

Digitized by the Internet Archive  
in 2011 with funding from  
Boston Library Consortium Member Libraries

<http://www.archive.org/details/flexiblesimulate00newe2>



DEWEY

HB31  
.M415

99-02

31

**working paper  
department  
of economics**

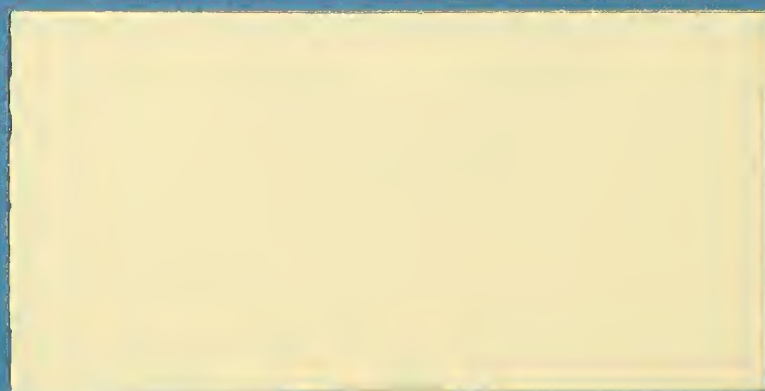
***FLEXIBLE SIMULATED MOMENT ESTIMATION OF  
NONLINEAR ERRORS-IN-VARIABLES MODELS***

**Whitney K. Newey**

**No. 99-02**

**February, 1999**

**massachusetts  
institute of  
technology  
50 memorial drive  
cambridge, mass. 02139**



**WORKING PAPER  
DEPARTMENT  
OF ECONOMICS**

***FLEXIBLE SIMULATED MOMENT ESTIMATION OF  
NONLINEAR ERRORS-IN-VARIABLES MODELS***

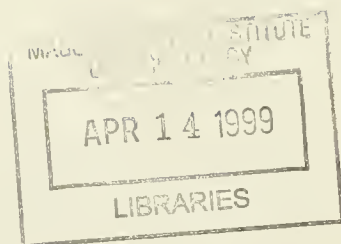
**Whitney K. Newey**

**No. 99-02**

**February, 1999**

**MASSACHUSETTS  
INSTITUTE OF  
TECHNOLOGY**

**50 MEMORIAL DRIVE  
CAMBRIDGE, MASS. 02142**





Flexible Simulated Moment Estimation of  
Nonlinear Errors-in-Variables Models<sup>1</sup>

Whitney K. Newey<sup>2</sup>

MIT

First Version, October 1989

Revised, February 1999

**Abstract**

Nonlinear regression with measurement error is important for estimation from microeconomic data. One approach to identification and estimation is a causal model, where the unobserved true variable is predicted by observable variables. This paper is about estimation of such a model using simulated moments and a flexible disturbance distribution. An estimator of the asymptotic variance is given for parametric models. Also, a semiparametric consistency result is given. The value of the estimator is demonstrated in a Monte Carlo study and an application to estimating Engle Curves.

*JEL Classification: C15, C21*

*Keywords: nonlinear regression, errors-in-variables, simulated moments.*

<sup>1</sup>This research was supported by the National Science Foundation and the Sloan Foundation. David Card and Angus Deaton provided helpful comments.

<sup>2</sup>Department of Economics, E52-262D, MIT, Cambridge, MA 02139. Phone: 617-253-6420, Fax: 617-253-1330, Email: wnewey@mit.edu.



## 1. Introduction

Nonlinear regression models with measurement error are important but difficult to estimate. Measurement error is a common problem in microeconomic data, where nonlinear models are often of interest. For example, flexible functional forms often lead to inherently nonlinear specifications. Instrumental variables estimators are not consistent for these models, as discussed in Amemiya (1985), so that alternative approaches must be adopted. The purpose of this paper is to develop an approach that is computationally feasible and also allows for flexibility in the distribution of disturbances. This purpose is accomplished by using simulated moments estimation with flexible distributions, an approach that may be useful for simulated moments estimation of other models.

The measurement error model considered here has a prediction equation for the true regressor with a disturbance that is independent of the predictors. The estimator is based on the conditional expectation of the dependent variable, and the conditional expectation of the product of the dependent variable and mismeasured regressor. This model for measurement error in nonlinear models has previously been considered by Hausman, Ichimura, Newey, and Powell (1991) and Hausman, Newey, and Powell (1995), but only for the case of polynomial regression or approximation, and simulated moments was not considered. This paper allows for general functional forms, significantly extending the scope of the previous work.

Much of the other work on measurement error in nonlinear models relies heavily on the assumption that the variance of the measurement error is small relative to the sample size. These papers include Wolter and Fuller (1982) and Amemiya (1985). In econometric practice the measurement error often seems quite large relative to the sample size, and has big effects on the coefficients. Thus, it seems important to consider approaches that allow for relatively large measurement error, as does the one here.

Simulated moments estimation provides a computationally convenient approach when estimating equations involve integrals, as discussed in Lerman and Manski (1981), Pakes (1986), McFadden (1989), and Pakes and Pollard (1989). This approach uses Monte Carlo methods to form unbiased estimators of integrals in moment equations. Allowing flexibility in disturbance distributions is desirable, because consistency of the estimator depends on correct specification of the distribution. Also, it is useful to preserve the computational convenience of simulated moments. These goals are accomplished by combining simulated moment estimation with a linear in parameters specification for distribution shape. The specification parameterizes ratio of the true density to the simulated one. This approach is similar to the importance sampling technique from the simulation literature.

The parametric simulated moments estimator we propose is essentially a generalized method of moment estimator. Here the moments are smooth in the parameters, so that standard asymptotic theory applies. For that reason we just give large sample inference procedures with an outline of the asymptotic theory for the parametric case. We pay more attention to conditions for consistency for the nonparametric case, giving a consistency result when the number of parameters in the distribution approximation is allowed to grow with sample size.

The paper also includes Monte Carlo and empirical applications, to evaluate the potential impact of this approach for applied work. The empirical application is estimation of Engel curves from household expenditure data. The measurement error correction makes a big difference in the application, with a Gaussian specification for prediction error sufficing in most cases. Also, the estimator seems quite accurate, having small standard errors. The results illustrate the usefulness of using simulated moment estimation to correct for measurement error, while allowing some flexibility in the distribution of the prediction error.

Section 2 describes the errors in variables model and some of its implications for conditional moments. Section 3 lays out the estimation method and discusses parametric

asymptotic inference for the estimator. Section 4 gives a semiparametric consistency result. Section 5 presents results of a small Monte Carlo study. Section 6 describes an empirical example of Engel curve estimation of the relationship between income and consumption.

