat{\phi}$. Finally, the characteristic function,

$$\hat{\psi}(t) = \exp\left(\int_0^t \hat{\phi}(s)ds\right),$$

is obtained by solving an ordinary differential equation (with the boundary condition that $\psi(0) = 1$). With $\hat{\psi}$ in hand, we can estimate any element $\mathbb{E}(q_{jk}(z))$ of the outer product $\mathbb{E}(f(z)f(z)')$ (recall that $f(z)$ was defined in (5)) as

$$\mathbb{E}[\widehat{q_{jk}(z)}] = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Xi(q_{jk}(\cdot), -s)\hat{\psi}(s)ds. \quad (12)$$

Since these functions $q_{jk}$ are much simpler than the kinds considered by Schennach (2004), I can compute $\Xi(q_{jk}(\cdot), -s)$ analytically rather than adopt the numerical integration routine suggested in her paper. As discussed in the identification section, evaluating the integral $\mathbb{E}(q_{jk}(z))$ reduces to evaluating the derivatives of the characteristic function $\psi$ at a finite number of points, so that in the sample, we compute the expectations by evaluating the sample counterparts of these derivatives. This results in a considerable simplification of the problem. To illustrate, the object $\mathbb{E}(\sin z)$ can be estimated by $(\hat{\psi}(1) + \hat{\psi}(-1))(i/2)$.

We can estimate $\mathbb{E}(yf_j(z))$ in a similar fashion, and the formula is given by (see Schennach, 2004, for more details)

$$\mathbb{E}(\widehat{yf_j(z))} = \frac{-i}{2\pi}\int_{-\infty}^{\infty} \Xi(f(t), t, -s)\hat{\psi}(s)\frac{\sum_{c=1}^{C}(i\exp(isv_{c2})y_{c1})}{\sum_{c=1}^{C}(\exp(isv_{c2}))}ds. \quad (13)$$

Here again, because of the polynomial-trigonometric choice of functions, the Fourier transforms take on particularly simple forms (see the appendix), and the problem reduces in the same manner as above. In fact, all of the estimates of elements of the matrices in the normal equations can be expressed as functions of the characteristic function and its derivatives evaluated at a finite set of points.

Given estimates of the moments above, we can form estimates of the matrices $\mathbb{E}(f(z)f(z)')$ and $\mathbb{E}(f(z)y)$ and estimate the parameter of interest by

$$\hat{\beta} = \mathbb{E}(\widehat{f(z)f(z)})'^{-1}\mathbb{E}(\widehat{f(z)y}), \quad (14)$$

where the elements of $\mathbb{E}(\widehat{f(z)f(z)})'$ are given by (12) and of $\mathbb{E}(\widehat{f(z)y})$ by (13). Each element of these matrices is a function of the estimated characteristic function and its derivatives evaluated at a finite set of points, that is, $\left\{\frac{\partial\hat{\psi}^k(s_0)}{\partial s^k}\right\}_{k=0}^{4}$ for $s_0 \in \{-C_{K,M}, -1, \ldots, 1, C_{K,M}\}$.[7] Further, it is possible to express these derivatives themselves as functions of $(\hat{\psi}(s_0), \hat{\phi}(s_0), \hat{\phi}'(s_0), \ldots, \hat{\phi}'''(s_0))$ for $s_0 \in \{-2, -1, \ldots, 1, 2\}$. Therefore, in order to characterize the limiting distribution of $\hat{\beta}$ above, it is enough to characterize the limiting distribution of these estimated quantities. The details for the asymptotic theory are relegated to the appendix, and the conclusion is

$$\sqrt{C}\ (\hat{\beta} - \beta) \Rightarrow N(0, V).$$

Although $\hat{\beta}$ converges at the regular rate and has a limiting normal distribution, it is difficult to construct a consistent estimator of the variance matrix $V$ because we do not have an explicit formula for it. One reason is that $\hat{\beta}$ is a complicated multistep estimator that depends on the preliminary steps in a complicated way. In addition, the limiting distribution of some of the preliminary estimates themselves involves variances that are hard to estimate (these distributions are functionals of gaussian processes). We will there-

[7] As the notation suggests, the constant $C_{K,M}$ depends on $K$ and $M$. For instance, with $K = 2$ and $M = 1$, $C_{K,M} = 2$ while for $K = 1$ and $M = 2$, $C_{K,M} = 4$.

TABLE 2.—MONTE CARLO RESULTS
(500 SIMULATIONS, C = 300)

| Signal / Noise | .1 | | .5 | | .9 | |
|---|---|---|---|---|---|---|
| Estimator | Deaton | HNIP | Deaton | HNIP | Deaton | HNIP |
| Mean ($\beta = 5$) | 4.974557 | 5.000918 | 4.974139 | 5.000687 | 4.975247 | 5.002013 |
| s.d. | .054903 | .0555145 | .051808 | .0525022 | .0533261 | .0537687 |

fore use the bootstrap. Although proving that the bootstrap works in this model is beyond the scope of this paper, the smoothness of the estimated parameters in the underlying distribution and the ($\sqrt{C}$) normality result suggest that the bootstrapped standard errors should be consistent.[8]

## V. Monte Carlo Results

The estimators themselves are fairly straightforward. The Monte Carlo (MC) illustrations serve to illustrate two points: the first is to offer a set of controlled comparisons between the estimators derived in this paper and the estimator proposed in Deaton (1988) in the linear context, and the second is to assess the possible effects of misspecification.

### A. *MC for First Specification*

The results of the MC simulations for the comparisons between the alternative estimators for specification 1 are given in tables 1 and 2. Both estimators have comparable mean squared errors (MSE) in the linear case, which we would expect to be the case since both are essentially correcting the pooled OLS estimates for the plim bias terms that are present due to measurement error. As robustness checks, varying the correlation between measurement errors or the signal-to-noise ratio of the mismeasured prices does not affect either estimator asymmetrically, and both perform well in these situations.

When these estimators are applied to data where the regression function is quadratic in prices, both are inconsistent, and there is no discernible pattern to be obtained from varying the correlation or signal-to-noise ratios on the distribution of the estimators. The distribution of the MC estimators is wildly off from the truth in both cases. The correctly specified estimator, which takes into account the quadratic form of the regression function, is well behaved, as we would expect from the previous section.