## 2. The Model

The model considered here is

$$\begin{aligned}
 (2.1) \quad y &= f(w^*, \delta_0) + \zeta, & E[\zeta|x, v] &= 0, \\
 w &= w^* + \eta, & E[\eta|x, v, \zeta] &= 0, \\
 w^* &= \pi_0' x + \sigma_0 v, & v &\text{ independent of } x.
 \end{aligned}$$

where  $y$  and  $\zeta$  are scalars,  $\delta_0$ ,  $w^*$ ,  $w$ ,  $\eta$ ,  $x$ , and  $v$  are vectors, and  $\pi_0$  and  $\sigma_0$  are conformable matrices. Here  $w^*$  represents true regressors,  $\eta$  measurement errors, and  $w$  observed regressors. The  $x$  are observed and  $w^*$ ,  $\zeta$ ,  $\eta$ , and  $v$  may be unobserved. The last equation is a prediction equation for the true regressors, where  $x$  are observed predictor variables,  $v$  is an unobserved prediction error, and  $\sigma_0$  is a scaling matrix, a square root of a variance matrix. Some of the true regressors can be allowed to be observed, with  $w^*$  equal to an element of  $x$ , by specifying that corresponding elements of  $\eta$  and  $\sigma_0 v$  are identically zero, and the corresponding element of  $\pi_0' x$  is  $w^* = w$ . This model was considered by Hausman, Ichimura, Newey, and Powell (1991) (HINP henceforth), for the special case where  $f(w^*, \delta)$  is a polynomial in  $w^*$ . As long as  $x$  includes a constant, the location and scale of  $v$  can be normalized, e.g. as  $E[v] = 0$  and  $\text{Var}(v) = I$  when the second moment of  $v$  exists.

Instrumental variables (IV) estimators can be used to estimate this model when



$f(w^*, \delta)$  is linear in  $w^*$ . Substituting  $w - \eta$  for  $w^*$  in the first equation leads to  $x$  being valid instruments, because the disturbance is linear in the measurement error  $\eta$ . In the nonlinear case this substitution leads to residuals that are nonlinear in  $\eta$ . Consequently,  $x$  will not be valid instruments, and another approach has to be adopted.

An approach to consistent estimation can be based on integrating out the prediction error. Let  $g_0(v)$  be the density of  $v$ . Integrating over the prediction error leads to three useful conditional expectation equations:

$$(2.2a) \quad E[y|x] = \int f(\pi'_0 x + \sigma_0 v, \delta_0) g_0(v) dv,$$

$$(2.2b) \quad E[w \cdot y|x] = \int [\pi'_0 x + \sigma_0 v] f(\pi'_0 x + \sigma_0 v, \delta_0) g_0(v) dv,$$

$$(2.2c) \quad E[w|x] = \pi'_0 x.$$

The first is a regression of  $y$  on  $x$ , analogous to the usual one, except that the unobserved variable  $v$  has been integrated out. The second equation is a regression of  $w \cdot y$  on  $x$ , that is less familiar. The third equation is a standard regression equation.

The second equation is important for identification of nonlinear models. The components of this equation corresponding to unobserved  $w^*$  (i.e. those not corresponding to observed covariates) provide information additional to the first equation. As shown in HINP for polynomial regression, the first equation does not suffice for identification. Intuitively, there are two functions that need to be identified, the regression function and the density of  $v$ , so that two equations are needed for identification. It was shown in HINP that the parameters of any polynomial regression equation are identified from these two equations, and one expects that identification of the regression parameters will hold more generally.

It is beyond the scope of this paper to develop fully primitive identification

conditions for this model, but some things can be said. First, the parameters  $\pi_0$  are identified from (2.2c) as the coefficients of a least squares regression of  $w$  on  $x$ , so  $\pi_0$  can be treated as known and identification of the other pieces of the model considered by focusing on equation (2.2a) and (2.2b). If  $\pi_0'x$  has a discrete distribution with a finite support and  $m$  points of positive probability, then (2.2 a - b) provide  $2m$  equations. Assuming that none are redundant, i.e. that a "rank condition" holds, one could identify  $2m$  parameters from these equations, including  $\delta$  and parameters of a parametric family of distributions for  $v$ . For example, HINP showed that, in the case where  $w^*$  is a scalar and  $f(w^*, \delta)$  is a polynomial of degree  $p$ , the  $\delta$  parameters are identified if the second moment matrix of  $(1, \pi_0'x, \dots, (\pi_0'x)^{p+1})'$  is nonsingular. Also, some of the moments of  $v$  are identified in this case. If  $\pi_0'x$  has a continuous distribution, then a simple counting argument suggests that  $f(w^*, \delta_0)$  and  $g_0(v)$  should be identified. Assuming that the left hand sides of (2.2a) and (2.2b) are distinct functions, these equations give two functional equations, and there are two functions to be identified. So, by an analogy with the finite dimensional case, it should be possible, under appropriate regularity conditions, to identify both the regression function for  $w^*$  and the density function for  $v$ . Making this intuition precise would be quite difficult, because of the nonlinear, nonparametric (i.e. functional) nature of these equations, but it is an important problem deserving of future attention.

Independence of  $x$  and  $v$  is a strong assumption, but in the general nonlinear model of equation (2.1) it is difficult to drop this assumption. Intuitively, if some moments of  $v$  can depend on  $x$ , then it is much more difficult to separate the regression function from the distribution.

### 3. Estimation

To describe the estimator it is helpful to embed the model in a more general conditional moment setup. Let  $z$  denote a data observation,  $\beta$  a  $q \times 1$  vector of parameters,  $g$  a density function of a random vector  $v$ ,  $\rho(z, \beta, g)$  a  $r \times 1$  residual vector, and  $H(z, \beta, v)$  a  $r \times 1$  vector of functions, related as in

$$(3.1) \quad \rho(z, \beta, g) = \int H(z, \beta, v) g(v) dv.$$

Suppose that there is a set of conditioning variables  $x$  such that for the true parameter value  $\beta_0$  and density  $g_0$ ,

$$(3.2) \quad E[\rho(z, \beta_0, g_0) | x] = 0.$$

The nonlinear errors-in-variables model is a special case of this one, where

$$(3.3) \quad \begin{aligned} H_1(z, \beta, v) &= y - f(\pi'x + \sigma v, \delta), \\ H_2(z, \beta, v) &= L\{w \cdot y - [\pi'x + \sigma v] \cdot f(\pi'x + \sigma v, \delta)\}, \\ H_3(z, \beta, v) &= w - \pi'x, \quad \beta = (\delta', \sigma, \pi')', \end{aligned}$$

and  $L$  is a selection matrix that picks out those elements of  $w$  that include measurement error.

The common approach to using equation (3.2) in estimation is nonlinear instrumental variables. One difficulty with this approach is that the density  $g(v)$  is unknown. Another difficulty is that the residual is an integral that may be difficult to compute. Here, these difficulties are dealt with simultaneously, by choosing a flexible parameterization for the density that makes it easy to use a simulation estimator of the integral. To describe this approach, we begin with a specification of the density function.