Both estimators stem from the same spirit of correcting OLS estimates for measurement error, and one might perhaps be able to extend directly the first method to account for nonlinearities in the regression function as well.

### B. *MC for Third Specification*

The primary purpose of the MC results for the third specification was to provide a check of the $\sqrt{C}$ convergence result (proved in the appendix) for the parameter coefficients. The true model is given by

$$y = 1 + 2z + 4z^2 + 3\sin(z) + 2\cos(z) + \varepsilon,$$

where $\varepsilon$ is standard normal and the $z$ are unobserved. Instead, we observe

$$v_j = z + \eta_j$$

for $j = 1, \ldots, H$ where $\eta$ has the Laplace distribution (and is independent of the $\varepsilon$ although that is not required). I also report results for the case where $\eta$ has the normal distribution. The MSE are presented in table 3. Encouragingly, MSE approximately halves when I double the sample size. However, in some of the cases, the MSE falls by much more than would be justified by a central limit theorem argument.

## VI. Empirical Results

The theory sketched above is a simplification of the data at hand in several respects. The first is that I do not observe positive quantities (and, hence, any unit values) for a subset of households, and therefore there is a question of how exactly to include such households in the analysis. In the first specification, since I work with a log-log specification,

[8] Note that strictly speaking this is not enough. See Abadie and Imbens (2008) or counterexample 1 in Bickel and Freedman (1981) for instances of $\sqrt{n}$ consistent and asymptotically normal estimators for which the bootstrap is invalid (at least without further assumptions).

TABLE 3.—MEAN SQUARE ERRORS FOR THE MONTE CARLO SIMULATIONS
(300 SIMULATIONS)

| Laplace Errors | | | Normal Errors | | |
|---|---|---|---|---|---|
| N = 1,000 | N = 2,000 | $MSE_{2000}/MSE_{1000}$ | N = 1,000 | N = 2,000 | $MSE_{2000}/MSE_{1000}$ |
| 292.67 | 117.7 | .402 | 131.58 | 55.445 | .42138 |
| 127.26 | 12.137 | .095 | 43.515 | 17.395 | .39975 |
| 35.303 | 14.108 | .399 | 19.6973 | 6.9773 | .3584 |
| 270.52 | 25.638 | .094 | 92.082 | 36.084 | .39186 |
| 334.46 | 137.58 | .411 | 141.68 | 63.561 | .44863 |

TABLE 4.—*F*-STATISTICS FROM A REGRESSION OF THE LOG OF UNIT VALUES

| RHS Variables | Region | Subround | Region × Subround | Village Dummies |
|---|---|---|---|---|
| *F*-statistics | 49.83 | 23.65 | 18.96 | 8.6 |

I must perforce drop any households with no recorded quantities. In the second and third specifications, I include households purchasing no wheat and set their unit value equal to the average unit value for the village they live in. This is a pragmatic imputation with no compelling justification. Second, for this analysis, I am ignoring cross-price effects, although a more complete analysis should include those as well. Third, the data are sampled using a two-stage stratified design, and I ignore weighting issues and treat the sample as if it were i.i.d. across clusters. Fourth, there are households that do not purchase any wheat but consume home-grown quantities, and one must take a position on how to treat such observations. Following the literature (Deaton, 1988), I have used the National Sample Survey (NSS) imputed harvest price of domestic production and calculated the unit value for such households from this imputation.

I work with a sample of 4,440 households from rural Maharashtra, of which 82% had consumed either market-bought or home-grown wheat in the past thirty days. The 3,641 households that consumed some wheat form the sample for my first specification. Of the remaining households, 759 lived in villages where at least one other household had consumed some wheat; these are added to the sample for my second and third specifications. The remaining 40 households lived in villages where no one consumed any wheat and are excluded from the analysis. Average total monthly consumption was Rs1,134 per household, and on average 3.4% of total household expenditure was spent on wheat for households that purchased a positive quantity. The unit values show that households paid an average of Rs5.79 per kilo for wheat. I have not as yet found an independent source for assessing whether these unit values are in fact close to the prevailing prices in rural Maharashtra during the survey, but the assumption for the paper is that these unit values do provide a measure of price, albeit a noisy one.

Table 4 highlights the spatial and temporal variation in unit values. Each column reports the *F*-statistic from a regression of the (log of the) unit values on a set of regressors. The regressors in the first column are the 5 NSS region dummies, and in the second column dummies for the four quarters (July–September, October–December, January–March, April–June). The *F*-statistics are all significant at conventional levels and provide some evidence of considerable price variation across space and time in our data. Finally, the last column reports the *F*-statistic from a regression onto a set of village dummies. Since our identification assumption is one of identifying the demand function through between-village variation in price, it is encouraging to see that the dummies together are jointly significant in the regression. Finally, table 5 presents the coefficients from a regression on region and subround dummies.

TABLE 6.—RESULTS FROM FIRST-STAGE REGRESSIONS

| depvar | ln(expend) | ln(hhsize) | Hindu | Schedule Caste |
|---|---|---|---|---|
| ln(unit value) | .073 (.008) | −.044 (.01) | −.010 (.04) | −.003 (.014) |
| ln(quantity) ($n$ = 3,641) | .354 (.04) | .257 (.05) | −.03 (.17) | .025 (.067) |
| Budget share ($n$ = 4,400) | −.013 (.001) | .012 (.002) | .0001 (.008) | −.006 (.002) |

Note: Regressors include log(total expenditure), log(household size), and sociodemographic characteristics.

Table 6 shows the results from the first-stage estimation for the various specifications. For each specification, the first stage estimates the parameters $\{\gamma_1, \gamma_2\}$ using a standard fixed-effect model where $x = \{$age-sex variables, log(total expenditure), religion, caste status, household size, labor type$\}$. As might be expected, household size and total expenditure turn out to be important regressors. The pattern of results is consistent with Engel's law in the sense that although richer households purchase more wheat (the second row of coefficients), the share of wheat in the total budget declines as households get richer (the second set of coefficients). Of course, the poorest households do not purchase any wheat, but the result is robust to their inclusion as well (the final row of coefficients). This does suggest, however, that the relationship between budget share of wheat and total expenditure is not necessarily monotonic.