For now, suppose that the density is a member of a parametric family, of the form

$$(3.4) \quad g(v, \gamma) = P(v, \gamma) \varphi(v), \quad P(v, \gamma) = \sum_{j=1}^J \gamma_j p_j(v),$$

where  $\varphi(v)$  is some fixed density function. For example, if  $\varphi(v)$  were standard normal and  $p_j(v) = v^{j-1}$ , then this would be an Edgeworth approximation. The function  $g(v, \gamma)$  need not be positive, but leads to residuals that are linear in the shape parameters  $\gamma$  and that can easily be estimated by simulation.

For a density like that of equation (3.4), a simulated residual can be constructed by drawing random variables from  $\varphi(v)$  and then evaluating the product of the linear combination  $P(v, \gamma)$  and the  $H$  functions. Let  $z_i$  denote a single observation and  $[v_{i1}, \dots, v_{iS}]$  denote a vector of random variables, each having marginal density  $\varphi(v)$ . For example, if  $\varphi(v)$  is a standard normal pdf, then  $[v_{i1}, \dots, v_{iS}]$  could be computer generated Gaussian random numbers. Then a simulated residual for the  $i^{\text{th}}$  observation is

$$(3.5) \quad \hat{\rho}_i(\theta) = S^{-1} \sum_{s=1}^S H(z_i, \beta, v_{is}) P(v_{is}, \gamma), \quad \theta = (\beta', \gamma')'.$$

This is essentially an importance sampling estimator of the residual, where  $\varphi(v)$  is sampling density and  $P(v, \gamma)$  approximates  $g(v)/\varphi(v)$ . The simulated residual is an unbiased estimator of the true residual, because

$$E[\hat{\rho}_i(\theta) | z_i] = \rho(z_i, \beta, g(\gamma)).$$

Therefore, by the results of McFadden (1989) and Pakes and Pollard (1989), an instrumental variables (IV) estimator with  $\hat{\rho}_i(\theta)$  as the residual should be consistent if the IV estimator with the true residual is. An IV estimator can be formed in a familiar way. Let  $\hat{A}(x)$  denote a  $q \times r$  vector of instrumental variables, that may be estimated. Suppose that  $\hat{\theta}$  solves

$$(3.6) \quad n^{-1} \sum_{i=1}^n \hat{A}(x_i) \hat{\rho}_1(\theta) = 0.$$

This is a simulated, nonlinear IV estimator like that of McFadden (1989).

Because equation (3.6) is linear in  $P(v, \gamma)$ , it is important to normalize the density  $P(v, \gamma)\phi(v)$  to integrate to one. Also, it may be important to impose a location and scale normalization on this density. There are different ways to impose normalizations by imposing constraints on the coefficients. For example, if  $\phi(v)$  is the standard normal density and  $p_1(v), p_2(v), \dots$  are the Hermite polynomials that are orthonormal with respect to the standard normal density (i.e.  $\int p_j(v)^2 \phi(v) dv = 1$  and  $\int p_j(v) p_k(v) \phi(v) dv = 0$  for  $j \neq k$ ), then  $\gamma_1 = 1, \gamma_2 = 0$ , and  $\gamma_3 = 0$  will imply that  $P(v, \gamma)\phi(v)$  integrates to one, and has zero mean and unit variance. It is also possible to impose such constraints using the simulated values, by requiring that  $\gamma$  satisfy  $\sum_{i=1}^n \sum_{s=1}^S (1, v_{is}, v_{is}^2) P(v_{is}, \gamma) = 0$ .

In the nonlinear errors in variables model, it is convenient to work with a two step estimator, where the first step consists of estimation of  $\pi$  by least squares (LS), and the second step is an instrumental variables estimator using the first two residuals of equation (3.3). The first order conditions for such an estimator can be formulated as a solution to equation (3.6), if  $\hat{A}(x)$  is chosen in a particular fashion. Let  $\hat{\pi}$  be the LS estimator and

$$(3.7) \quad \hat{A}(x) = \text{diag}[\hat{B}(x), x],$$

where  $\hat{B}(x)$  has two columns and number of rows equal to the number of elements of  $(\delta', \sigma, \gamma')$ . Suppose that constraints are imposed on the  $\gamma$  coefficients such that  $\sum_{s=1}^S P(v_{is}, \gamma) = 0$ . Then the solution to equation (3.6), with  $\hat{A}(x)$  specified as in equation (3.7), requires that  $\hat{\pi}$  be the least squares estimator, and that the other parameters solve the equation



$$(3.8) \quad 0 = \sum_{i=1}^n \hat{B}(x_i) \hat{\rho}_i(\alpha), \quad \alpha = (\delta', \sigma, \gamma')',$$

$$\hat{\rho}_i(\alpha) = \begin{pmatrix} y_i \\ Lw_i y_i \end{pmatrix} - \frac{1}{S} \sum_{s=1}^S \begin{pmatrix} f(\hat{\pi}' x_i + \sigma v_{is}, \delta) \\ L(\hat{\pi}' x_i + \sigma v_{is}) f(\hat{\pi}' x_i + \sigma v_{is}, \delta) \end{pmatrix} P(v_{is}, \gamma)$$

In the empirical example the estimator minimizes a quadratic form that has this type of equation as its first order condition, although the normalization  $\sum_{s=1}^S P(v_{is}, \gamma) = 0$  is not imposed. Specifically, for  $\hat{C}(x)$  equal to a vector of instrumental variables and  $\hat{W}$  a positive definite matrix,  $\hat{\alpha}$  solves

$$(3.9) \quad \min_{\alpha} [\sum_{i=1}^n \hat{C}(x_i) \hat{\rho}_i(\alpha)]' \hat{W} [\sum_{i=1}^n \hat{C}(x_i) \hat{\rho}_i(\alpha)].$$

The first order conditions to this minimization problem are as given in equation (3.8), with

$$(3.10) \quad \hat{B}(x) = [\partial \sum_{i=1}^n \hat{C}(x_i) \hat{\rho}_i(\hat{\alpha}) / \partial \alpha]' \hat{W}.$$

Standard large sample theory for IV can be used for asymptotic inference procedures. If the simulated values  $(v_{i1}, \dots, v_{iS})$  are included with the data to form an augmented observation for the  $i^{\text{th}}$  data point, then the usual IV formulae can be used to form a consistent variance estimator. For example, suppose that  $(z_i, v_{i1}, \dots, v_{iS})$  are independent observations as  $i$  varies. Then under standard regularity conditions (e.g. see Newey and McFadden, 1994), the asymptotic variance of  $\sqrt{n}(\hat{\theta} - \theta_0)$  can be estimated by

$$(3.11) \quad \hat{V} = \hat{G}^{-1} \hat{\Omega} \hat{G}^{-1}, \quad \hat{G} = n^{-1} \sum_{i=1}^n \hat{A}(x_i) \partial \hat{\rho}_i(\hat{\theta}) / \partial \theta, \quad \hat{\Omega} = n^{-1} \sum_{i=1}^n \hat{A}(x_i) \hat{\rho}_i(\hat{\theta}) \hat{\rho}_i(\hat{\theta})' \hat{A}(x_i)'.$$

If the simulation draws  $v_{is}$  are mutually statistically independent as  $s$  varies for a given  $i$ , then one could also use the estimator

$$(3.12) \quad \tilde{V} = \hat{G}^{-1} \tilde{\Omega} \hat{G}^{-1}, \quad \tilde{\Omega} = n^{-1} \sum_{i=1}^n \hat{A}(x_i) \hat{A}_i' \hat{A}(x_i),$$

$$\hat{A}_i = S^{-1} \sum_{s=1}^S H(z_i, \hat{\beta}, v_{is}) H(z_i, \hat{\beta}, v_{is})' P(v_{is}, \hat{\gamma})^2.$$

Both of these variance estimators ignore estimation of the instruments, which is valid under standard regularity conditions. Because the large sample theory for these estimators is straightforward, we do not give regularity conditions here.

#### 4. Consistent Semiparametric Estimation

If the functional form of the density  $g_0(v)$  is left unspecified, then the model becomes semiparametric. Models where identification is achieved by conditional moment restrictions like those of equation (3.1) are nonlinear, nonparametric simultaneous equations models. Newey and Powell (1991) have considered estimation of such models, and their result can be applied here. The basic idea is to apply the previous estimator, but with  $P(v, \gamma)$  chosen to be a member of an increasing sequence of approximating families and the IV equation (3.6) replaced by a nonparametric conditional expectation equation.

Let  $\mathcal{S}$  be a set of functions of  $v$  that will be assumed to include the true density  $g_0(v)$  and satisfy other regularity conditions given below. Also, let  $\{P_J(v, \gamma)\}_{J=1}^{\infty}$  be a sequence of families such that  $P_J(v, \gamma)\varphi(v)$  can approximate any function. Let  $\mathcal{S}_J = \{P_J(v, \gamma)\varphi(v)\} \cap \mathcal{S}$ ,  $\theta = (\beta, g)$  be the parameter consisting of the Euclidean vector  $\beta$  and a density  $g$ , and  $\hat{\rho}_i(\theta)$  be the simulated residual of equation (3.5) with  $P_J(v, \gamma)$  replacing  $P(v, \gamma)$ . Let  $\hat{E}[\hat{\rho}_i(\theta)|x_i]$  be a nonparametric estimator of the conditional expectation of  $\hat{\rho}_i(\theta)$  given  $x_i$ , such as a series or kernel estimator. Then a minimum distance estimator of  $\theta_0 = (\beta_0, g_0)$  is

$$(4.1) \quad \hat{\theta} = \underset{\theta \in \mathcal{B} \times \mathcal{G}}{\operatorname{argmin}} \hat{Q}(\theta); \quad \hat{Q}(\theta) \equiv \sum_{i=1}^n \hat{E}[\hat{\rho}_i(\theta) | x_i]' \hat{D} \hat{E}[\hat{\rho}_i(\theta) | x_i] / n,$$

where  $\hat{D}$  is a positive definite matrix and  $\hat{J}_n$  can depend on the data and on sample size. The objective function in equation (4.1) is a sample analog of

$$Q(\theta) = E[E[\rho(z, \theta) | x]' D E[\rho(z, \theta) | x]],$$

where  $D$  is the limit of  $\hat{D}$  and  $\rho(z, \theta) = \int H(z, \beta, v) g(v) dv$ . If  $D$  is positive definite and  $\theta_0$  is identified from the conditional moment equation (3.2) (i.e. that equation has a unique solution), then  $Q(\theta)$  will have a unique minimum of zero at  $\theta_0$ . The general extremum estimator reasoning (e.g. Newey and McFadden, 1994) then suggests that  $\hat{\theta}$  should be consistent.

The estimator can be shown to be consistent if  $\beta$  and  $g$  are restricted to compact sets, similarly to Gallant (1987). The compact function set assumption is a strong one, but the results of Newey and Powell (1991) indicate its importance for minimum distance estimators of the form considered here. For a matrix  $A = [a_{ij}]$ , let  $\|A\| = [\operatorname{trace}(A'A)]^{1/2}$ , and for a function  $g(v)$  let  $\|g\|$  denote a function norm, to be further discussed below.

Assumption 4.1:  $\beta_0 \in \mathcal{B}$ , which is compact, and  $g_0 \in \mathcal{G}$ , a compact set in a norm  $\|g\|$ .

In the primitive regularity conditions given below,  $\|g\|$  will be a Sobolev norm,

The following dominance condition will be useful in showing uniform convergence.

Assumption 4.2: There exists  $M(z, v)$  such that for  $\tilde{\beta}, \beta \in \mathcal{B}$ ,  $\|H(z, \tilde{\beta}, v)\| \leq M(z, v)$ ,  $\|H(z, \tilde{\beta}, v) - H(z, \beta, v)\| \leq M(z, v) \|\tilde{\beta} - \beta\|$ .

Moment conditions for the dominating function  $M(z, v)$  will be specified below.

To show uniform convergence of the objective function of equation (4.1), it is useful to impose a strong condition on function norm, that it dominates a weighted

supremum norm. Let  $V$  denote the support of  $\varphi(v)$ , and

$$\|g\|_{V,\omega} \equiv \sup_V |g(v)|\omega(v), \quad \omega(v) > 0.$$

Below it will be assumed that  $\|g\|_{V,\omega}$  is dominated by  $\|g\|$ , so that Assumption 4.1 implies that  $\|g\|_{V,\omega}$  is bounded on  $\mathcal{G}$ . The import of this assumption is a uniform bound on the tail behavior of  $g(v)$ , imposed by the presence of the weight function; the faster  $\omega(v)$  grows as  $v$  moves outward, the faster the tails of  $g(v)$  must go to zero in order to guarantee that  $\sup_V \{|g(v)|\omega(v)\}$  is finite. Also, the nature of importance sampling imposes a restriction on the tail thickness of the true density relative to the baseline density. For second moment dominance this restriction will translate into a restriction on  $\omega(v)$  relative to  $\varphi(v)$ . These considerations lead to the following assumption:

Assumption 4.3:  $\|g\|_{V,\omega} \leq \|g\|$  for  $g \in \mathcal{G}$  and there exists  $\epsilon > 0$  such that  $E[\int_V M(z,v)^{2+\epsilon} [\omega(v)\varphi(v)]^{-1-\epsilon} \omega(v)^{-1} dv] < \infty$ .

In order to guarantee that the parametric approximation suffices for consistency, the following denseness condition will be imposed.

Assumption 4.4: For any  $g \in \mathcal{G}$  and  $J$  there exists  $P_J(v,\gamma)\varphi(v) \in \mathcal{G}_J$  such that  $\lim_{J \rightarrow \infty} \|P_J(\cdot,\gamma)\varphi - g\| = 0$ .

This condition specifies that  $g_0$  can be approximated by the family.