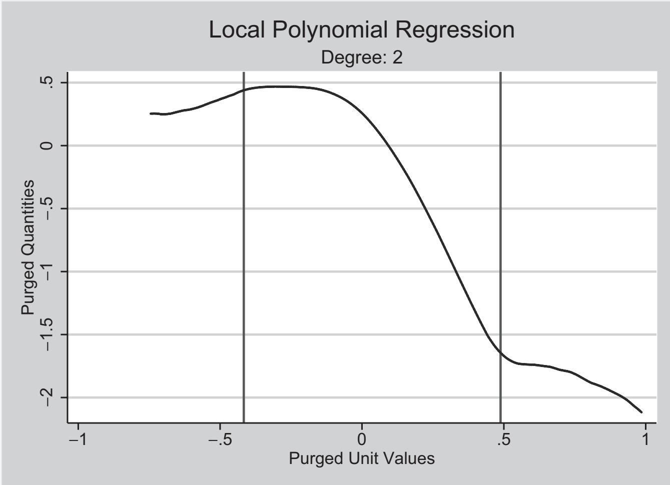
Figure 1 depicts the relationship between purged (log) quantities and unit values using local polynomial kernel regression. The function is not clearly linear, and the object of the first specification is to evaluate the implications of approximating the true regression function (recall that the displayed regression line is not of interest since it is the regression on mismeasured prices) solely with a polynomial specification.

Following the strategy outlined in the identification section above, I next construct sample moments using the purged unit values and quantities to obtain estimates for the unobserved moments. For instance, to estimate the mean of the unobserved price, I take the average of the observed unit values across all households. Higher moments are estimated by taking all possible combinations of households that are consistent for the moment of interest and taking their

TABLE 5.—REGIONAL AND SEASONAL DIFFERENCES IN UNIT VALUES

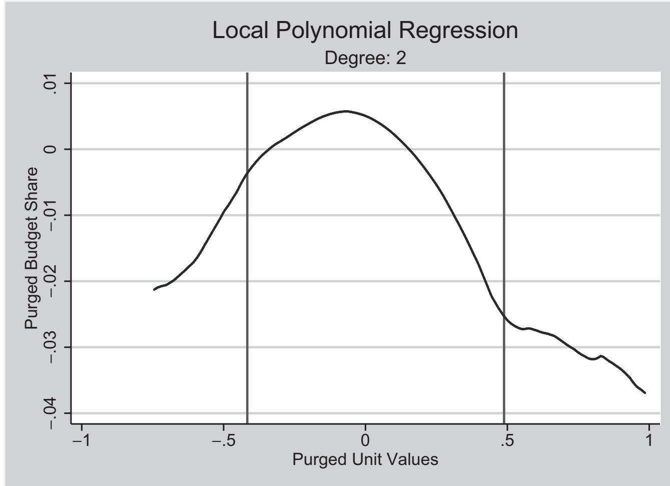
| depvar | R2 | R3 | R4 | R4 | R5 | SR2 | SR3 | SR4 |
|---|---|---|---|---|---|---|---|---|
| ln(unit value) | −.017 | −.19 | −.159 | −.15 | −.09 | .031 | .087 | .097 |
| s.e. | (.015) | (.018) | (.016) | (.016) | (.02) | (.012) | (.013) | (.013) |

FIGURE 1.—LOCALLY QUADRATIC REGRESSION OF PURGED QUANTITIES ON PURGED UNIT VALUES USING A GAUSSIAN KERNEL WITH A BANDWIDTH OF .2




Note: The vertical lines represent the 5th and 95th percentiles, respectively.

FIGURE 2.—LOCALLY QUADRATIC REGRESSION OF PURGED BUDGET SHARE ON PURGED UNIT VALUES USING A GAUSSIAN KERNEL WITH A BANDWIDTH OF .2




Note: The vertical lines represent the 5th and 95th percentiles, respectively.

weighted average (where the weights are given by the relative fraction of observations for a particular sample moment). For instance in calculating $E[z^2]$, we use all $\binom{H}{2}$ possible choices of $i \neq j$ and compute $E_{F_c} v_i v_j$. Since some villages may not have $H$ households, the product $v_{ci}v_{cj}$ may not exist for all clusters, and the average is, naturally, over all nonmissing values.

The first specification does not use households that report zero quantities. This leaves a total of 3,641 households spread across 440 villages. I estimate a series of possible models starting with the linear model. In this case, I can compare my estimates with estimates obtained from the method outlined in Deaton (1997). The results are displayed in table 7. The point estimates for the elasticity are not too dissimilar in both cases, with the Deaton method slightly larger than the linear specification of our method. The most important difference is in the bootstrapped standard errors, which are much larger under the newer method, so that the corresponding confidence intervals are also much larger (I report the 2.5th and 97.5th percentile based on 1,000 bootstrapped resamples at the cluster level for the second stage of estimation).

For models with higher-order polynomials,

$$\frac{\partial}{\partial \ln p}\mathbb{E}[q|\ln p] \qquad (15)$$

is a function of the price, and one has to take a position on where to evaluate these derivatives. I calculate the sample version of the average derivative,

$$\mathbb{E}\left(\frac{\partial}{\partial \ln p}\mathbb{E}[q|\ln p]\right),$$

and report results for the derivative evaluated at the 25th and 75th percentile of the purged unit value distribution. I estimate polynomials of order up to five, although the magnitude of the bootstrapped standard errors for specifications higher than the quadratic calls into question any conclusions that one could derive from the higher-order polynomial models. Specifically, in these models, we cannot reject the null that the observed elasticity (evaluated at the candidate values of the price variable) is not significantly different from 0.

The specifications that are nonlinear in price seem to provide better approximations to the regression function graphed above, in that the elasticities are lower at the lower end of the distribution and higher at the higher end of the distribution. Note that the plot above is the regression function for quantities conditional on unit values (mismeasured prices), so its usefulness as a guide for the regression function of quantities on the true prices may be limited.