It is necessary to make some assumption concerning the conditional expectations estimator. The following condition is lifted from Newey and Powell (1991). Without changing notation assume that the data observation  $z_i$  includes the simulation draws  $(v_{i1}, \dots, v_{iS})$ . Assume that the data are stationary.



Assumption 4.5: For  $\epsilon > 0$  from Assumption 4.3, i)  $\hat{D} \xrightarrow{P} D$ ,  $D$  is positive definite, and if  $E[|\psi(z_i)|^{1+\epsilon/2}] < \infty$  then  $\sum_{i=1}^n \psi(z_i)/n \xrightarrow{P} E[\psi(z_i)]$ ; ii) if  $E[\psi(z)^{2+\epsilon}]$  is finite,  $\sum_{i=1}^n \|\hat{E}[\psi(z)|x_i] - E[\psi(z)|x_i]\|^2/n \xrightarrow{P} 0$ ; iii) either a)  $\hat{E}[\psi(z)|x_i] = \sum_{j=1}^n w_{ij} \psi(z_j)$ ,  $w_{ij} \geq 0$ ,  $\sum_{j=1}^n w_{ij} = 1$ ,  $(i,j=1,\dots,n)$ , and and if  $E[|\psi(z)|^{1+\epsilon/2}] < \infty$ ,  $\sum_{i=1}^n \hat{E}[\psi(z)|z_i]/n = O_p(1)$ ; or b)  $\hat{E}[\psi(z)|z_i] = P'_i (\sum_{j=1}^n P_j P'_j)^{-1} \sum_{j=1}^n P_j \psi(z_j)$ .

Assumption 4.5 can easily be checked in some cases and is quite general. For instance, if  $z_i$  is i.i.d. then it is easy to use known results to show that Assumption 4.5 holds for nearest neighbor and series estimators. For K-nearest-neighbor estimators with  $K \rightarrow \infty$ ,  $K/n \rightarrow 0$ , ii) follows by Lemma 8 of Robinson (1987) and Proposition 1 of Stone (1977), while iii) a) holds by construction. For a series estimator of the form given in iii) b), with  $P_t$  containing  $K$  elements such that any function with finite mean square can be approximated arbitrarily well in mean-square for large enough  $K$ , ii) follows from Lemma A.10 of Newey (1994a) and the arguments for Lemma A.11 as long as  $K \rightarrow \infty$  and  $K/n^{\epsilon/(\epsilon+2)} \rightarrow 0$ .

Neither of these results allow for data based  $K$ . It should be noted that Assumption 4.5 restricts the form of randomness. Implicitly the form of the weights  $w_{st}$  in Assumption 4.5 and the approximating functions  $P_t$  are restricted to not depend on  $\psi$ . Thus, while they could be chosen based on some fixed  $\psi$  (e.g. a linear combination of  $\rho(z, \bar{\theta})$  for some preliminary estimator  $\bar{\theta}$ ), they are not allowed to vary with  $\psi$  (i.e. with  $\theta$  in  $\rho(z, \theta)$ ). Assumption 4.5 should also be "plug-compatible" with future results on nonparametric conditional expectation estimators, such as those for time series.

The last assumption specifies that  $\hat{J}_n$  must go to infinity with the sample size.

Assumption 4.6:  $\hat{J}_n \xrightarrow{P} \infty$  as  $n \rightarrow \infty$ .

As mentioned earlier, the degree of approximation  $J$  can be random, in a very general



way. However, it should be noted that it is not restrictions on the growth rate of  $J$  that are used to obtain consistency, but rather the restriction of the function to a compact set. Often, the compactness condition will require that higher order derivatives be uniformly bounded, a condition that will have more "bite" for large values of  $J$ , imposing strong constraints on the coefficients of higher order terms.

These assumptions and identification deliver the following consistency result:

*Theorem 4.1: If  $E[p(z, \theta)|x] = 0$  has a unique solution on  $\mathcal{B} \times \mathcal{S}$  at  $\theta_0$  and Assumptions 4.1 - 4.6 are satisfied, then  $\|\hat{\beta} - \beta_0\| \xrightarrow{P} 0$  and  $\|\hat{g} - g_0\| \xrightarrow{P} 0$ .*

It should be noted that the hypotheses of this theorem are not very primitive until the norm  $\|g\|$  is specified. Once that is specified, it may require some work to check the the other assumptions.

The following set of Assumptions is sufficient to demonstrate that the assumptions are not vacuous, and do cover cases of some interest.

*Assumption 4.7:* i)  $v$  is one-dimensional; ii) There is a compact interval  $V$  and a fixed constant  $B$  such that  $\mathcal{S} = \{g(v) : g(v) = 0 \text{ for } v \notin V, \sup_v |g(v)| \leq B, |g(\tilde{v}) - g(v)| \leq B|\tilde{v} - v| \text{ for all } \tilde{v}, v \in V\}$ ; iii)  $\|g\| = \sup_V |g(v)|$ ; iv) The support of  $\varphi(v)$  is  $V$  and  $\varphi(v)$  is continuous and bounded away from zero on  $V$ ; v)  $P(v, \gamma) = \sum_{j=0}^J \gamma_j v^j$ ; vi)  $E[\sup_{v \in V} M(z, v)^{2+\epsilon}] < \infty$ .

*Corollary 4.2: If Assumptions 3.1, 4.2, 4.5 - 4.7 are satisfied,  $\beta_0 \in \mathcal{B}$ , satisfied, and  $\mathcal{B}$  is compact then  $\|\hat{\beta} - \beta_0\| \xrightarrow{P} 0$  and  $\|\hat{g} - g_0\| \xrightarrow{P} 0$ .*

This result is restrictive in several ways. It is easy to relax the assumption that  $v$  is one dimensional, using the results of Elbadawi, Gallant, and Souza (1983). It is more difficult to allow for noncompact support for  $v$ , although this extension is possible using the results of Gallant and Nychka (1987). Unfortunately, their result allows for quite thick tails, with  $\omega(v) = C(1+v'v)^\Delta$  in Assumption 4.3. This tail behavior does

not allow Assumption 4.3 to be satisfied when  $\varphi(v)$  is the standard normal density. Of course, there are fast computational methods for generating data from densities proportional to  $(1+v'v)^{-\Delta}$ , so that one could easily use such thick-tailed baseline densities. Also, it should be possible to develop intermediate conditions that allow for more general simulators.