TABLE 7.—AVERAGE PRICE ELASTICITIES FOR DIFFERENT MODELS UNDER SPECIFICATION 1

| Model | 25th Percentile | Average Derivative | 75th Percentile |
|---|---|---|---|
| Deaton | — | −1.438 (−1.84, −1.35) | — |
| Linear | −1.575 (−2.13, −1.22) | −1.575 (−2.13, −1.22) | −1.575 (−2.13, −1.22) |
| Quadratic | −.5607 (−1.30, .32) | −1.628 (−2.21, −1.26) | −2.393 (−3.59, −1.84) |
| Cubic | −.6626 (−7.82, 3.52) | −1.723 (−4.24, .52) | −2.536 (−10.47, 4.86) |
| Quartic | −.7785 (−6.86, 5.56) | −1.769 (−5.01, .45) | −2.555 (−13.01, 6.56) |
| Quintic | −.4848 (−7.39, 4.20) | −1.531 (−3.16, −.23) | −2.141 (−8.71, 4.03) |

Note: All confidence intervals are based on 1,000 bootstrapped replications.

TABLE 8.—ELASTICITIES ESTIMATED FROM SPECIFICATION 2

| Model | 25th Percentile | Average Derivative | 75th Percentile |
|---|---|---|---|
| Deaton | — | −1.465 (−2.15, −1.13) | — |
| Linear | −1.463 (−1.95, −.89) | −1.463 (−1.95, −.88) | −1.463 (−1.95, −.89) |
| Quadratic | −.965 (−1.58, −.25) | −1.424 (−1.89, −.89) | −1.765 (−2.35, −1.21) |
| Cubic | −.988 (−1.84, .175) | −1.445 (−2.20, −.69) | −1.803 (−3.30, −.85) |
| Quartic | 11.94 (−13.19, 6.74) | −.089 (−2.43, .68) | −10.10 (−6.32, 4.66) |
| Quintic | −1.28 (−9.31, 8.78) | −1.50 (−2.44, −.48) | −1.532 (−6.48, 4.65) |

Note: Elasticities are evaluated at the average budget share for three different points in the unit value distribution. All confidence intervals based on 1,000 bootstrapped replications.

Figure 2 presents the locally weighted quadratic regression of purged budget shares on purged unit values. The lines correspond to the 5th and 95th percentiles of the distribution for purged unit values. Again, the relationship looks very similar to the relationship in figure 1. Specifications 2 and 3 attempt to model the relationship between these purged budget shares and the true (unobserved) prices using a polynomial model and a Fourier flexible form approach.

The results from the polynomial models are presented in table 8. The elasticity estimated in these models (apart from the linear model) is a function of the budget share and the price. I fix the budget share equal to the average budget share for the sample and evaluate the elasticity for two different values of the price variable (the 25th and the 75th percentile of the purged unit values) as well as the average derivative calculation. As might be suggested by figure 2, the elasticities are lower at the 25th percentile of the purged unit values and higher when evaluated at the 75th percentile of the unit value distribution. The average elasticity (evaluated at the mean of the purged unit value distribution) lies between these two. All the models in table 8 exhibit this pattern. However, note that the standard errors become extremely large after the cubic specification, and indeed for all but one of the quartic and quintic specifications, we cannot reject the null of zero elasticity. The large standard errors for the higher degree specifications reflect the imprecision with which the higher moments are estimated (as reflected in their bootstrapped standard errors which are not reported here). Finally, the results from the last specification (setting $K = 2$ and $M = 2$) are displayed in table 9. The point estimates are broadly similar to the ones from the previous specification, and here again the elasticities are lower at the lower end of the price distribution relative to the higher end.

Looking across all the specifications, we see that the results across various nonlinear models are broadly consistent with each other. While there is no compelling reason to believe any of the regression functions is correctly specified, the relative stability of the elasticity estimates across specifications is encouraging and suggests that misspecification is not driving the results. Overall, the elasticities derived from the budget share specification are more precisely estimated (as expected) relative to the log-quantity specification. While no specification clearly dominates the others, the quadratic choice for all three specifications is relatively straightforward to implement and provides broadly consistent and reasonably precise estimates of the price elasticities.

For comparison, we provide results from estimating a specification without accounting for measurement error. We implement a quadratic version of the first specification. The results are presented in table 10 and should be compared against the third row of results in table 7. Both the point estimates and the standard errors for the naive estimator differ from those of the estimator that accounts for measurement error. Two differences emerge. First, the estimates without correction suggest much less sampling uncertainty than the results in table 7, which is reasonable since the former do not account for measurement error in any way. If the proposed model in the paper is correct, the naive estimator seems to significantly underestimate sampling uncertainty. Second, the point estimates are quite different as well and suggest much smaller differences in the elasticities across the price distribution. The differences in this case seem to stem from the fact that the estimated higher moments of the purged price distribution are quite different for the two models. If specification 1 is correct, then the naive estimator's computation of these moments will be inconsistent since it does not purge the measurement error.

To conclude, the most substantive point, and one that comes across in all the specifications, is the evidence that price elasticity does vary across the price distribution. Estimated price elasticities are generally lower (in absolute magnitude) in the lower tails of the distribution and are generally larger in the upper tails, something that a linear specification of the demand equation would miss.

TABLE 9.—ELASTICITIES ESTIMATED FROM SPECIFICATION 3

| 25th Percentile | Average Derivative | 75th Percentile |
|---|---|---|
| −1.3724 (−2.09, −.43) | −.8806 (−1.41, −.28) | −1.5943 (−3.40, −.88) |

Note: Elasticities are evaluated at the average budget share for three different points in the unit value distribution. All confidence intervals based on 200 bootstrapped replications.

TABLE 10.—ELASTICITIES ESTIMATED WITHOUT ACCOUNTING FOR MEASUREMENT ERROR

| 25th Percentile | Average Derivative | 75th Percentile |
|---|---|---|
| −1.4732<br>(−1.48, −1.46) | −1.7614<br>(−1.768, −1.75) | −1.9492<br>(−1.95, −1.942) |

Note: Elasticities are evaluated at the average budget share for three different points in the unit value distribution.

## VII. Conclusion

This paper estimates price elasticities by modeling flexibly the regression function for quantities demanded (and budget shares) as a function of price while at the same time accounting for the effect of measurement error in prices that is fundamentally nonclassical in nature. In such a setting, standard instrumental variable techniques are no longer available. We adapt the strategy suggested by Hausmann et al. (1991) and Schennach (2004) to account for the nonclassical nature of the measurement error and exploit the clustered nature of the data to achieve identification of the price elasticity. We next propose an estimator based on the identification strategy and discuss its large sample properties. Next, we discuss its performance in MC simulations and finally apply it to estimate a price elasticity using data from rural India. Kernel regressions from the data suggest potential evidence of nonlinearity in the true conditional expectation, and we estimate a series of alternative specifications to account for this. The estimator confirms the initial presumption, and for all the specifications considered, the estimated price elasticities are somewhat lower at the bottom of the price distribution and higher at the top.

Two potential further directions for research are to include and estimate cross-price effects and also to consider estimation in the case where the number of polynomial (or trigonometric) terms increases with the sample size.

## REFERENCES


Abadie, Alberto, and Guido W. Imbens, "On the Failure of the Bootstrap for Matching Estimators," *Econometrica* 76 (2008), 1537–1557.

Bickel, Peter J., and David A. Freedman, "Some Asymptotic Theory for the Bootstrap," *Annals of Statistics* 9:6 (1981), 1196–1217.

Deaton, Angus, "Estimation of Own and Cross Price Elasticities from Household Survey Data," *Journal of Econometrics* 36 (1987), 7–30.

——— "Quality, Quantity and the Spatial Variation of Price," *American Economic Review* 78:3 (1988), 418–430.

——— *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy* (Baltimore, MD: Johns Hopkins University Press, 1997).

Deaton, Angus, and John Muellbauer, "An Almost Ideal Demand System," *American Economic Review* 70:3 (1980), 312–326.

Deaton, A., and Serena Ng, "Parametric and Non-Parametric Approaches to Price and Tax Policy Reform," *Journal of the American Statistical Association* 93 (1998), 900–909.

Eubank, Randy L., and Paul Speckman, "Curve Fitting by Polynomial-Trigonometric Regression," *Biometrika* 77 (1990), 1–9.

Feuerverger, A., and R. A. Mureika, "The Empirical Characteristic Function and Its Applications," *Annals of Statistics* 5 (1977), 88–97.

Gallant, A. Ronald, "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form," *Journal of Econometrics* 15 (1981), 211–245.

Hardle, Wolfgang K., and Thomas M. Stoker, "Investigating Smooth Multiple Regression by the Method of Average Derivatives," *Journal of the American Statistical Association* 84:408 (1989), 986–995.

Hausmann, Jerry A., Whitney K. Newey, Hidehiko Ichimura, and James L. Powell, "Identification and Estimation of Polynomial Errors-in-Variables Models," *Journal of Econometrics* 50:3 (1991), 273–296.

Kennan, John, "Simultaneous Equations Bias in Disaggregated Econometric Models," *Review of Economic Studies* 56 (1989), 151–156.

Lighthill, Michael J., *Introduction to Fourier Analysis and Generalized Functions* (Cambridge: Cambridge University Press, 1996).

Newey, Whitney K., and Daniel L. McFadden, "Large Sample Estimation and Hypothesis Testing" (pp. 2111–2245), in R. Engle and D. McFadden (Eds.), *Handbook of Econometrics* (Amsterdam: Elsevier Science, 1994).

Schennach, Susanne M., "Estimation of Non Linear Models with Measurement Error," *Econometrica* 72:1 (2004), 33–75.

van der Vaart, Aad W., *Asymptotic Statistics* (Cambridge: Cambridge University Press, 1998).

van der Vaart, Aad W., and Jon A. Wellner, *Weak Convergence and Empirical Processes with Applications to Statistics* (New York: Springer, 1996).


## APPENDIX

### A1 Identification

**Proof of lemma 1.** Consider the directly identified quantity (for $h_a \neq h_b$ for all $(a, b) \in \{1, \ldots, j\}$,

$$\begin{aligned}\mathbb{E}(v_{ch_1} \cdots v_{ch_j}) &= \mathbb{E}((z_c + \eta_{ch_1}) \cdots (z_c + \eta_{ch_j})) \\ &= \mathbb{E}\Bigg( z_c^j + z_c^{j-1} \sum_{h_1=1}^{j} \eta_{ch_1} + \cdots \\ &\quad + z_c \sum_{h_1=1\ldots}^{2} \sum_{h_{j-1}=h_{j-2}+1}^{j} (\eta_{ch_1} \cdots \eta_{ch_{j-1}}) + (\eta_{ch_1} \cdots \eta_{ch_j}) \Bigg) \\ &= \mathbb{E}(z_c^j),\end{aligned}$$

where the last equality follows from the independence between $z_c$ and $\eta_{ch}$ and the fact that $\mathbb{E}(\eta_{ch}) = 0$ (assumption 2). Note that for this result alone, mean independence rather than independence will suffice. Further, we can consider all $\binom{H}{j}$ distinct subsets of size $j$ from the set $\{v_1, \ldots, v_H\}$, each of whose expectation will be equal to that of $\mathbb{E}(z_c^j)$ by the same argument as above so that

$$\mathbb{E}(z_c^j) = \mathbb{E}\left[ \frac{1}{\binom{H}{j}} \sum_{h_1=1}^{H-j+1} \sum_{h_2=h_1+1}^{H-j+2} \cdots \sum_{h_j=h_{j-1}}^{H} v_{ch_1} v_{ch_2} \cdots v_{ch_j} \right].$$

**Proof of lemma 2.** As in lemma 1, consider $j + 1$ distinct households and the directly identified quantity $\mathbb{E}(v_{ch_1}, \ldots, v_{ch_j} y_{ch_{j+1}})$. Rewriting this, we obtain

$$\begin{aligned}&\mathbb{E}\Bigg( y_{ch_{j+1}} \Bigg( z_c^j + z_c^{j-1} \sum_{h_1=1}^{j} \eta_{ch_1} + \cdots \\ &\quad + z_c \sum_{h_1=1\ldots}^{2} \sum_{h_{j-1}=h_{j-2}+1}^{j} (\eta_{ch_1} \cdots \eta_{ch_{j-1}}) + (\eta_{ch_1} \cdots \eta_{ch_j}) \Bigg) \Bigg) \\ &\quad = \mathbb{E}(z_c^j y_{ch_{j+1}}) + \mathbb{E}\left( y_{ch_j} z_c^{j-1} \sum_{h_1=1}^{j} \eta_{ch_1} \right) \\ &\quad + \mathbb{E}\left( y_{ch_j} \sum_{h_1=1\ldots}^{2} \sum_{h_{j-1}=h_{j-2}+1}^{j} (\eta_{ch_1} \cdots \eta_{ch_{j-1}}) \right) \\ &\quad + \mathbb{E}(y_{ch_j}(\eta_{ch_1} \cdots \eta_{ch_j})) \\ &\quad = \mathbb{E}(z_c^j y_{ch_{j+1}}),\end{aligned}$$