## 5. A Sampling Experiment

A small Monte Carlo study is useful in a rough check of whether the estimator can give satisfactory results in practice. Consider the model

$$(5.1) \quad y = \delta_1 + \delta_2 w^* + \delta_3 e^{-w^*} + \varepsilon, \quad \delta_1 = \delta_2 = 1, \quad \zeta \text{ is } N(0,1),$$

$$w = w^* + \eta, \quad \eta \text{ is } N(0,1),$$

$$w^* = \pi_1 + \pi_2 x + v, \quad \pi_1 = \pi_2 = 1, \quad x \text{ and } v \text{ are } N(0,.5).$$

The regression equation for this model is one that is useful in estimating the relationship between consumption and income. This specification will be further discussed in Section 6, where it is used in the empirical example. The parameter values were set so that the r-squared for the prediction equation for  $w^*$  was 1/2, and so the signal to noise ratio was 1. The number of observations was set to 100. The number of observations was chosen to be small relative to typical sample sizes in economics, to make computation easier. The r-squared for the regression of  $w^*$  on  $x$  was set higher than typical in order to offset the small sample size, so that the estimator might be informative.

Table One reports the results from 100 replications. Results for three different

estimators of  $\delta_1$ ,  $\delta_2$ , and  $\delta_3$  are reported. The first estimator is an ordinary least squares (OLS) regression of  $y$  on the right-hand side variables  $(1, w, e^{-w})$  that are measured with error. The second estimator is an instrumental variables estimator (IV) with the same right-hand side but with instruments  $\hat{R}_i = (1, \hat{h}_i, \hat{h}_i^2, \hat{h}_i^3)'$ , where  $\hat{h}_i = \hat{\pi}_1 + \hat{\pi}_2 x_i$ . The third estimator is a simulated moment estimator (SM) from equation (3.9), with  $\hat{C}(x_i) = I \otimes \hat{R}_i$  and  $\hat{W} = I \otimes (\sum_{i=1}^n \hat{R}_i \hat{R}_i')^{-1}$ , where  $I$  is a two dimensional identity matrix. This estimator is a system two-stage least squares estimator where the instrumental variables are  $\hat{R}_i$ . Also,  $P(v, \gamma)$  was a Hermite polynomial of the third-order, where  $\gamma_1 = 1$ ,  $\gamma_2 = \gamma_3 = 0$ , and  $\gamma_4$  was estimated. There were two simulations per observation.

In one replication out of the 100 the estimator did not converge to a stationary point. This replication was excluded from the results, that are reported in Table One. The estimator shows promise. The standard errors of the IV and simulated moment estimators are much larger than the OLS estimator, but the biases are substantially smaller. As previously noted the IV estimator is inconsistent, although in this example it leads to bias reduction. It is interesting to note that the standard error of the SM estimator is smaller than that of the IV estimator. Thus, in this example the valid SM correction for measurement error leads to both smaller bias and variance than the inconsistent IV correction.

## 6. An Application to Engel Curve Estimation

The application presented here is estimation of Engel curves, a subject that has long been of interest in econometrics. Measurement error has recently been shown in Hausman, Newey, and Powell (1995) to be important in the estimation of nonlinear Engel curves. This section adds to that work by estimating a nonlinear, nonpolynomial Engel

curve for the model of equation (2.1), a specification that was not estimated in Hausman, Newey, and Powell (1995). Also, the results here take account of measurement error in the denominator of the share equation.

The functional form considered here is that preferred by Leser (1963),

$$(5.1) \quad S_i = \delta_1 + \delta_2 \ln(I_i^*) + \delta_3 (1/I_i^*) + \varepsilon_i,$$

where  $S_i$  is the share of expenditure on a commodity and  $I_i^*$  is the true total expenditure. As suggested by the Hausman, Newey, and Powell (1993) tests of the Gorman (1981) rank restriction, a rank two specification such as this may be a good specification, once the measurement error has been accounted for.

In addition, a specification is considered that accounts for the presence of  $I_i^*$  in the denominator of the left-hand side of this equation. This "denominator problem" results from the fact that  $S_i = Y_i / I_i^*$ , where  $Y_i$  is the expenditure on the commodity. Thus, if  $I_i^*$  is measured with error, another nonlinear measurement error problem results from using the measured shares. This problem can be dealt with by bringing  $I_i^*$  out of the denominator, giving

$$(5.2) \quad Y_i = \delta_1 I_i^* + \delta_2 I_i^* \ln(I_i^*) + \delta_3 + I_i^* \varepsilon_i.$$

If  $\varepsilon_i$  satisfies the usual restriction  $E[\varepsilon_i | I_i^*] = 0$ , then equations (5.1) and (5.2) are equivalent statistical specifications, in that running least squares on either equation should give a consistent estimator. Covariates will also be allowed in this specification by allowing additional variables  $I_i^* x_{li}$  to enter linearly in this equation, corresponding to inclusion of  $x_{li}$  as additional regressors in the share equation (5.1).

The measurement error will be assumed to be multiplicative, i.e. for  $I_i$  equal to the observed total expenditure,



$$(5.3) \quad \ln(I_i^*) = w_i^* = \pi_0' x_i + v_i, \quad \ln(I_i) = \ln(I_i^*) + \eta_i = w_i = w_i^* + \eta_i.$$

In the empirical work the predictor variables  $x_i$  will be a constant, age and age squared for household head and spouse, and dummies for educational attainment, spouse employment, home ownership, industry, occupation, region, and black or white, a total of 19 variables, including the constant. With this specification for the measurement and prediction equations,  $f(w^*, \delta) = \delta_1 + \delta_2 w^* + \delta_3 \exp(-w^*)$ , as in the Monte Carlo example.

The measurement error in the left-hand side denominator can be accounted for as in equation (5.2), leading to a specification with  $f(w^*, \delta) = \delta_1 \exp(w^*) + \delta_2 \exp(w^*) w^* + \delta_3$ . It is interesting to note that even if the share equation is linear in  $\ln(I_i^*)$ , so that  $\delta_3 = 0$ , this equation is nonlinear, so that IV will not be consistent. Thus, measurement error in the denominator of the share suggests the need for the estimators developed here.

The data used in estimation are from the 1982 Consumer Expenditure Survey (CES). The basic data we use are total expenditure and expenditure on commodity groups from the first quarter of 1982. Results were obtained for four commodity groups, food, clothing, transportation, and recreation. The number of observations in the data set is 1321. The empirical results were reported as elasticities, i.e.  $d \ln f(x) / d \ln x$ , as is common in econometrics. To compare shapes, elasticities were calculated at the quartiles of observed expenditure.

The results are given in Tables Two through Five. Table Two gives some sample statistics, including the quartiles of the income distribution. The other tables will include estimated expenditure elasticities at these quartiles. Table Two also gives information on the prediction regression. The  $R^2$  in this regression is .23, which is quite sizable for such a cross-section data set. The other information is useful in calculating the magnitude of the measurement error and bounding the size of the variance of the prediction error  $v$ . In particular, the model we have assumed implies that the



standard error .45 of the residual is an upper bound on the standard deviation of both the measurement error and the variance of the prediction error  $v$ .

Also, given an estimator  $\hat{\sigma}$  of  $\text{Var}(v)^{1/2}$ , an estimator of the  $R^2$  of the measurement equation, that determines the magnitude of the measurement error bias in a linear model, is  $\hat{\text{Var}}(\pi'x + v)/\hat{\text{Var}}(w) = [\hat{\text{Var}}(\pi'x) + \hat{\sigma}^2]/\hat{\text{Var}}(w) = [(.25)^2 + \hat{\sigma}^2]/(.51)^2 \cong .24 + (3.8)\hat{\sigma}^2$ .