where the last equality follows from the fact that all the terms are of the form

$$\mathbb{E}(y_{ch_{j+1}} g(z_c) r(\eta_{ch_1}, \ldots, \eta_{ch_j}))$$

for appropriately specified functions $g(\cdot)$ and $r(\cdot)$. This expectation is identically 0 since

$$\mathbb{E}(y_{ch_{j+1}}g(z_c)r(\eta_{ch_1},\ldots,\eta_{vh_j})) = \mathbb{E}((f(z_c)+f_c + \varepsilon_{ch_{j+1}})g(z_c)r(\eta_{ch_1},\ldots,\eta_{vh_j}))$$
$$= \mathbb{E}(f(z_c)g(z_c)r(\eta_{ch_1},\ldots,\eta_{vh_j})) + \mathbb{E}(f_c g(z_c)r(\eta_{ch_1},\ldots,\eta_{ch_j})) + \mathbb{E}(\varepsilon_{ch_{j+1}}r(\eta_{ch_1},\ldots,\eta_{ch_j}))$$
$$= 0,$$

where the first two terms are 0 since $(\eta_{ch_s})_{s=1}^{j}$ is independent of $(z_c, f_c)$ and $\mathbb{E}(r(\eta_{ch_1},\ldots,\eta_{ch_j})) = 0$ and the last term is 0 since $\varepsilon_{ch_{j+1}}$ is independent of $(\eta_{ch_s})_{s=1}^{j}$ (all of which follow from assumption 4).

**Proof of lemma 3.** Under assumption 1, the parameters $\gamma = (\gamma_1, \gamma_2)$ are identified, and therefore $\{y_{ch}, v_{ch}\}$ are identified. Next, under the assumptions stated in the lemma, $\beta$ can be obtained as

$$\beta = \mathbb{E}(\tilde{z}_{cK}\tilde{z}'_{cK})^{-1}\mathbb{E}\left(\frac{1}{H}\sum_{h=1}^{H}\tilde{z}_{cK}y_{ch}\right),$$

and each element on the right-hand side is identified by the previous two lemmas.

### A2 Large Sample Theory

The asymptotic theory for the first two specifications is straightforward and can be derived using standard results from two-step estimation theory (e.g., section 6 of Newey & McFadden, 1994). In the first step, the elements of $\gamma$ can be consistently estimated (up to a constant) using the sample version of the moment conditions[9]

$$\mathbb{E}\left(\underbrace{\begin{matrix}\sum_{h=1}^{H}(\ddot{w}_{ch}-\ddot{x}'_{ch}\gamma_1)\ddot{x}_{ch}\\ \sum_{h=1}^{H}(\ddot{u}_{ch}-\ddot{x}'_{ch}\gamma_1)\ddot{x}_{ch}\end{matrix}}_{m((w_{ch},x_{ch},u_{ch})_{h=1}^{H},\gamma)}\right) = 0.$$

Under the assumptions for lemma 3 and that the second moments of the bivariate vector of the composite error terms $(\ddot{\varepsilon}_{ch}, \ddot{\eta}_{ch})$ exist,[10] standard arguments yield

$$\sqrt{C}\,(\hat{\gamma}-\gamma) \Rightarrow \mathcal{N}(0, V_\gamma),$$

where

$$V_\gamma = \left(\mathbf{I}_2 \otimes \sum_{h=1}^{H}\mathbb{E}(\ddot{x}_h\ddot{x}'_h)\right)^{-1} M \left(\mathbf{I}_2 \otimes \sum_{h=1}^{H}\mathbb{E}(\ddot{x}_h\ddot{x}'_h)\right)^{-1},$$

where $M$ is the second moment matrix of the moment conditions.

In the second step, we estimate the elements of the parameter $\theta$—defined in (9) by (10) and (11). Rewriting these to reflect the dependence on the first-stage estimates $\gamma$ and writing them out in a common format,

$$\frac{1}{C}\sum_{c=1}^{C}\left[\frac{1}{\binom{H}{j}}\sum_{h_1=1}^{H-j+1}\sum_{h_2=h_1+1}^{H-j+2}\cdots\sum_{h_j=h_{j-1}}^{H}(u_{ch_1}-x'_{ch_1}\hat{\gamma}_2)\cdots(u_{ch_j}-x'_{ch_j}\hat{\gamma}_2)\right] - \hat{\zeta}_j = 0$$

$$\frac{1}{C}\sum_{c=1}^{C}\left[\frac{1}{\binom{H}{j}}\sum_{h_1=1}^{H-j+1}\sum_{h_2=h_1+1}^{H-j+2}\cdots\sum_{h_j=h_{j-1}}^{H}(u_{ch_1}-x'_{ch_1}\hat{\gamma}_2)\cdots(u_{ch_{j-1}}-x'_{ch_j}\hat{\gamma}_2)\times(w_{ch_j}-x'_{ch_j}\hat{\gamma}_1)\right] - \hat{\xi}_j = 0,$$

where for the first specification, we replace $w_{ch}$ with $q_{ch}$. We collect these $2K + K$ moment conditions and, denoting them collectively by the moment, vector $g((w_{ch}, x_{ch}, u_{ch})_{h=1}^{H}, \hat{\theta}, \hat{\gamma})$. Note that the moment conditions are in fact linear in the second-stage parameters. We can derive the asymptotic normality of the parameter vector under the following additional assumption:

**Assumption 5.** The error terms $(\ddot{\varepsilon}_{ch}, \ddot{\eta}_{ch})$ have finite second moments. The true parameter $(\theta, \gamma)$ lies in the interior of the Euclidean parameter space. The moment function $g$ satisfies $\mathbb{E}$, $(\|g((w_{ch}, x_{ch}, u_{ch})_{h=1}^{H}, \theta, \gamma)\|^2)$ is finite and its first derivative (with respect to $(\theta, \gamma)$) is uniformly bounded over a neighborhood of the truth in expectation. The expressions in the first three displays after equation (16) exist, and the matrix $V_\theta$ is nonsingular.

We can now verify the conditions of theorem 6.1 of Newey and McFadden (1994) to conclude that

$$\sqrt{C}\,(\hat{\theta}-\theta) \Rightarrow \mathcal{N}(0, V_\theta),$$

where

$$V_\theta = \mathbb{E}(rr'),$$

where

$$r = g((w_{ch}, x_{ch}, u_{ch})_{h=1}^{H}, \theta, \gamma) + G_\gamma m((w_{ch}, x_{ch}, u_{ch})_{h=1}^{H}, \gamma)$$

and

$$G_\gamma = \mathbb{E}\left(\frac{\partial g((w_{ch}, x_{ch}, u_{ch})_{h=1}^{H}, \theta, \gamma)}{\partial\gamma}\right).$$

### A3 Asymptotic Distribution for the Third Specification

Under the assumptions outlined above, we derive the limiting distribution for the case where $K = 2$ and $M = 2$. As mentioned in the text, in order to characterize the limiting distribution of the estimator $\hat{\beta}$, it suffices to characterize the limiting distribution of the vector

$$(\hat{\psi}(s_0), \hat{\phi}(s_0), \hat{\phi}'(s_0), \ldots, \hat{\phi}'''(s_0)) \quad (16)$$

for $s_0 \in \{-4, -1, \ldots, 1, 4\}$. Apart from the first term, each term can be estimated by its sample average; this is because we can calculate the derivatives of $\frac{\mathbb{E}[v_1\exp(itv_2)]}{\mathbb{E}[\exp(itv_2)]}$ analytically, invoking the appropriate DCT to allow interchange of integration and differentiation. Each derivative can then be consistently estimated by the analogy principle, and the analysis of the limiting distribution is considerably simplified by the fact that both the numerator and denominator are Lipschitz continuous in $t$. However, there is an added complication due to the fact that we do not observe $v_s$ directly. Consider, for instance, the estimator of the parameter

$$\phi(t) = \frac{\mathbb{E}[v_1\exp(itv_2)]}{\mathbb{E}[\exp(itv_2)]},$$

which we estimate by

[9] Note that no claim of optimality for the choice of moment conditions is being made. It is possible to construct more efficient estimators of the $\gamma$'s given the assumptions above; however, this is not pursued here. The double dots denote that the variations are in deviations from their cluster means.

[10] As usual, the notation $f(z_c)$ denotes the flexible specifications of the price variable in the demand equation.

$$\hat{\phi}(t) = \frac{\sum_{c=1}^{C} (\hat{v}_{c1} \exp(it\hat{v}_{c2}))}{\sum_{c=1}^{C} \exp(it\hat{v}_{c2})} \equiv \frac{\hat{m}_1(t)}{\hat{m}_2(t)}.$$

In order to study the asymptotic properties of this estimator, we first introduce an easier (but infeasible) estimator:

$$\tilde{\phi}(t) = \frac{\frac{1}{C}\sum_{c=1}^{C} v_{c1} \exp(itv_{c2})}{\frac{1}{C}\sum_{c=1}^{C} \exp(itv_{c2})} \equiv \frac{\tilde{m}_1(t)}{\tilde{m}_2(t)}.$$

We will study the limiting properties of $\sqrt{C}\,(\hat{\phi}(t) - \phi(t))$ by looking at

$$\sqrt{C}\,(\hat{\phi}(t) - \phi(t)) = \sqrt{C}\,(\hat{\phi}(t) - \tilde{\phi}(t)) + \sqrt{C}\,(\tilde{\phi}(t) - \phi(t)),$$

and we will characterize the joint limiting properties of $\sqrt{C}\,(\hat{\phi}(t) - \tilde{\phi}(t))$ and $\sqrt{C}\,(\tilde{\phi}(t) - \phi(t))$.

Consider first the second term $\sqrt{C}\,(\tilde{\phi}(t) - \phi(t))$, and consider first the convergence properties of the numerator $\tilde{m}_1(t)$ and the denominator $\tilde{m}_2(t)$ individually (we will then apply the delta method to obtain the limiting properties of $\tilde{\phi}$). Standard arguments can be used to show that for a fixed $t$, the normalized sample averages $\sqrt{C}\,(\tilde{m}_1(t) - m_1(t), \tilde{m}_2(t) - m_2(t))' \equiv (\sqrt{C}\,(\tilde{m}(t) - m(t)))$ converge in distribution to a bivariate normal distribution. In fact, they can be used to show that for any finite set of $\{t_1, \ldots, t_k\}$, the corresponding vector $\sqrt{C}\,((\tilde{m}(t_1), \ldots, \tilde{m}(t_k))' - (m(t_1), \ldots, m(t_k)))$ converges to a multivariate normal distribution.[11] For any fixed $t$, then, the asymptotic normality of $\sqrt{C}\,(\tilde{\phi}(t) - \phi(t))$ follows from an application of the delta method and the additional assumption that $m_2(t) \neq 0$.

Consider the first term now. We can rewrite the difference between $\hat{m}_2(t)$ and $\tilde{m}_2(t)$ as

$$\frac{1}{\sqrt{C}} \sum_{c=1}^{C} \exp(it(u_{ch_2} - x'_{ch}\hat{\gamma}_2)) - \exp(it(u_{ch_2} - x'_{ch}\gamma_2)),$$

and a standard mean-value argument and the results for the first-stage estimation in (14) will yield that $\sqrt{C}\,(\hat{m}_2(t) - \tilde{m}_2(t))$ will converge weakly to a normal distribution with zero mean and variance:

$$\mathbb{E}(x_{ch} \exp(itv_{ch}))V_{\gamma}\mathbb{E}(x'_{ch} \exp(itv_{ch})).$$

Similar arguments yield the asymptotic normality for $\sqrt{C}\,(\hat{m}_1(t) - \tilde{m}_1(t))$. The delta method will then yield the asymptotic normality of $\sqrt{C}\,(\hat{\phi}(t) - \tilde{\phi}(t))$ and hence the asymptotic normality of $\sqrt{C}\,(\hat{\phi}(t) - \phi(t))$ (note that the variance of the limiting distribution will require an additional calculation of the covariance between the two limit distributions). Similar arguments can be adduced to show the convergence of the expressions $(\hat{\phi}'(s_0), \ldots, \hat{\phi}'''(s_0))$.