Tables Three to Five give results for each commodity for three different specifications of the share equation and four different estimators. Table Three gives results for the share equation, where measurement error in the denominator of the left-hand side is ignored. This specification is the same as in the Monte Carlo study. Table Four changes the specification to account for the left-hand side denominator by multiplying through the original equation by total expenditure, as described above. Table Five adds covariates  $x_1$  to the share equation to allow for demographic and regional price effects. There are six covariates; own and spouse age, family size, and three regional dummy variables. The equation estimated is analogous to that of Table Four in accounting for the left-hand side denominator, with  $f(w, x_1, \delta) = \delta_1 + \delta_2 \exp(w^*)w^* + \delta_3 + \exp(w^*)x_1' \delta_4$ . It should be noted that this specification restricts family size to be absent from the prediction equation.

Tables Three to Five report results for four different estimators, ordinary least squares (LS), two stage least squares (IV) with instruments described below, the simulated moment estimator with Gaussian  $v$  (SMO), and the simulated moment estimation with one Hermite polynomial term (SM1), of the third order, included in the moment functions. The simulated moment estimators are each obtained as in equation (3.9), with  $\hat{\rho}_1(\alpha)$  as given in equation (3.8), 10 simulation draws, and  $\hat{W}$  equal to the inverse of an estimated asymptotic variance of  $\sum_{i=1}^n \hat{C}(x_i) \hat{\rho}_1(\alpha) / \sqrt{n}$ . Specifically,

$$(5.4) \quad \hat{W} = \hat{\Sigma}^{-1}, \quad \hat{\Sigma} = n^{-1} \sum_{i=1}^n \hat{U}_i \hat{U}_i',$$

$$\hat{U}_i = \hat{C}(x_i) \hat{\rho}_1(\tilde{\alpha}) + [\partial \sum_{j=1}^n \hat{C}(x_j) \hat{\rho}_j(\tilde{\alpha}) / \partial \pi] (\sum_{j=1}^n x_j x_j')^{-1} x_i (w_i - \hat{\pi}' x_i),$$

where  $\tilde{\alpha}$  is an initial consistent estimator.<sup>3</sup> This is an asymptotic variance minimizing choice of  $\hat{W}$ , that accounts for the presence of  $\hat{\pi}$  in  $\hat{\rho}_1$ .

The standard errors for LS and IV were calculated from heteroskedasticity consistent formulae, e.g. as given in White (1982). The standard errors for simulated moment estimators were calculated from the GMM asymptotic variance estimator  $(\hat{H}'\hat{\Sigma}^{-1}\hat{H})^{-1}$ , where  $\hat{H} = \partial \sum_{i=1}^n \hat{C}(x_i)\hat{\rho}_1(\tilde{\alpha})/\partial \alpha$ .

A selection process was used to choose the order of powers of the predicted value to include in the instruments. Starting at the second order, the minimum needed to have enough moments to allow estimation of distribution parameters, the order was chosen by cross-validation on the food equation, Gaussian, simulated moment estimator (SMO), using the cross-validation criteria for choice of instruments suggested in Newey (1994b). Inclusion of higher order powers did not result in any decrease in the cross-validation criteria. Consequently, in Tables Three and Four the instrumental variables were  $(1, x'\hat{\pi}, (x'\hat{\pi})^2)$ . In Table Five  $\exp(x'\hat{\pi}) \cdot x_1$  was added to the instruments, because of the presence of the covariates.

The number of Hermite polynomial terms to include was chosen essentially by an upwards testing procedure, applied in the model of Table Three. Inclusion of a third order term was tried in each case, as reported in Table Three. This term allows for asymmetry in the distribution of  $v$ . If it was statistically significant, a fourth order term was tried. In none of the cases was this term significant, so only results for the one, third order, Hermite polynomial term are reported in the tables.

For each estimator, elasticities at the quartiles, the estimate of  $\sigma = \text{Var}(v)^{1/2}$ , and the estimator of the coefficient  $\gamma$  of the Hermite polynomial term, as well as standard errors (in parentheses below the estimates) are reported. The (asymptotic) t-statistic on the coefficient of inverse expenditure (t-stat) and the overidentification

<sup>3</sup> The procedure used to obtain the initial consistent estimators was to begin with an identity weighting matrix, use a few iterations to obtain "reasonable" parameter values, choose  $\hat{W}$  as in equation (5.4), and then minimize to get  $\tilde{\alpha}$ .

(minimum chi-square) test (Q) statistic for the simulated moment estimator are also reported. The t-statistic is particularly relevant in Table Three because the 2SLS estimator would be consistent if the coefficient on inverse expenditure were zero. The degrees of freedom of the overidentification test statistic are 2 and 1 respectively for SMO and SM1, in Tables Three and Four, and 8 and 7 respectively in Table Five. The difference between these statistics for SMO and SM1 is a one-degree of freedom chi-squared test of the Hermite coefficient being zero.

Even though the IV estimator is inconsistent, it gives results similar to the SM estimator in a number of cases. When the share denominator is allowed to be measured with error there are larger differences between IV and SM. The standard errors of SM are smaller than those of IV, which is consistent with the Monte Carlo results of Section 4. There are large differences between the OLS and SM estimators, as is consistent with the presence of measurement error. It is interesting to note that the elasticities for transportation go down rather than up, unlike linear regression with measurement error.

In comparing Tables 3 and 4, it is apparent that accounting for measurement error in the denominator leads to some changes in the results. There is more nonlinearity in the food equation in Table 4 than in Table 3. The prediction error standard deviation  $\sigma$  is more precisely estimated in these equations. The overidentification test statistics are larger in Table 4. Surprisingly, the estimated standard errors in Table 4 are not much larger than those in Table 3, although Table 4 is a levels equation that is sometimes thought to be more heteroskedastic than the share equation. There is little evidence of nonnormality. In most cases SMO is quite similar to SM1, except for much larger standard errors.

In summary, although allowing for nonnormality does not change the empirical results, correcting for measurement error makes a big difference. In several cases the simulated moments estimator is quite different than the inconsistent IV estimator, suggesting that the inconsistency of IV estimator may not be uniformly small. Furthermore, the simulated moment estimators seem quite accurate, having small standard

errors. These results illustrate the usefulness of using simulated moment estimation to correct for measurement error, while allowing some flexibility in the distribution of the prediction error to assess the impact of allowing for nonnormality.