The large sample distribution for the estimator $\hat{\pi}(s)$ is somewhat more complicated since it depends on the value of the estimator $\hat{\phi}$ at all points $[0, s]$ rather than just a finite number of them. This implies that we need to consider not just the pointwise convergence of $\hat{\phi}$ (in $t$) but rather convergence that is uniform in $t$ in some appropriate sense.

As a first step, we will need to show that $\sqrt{C}\,(\hat{m}(\cdot) - m(\cdot))$ when viewed as stochastic processes (in $t$ for $t \in [0, s]$) converge weakly. As before, we consider the decomposition

$$\sqrt{C}\,(\hat{m}(t) - m(t)) = \sqrt{C}\,(\hat{m}(t) - \tilde{m}(t)) + \sqrt{C}\,(\tilde{m}(t) - m(t)). \quad (17)$$

We first use results from the empirical process literature to study the limiting properties of the second term. One method is to show that the function classes (suppressing the dependence on $h$)

$$\mathcal{F} = \{t \in [0, s] : f(v_c, t) = \exp\{iv_c t\}\}$$

$$\mathcal{G} = \{t \in [0, s] : f(v_{1c}, v_{2c}, t) = v_{1c} \exp\{iv_{2c}t\}\}$$

are Donsker. In particular, it is easy to show that conditions of example 19.7 of van der Vaart (1998) are satisfied since the functions are Lipschitz in $t$ with appropriate moments existing because of the identification assumptions. Therefore, we can conclude that the class $\mathcal{F}$ above is Donsker and hence converge weakly to a tight (gaussian) process $W_1$ with zero mean covariance given by

$$\text{Cov}\,(W_1(s), W_1(t)) = \mathbb{E}(\exp(i(t + s)v_c)) - \mathbb{E}(\exp(itv_c))\mathbb{E}(\exp(isv_c))$$

(see also Feuerverger & Mureika, 1977, for more details). Similarly, we can show that the class $\mathcal{G}$ is also Donsker and also converges weakly to a tight gaussian process $W_2$ with zero mean and covariance given by

$$\text{Cov}\,(W_2(s), W_2(t)) = \mathbb{E}(v_{1c}^2 \exp(i(t + s)v_{2c})) - \mathbb{E}(v_{1c} \exp(itv_{2c}))\mathbb{E}(v_{1c} \exp(isv_{2c})).$$

Since showing the Donsker property element by element implies the Donsker property for the vector of elements, we can conclude that the process $\sqrt{C}\,(\tilde{m}(\cdot) - m(\cdot))$ converges weakly to to gaussian limit process $X(\cdot)$.

Finally, we deal with the first term in the decomposition (17). Consider the term

$$\tilde{W}_{1C}(t) = \hat{m}_2(t) - \tilde{m}_2(t) = \frac{1}{\sqrt{C}} \sum_{c=1}^{C} \exp(it(u_{ch} - x'_{ch}\hat{\gamma}_2)) - \exp(it(u_{ch} - x'_{ch}\gamma_2)).$$

Using the previous results, it follows that for any finite collection $(t_s)_s$ the vector $(\tilde{W}_{1C}(t_s))_s$ converges to a normal distribution $\tilde{W}_1$ with zero mean and covariance given by

$$\text{Cov}\,(\tilde{W}_1(t_a), \tilde{W}_1(t_b)) = \mathbb{E}(\exp(it_a v_{ch})x'_{ch})V_{\gamma}\mathbb{E}(\exp(it_b v_{ch})x_{ch}).$$

In addition, the space $[0, s]$ with the usual metric is totally bounded. In order to show a weak convergence result for the process $\tilde{W}_{1C}(\cdot)$, we need only to verify that condition ii (asymptotic tightness) of theorem 18.14 of van der Vaart (1998) holds. To see this, for any given $(\varepsilon, \eta)$, partition $[0, s]$ into intervals of length no larger than $\frac{\varepsilon}{c\Phi^{-1}(1 - \eta)}$ where $\Phi^{-1}$ is the inverse of the standard normal CDF. Then, as $C \to \infty$, the probability that the maximum value of the difference $\tilde{W}_{1C}(s) - \tilde{W}_{1C}(t)$ over any partition (i.e., $(s, t)$ are in the same partition) will exceed $\varepsilon$ will be less than $\eta$. Therefore, the process is also asymptotically tight, and the weak convergence result follows. Similar arguments can be used to show that the process $\sqrt{C}\,(\hat{m}_1(\cdot) - \tilde{m}_1(\cdot))$ also converges to a gaussian process. Therefore, by the continuous mapping theorem, each element of the vector-valued process $\sqrt{C}\,(\hat{m}(\cdot) - m(\cdot))$ converges to a gaussian process, which we denote by $X(\cdot)$.

Having shown that the process $\sqrt{C}\,(\hat{m}(\cdot) - m(\cdot))$ converges to a gaussian process, we next apply the functional delta method to the function

$$g(x(\cdot)) = \int_0^s \frac{x_1(t)}{x_2(t)}\,dt.$$

This function is Frechet differentiable at $m(\cdot)$ with continuous derivative $\mathcal{L}_m(\cdot)$ given by

$$\mathcal{L}_m(h) = \int_0^s \frac{1}{m_2(t)}\,h_1(t) - \frac{m_1(t)}{(m_2(t))^2}\,h_2(t)dt.$$

Then we can apply theorem 20.8 of van der Vaart (1998) to conclude that $\sqrt{C}\,(\hat{\pi}(s) - \pi(s))$ converges to the random variable $\mathcal{L}_m(X(\cdot))$, and since $\mathcal{L}_m(\cdot)$ is a continuous linear functional applied to a tight gaussian process, $\mathcal{L}_m(X(\cdot))$ will be normally distributed (see, e.g., lemma 3.9.8 of van der Vaart & Wellner, 1996).

[11] In fact, we will show below that the process (indexed by $t \in [0, s]$, $\sqrt{C}\,(\tilde{m}(\cdot) - m(\cdot))$ converges to a gaussian process because that is needed to study the limiting distribution of $\hat{\psi}$.