## Appendix

Proof of Theorem 4.1: The proof proceeds by verifying the hypotheses of Theorem 5.1 of Newey and Powell (1991). Let the norm for  $\theta = (\beta, g)$  be  $\|\theta\| = \|\beta\| + \|g\|$ . Note that  $\Theta = \mathcal{B} \times \mathcal{G}$  is compact by  $\mathcal{B}$  and  $\mathcal{G}$  compact. For  $S$  simulations let  $Z$  denote the augmented data vector, with  $Z = (z, v_1, \dots, v_S)$ . Also, let  $\rho(Z, \theta) = S^{-1} \sum_{s=1}^S H(z, \beta, v_s) P(v_s, \gamma)$ . Note that for  $p_0(v) = g_0(v)/\varphi(v)$ , it follows by Assumptions 4.2 and 4.3 that by the triangle inequality

$$(A.1) \quad \|\rho(Z, \theta_0)\| \leq \sum_{s=1}^S \|H(z, \beta_0, v_s)\| |p_0(v_s)| / S \leq \left\{ \sum_{s=1}^S M(z, v_s) [\omega(v_s) \varphi(v_s)]^{-1} / S \right\} \|g_0\|,$$

$$(A.2) \quad \{E[\|\rho(Z, \theta_0)\|^{2+\epsilon}]\}^{1/(2+\epsilon)} \leq C \cdot \sum_{s=1}^S \{E[\{M(z, v_s) [\omega(v_s) \varphi(v_s)]^{-1}\}^{2+\epsilon}]\}^{1/(2+\epsilon)} / S \\ = C \cdot \{E[\int_V [M(z, v) \omega(v)]^{-2-\epsilon} \varphi(v)^{-1-\epsilon} dv]\}^{1/(2+\epsilon)} < \infty.$$

It follows similarly to equation A.1 that that for  $\tilde{\theta}$ ,  $\theta \in \Theta$ ,

$$(A.3) \quad \|\rho(Z, \tilde{\theta}) - \rho(Z, \theta)\| \leq C \cdot \left\{ \sum_{s=1}^S M(z, v_s) [\omega(v_s) \varphi(v_s)]^{-1} / S \right\} \|\tilde{\theta} - \theta\|,$$

so that Assumption 5.1 of Newey and Powell (1991) follows by eq. (A.2). Assumptions 5.2 and 5.3 then follow by Assumptions 4.4 and 4.5. Furthermore, by the fact that  $E[\rho(Z, \theta) | x] = E[\rho(z, \beta, g) | x]$  for an unbiased simulator, as noted in the text, Newey and Powell's (1991) Assumption 3.1 holds by Assumption 2.1. The conclusion then follows by the conclusion of Theorem 5.1 of Newey and Powell (1991). QED.

Proof of Corollary 4.2: The proof proceeds by verifying the hypotheses of Theorem 4.1. Assumption 4.1 follows by hypothesis and the Arzela theorem, which gives compactness of  $\mathcal{G}$  in the sup norm. Assumption 4.3 follows with  $\omega(v) = 1$  by  $\varphi(v)$  bounded away from zero and Assumption 4.6, vi). Assumption 4.4 follows by a Weierstrass approximation of  $g(v)/\varphi(v)$  by  $P_J(\gamma)$ . The proof then follows by the conclusion of Theorem 4.1. QED.



## References

- Amemiya, Y., 1985, Instrumental variables estimator for the nonlinear errors in variables model, *Journal of Econometrics* 28, 273-289.
- Elbadawi, I., A.R. Gallant, and G. Souza, 1983, An elasticity can be estimated consistently without a priori knowledge of functional form, *Econometrica* 51, 1731-1751.
- Gallant, A.R., 1987, Identification and consistency in nonparametric regression, in T.F. Bewley, ed., *Advances in econometrics: fifth world congress*, Cambridge: Cambridge University Press, 145-169.
- Gallant, A.R. and D.W. Nychka, 1987, Semi-nonparametric maximum likelihood estimation, *Econometrica* 55, 363-390.
- Gorman, W.M., 1981, Some engel curves, in A. Deaton, ed., *essays in the theory and measurement of consumer behaviour in honor of Sir Richard Stone*, Cambridge: Cambridge University Press.
- Hausman, J.A., H. Ichimura, J.L. Powell, and W.K. Newey, 1991, Estimation of polynomial errors in variables models, *Journal of Econometrics* 50, 273-295.
- Hausman, J.A., W.K. Newey, and J.L. Powell, 1995, Nonlinear errors in variables: Estimation of some Engel curves, *Journal of Econometrics* 65, 205-233.
- Leser, C.E.V., 1963, Forms of Engel functions, *Econometrica* 31, 694-703.
- Lerman, S. and C. Manski, 1981, On the use of simulated frequencies to approximate choice probabilities, in C. Manski and D. McFadden, eds., *Structural analysis of discrete data with econometric applications*, Cambridge: MIT Press.
- McFadden, D., 1989, A method of simulated moments for estimation of discrete response models without numerical integration, *Econometrica* 57, 995-1026.
- Newey, W.K., 1994a, Series estimation of regression functionals, *Econometric Theory* 10, 1-28.
- Newey, W.K., 1994b, Efficient estimation of models with conditional moment restrictions, in G.S. Maddala and C.R. Rao, eds., *Handbook of statistics, volume 11: Econometrics*. Amsterdam: North-Holland.
- Newey, W.K. and D. McFadden, 1994, Large sample estimation and hypothesis testing, in R. Engle and D. McFadden, eds., *Handbook of Econometrics, Vol. 4*, Amsterdam: North-Holland.
- Newey, W.K. and J.L. Powell, 1991, Instrumental variables estimation for nonparametric models, working paper, Department of Economics, MIT.
- Pakes, A., 1986, Patents as options: Some estimates of the value of holding European patent stocks, *Econometrica* 54, 755-785.
- Pakes, A. and D. Pollard, 1989, Simulation and the asymptotics of optimization estimators, *Econometrica* 57, 1027-1057.
- Robinson, P., 1987, Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form, *Econometrica* 55, 875-891.

Stone, C.J., 1977, Consistent nonparametric regression (with discussion), *Annals of Statistics* 5, 595-645.

White, H., 1982, Instrumental variables regression with independent observations, *Econometrica* 50, 483-499.

Wolter, K.M. and W.A. Fuller, 1982, Estimation of nonlinear errors in variables models, *Annals of Statistics* 10, 539-548.

Table One: Monte Carlo Results

	$\delta_1$			$\delta_2$			$\delta_3$		
	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
OLS	-1.10	.25	1.13	.67	.13	.67	.85	.12	.86
IV	-.30	3.31	3.32	.13	1.25	1.26	.47	2.01	2.07
SM	-.09	2.27	2.27	.07	1.05	1.05	.31	1.62	1.65

Table Two: Some Sample Statistics

	25th	50th	75th
Income Quartiles	3373	4574	6417
Sample standard error of log of expenditure			.51
Standard error of predicted			.25
Standard error of residual			.45
R-squared			.23

Table Three: Elasticity Estimates for Share Equations

				Food		t-stat	Q
	25th	50th	75th	$\sigma$	$\gamma$		
LS	.72 (.02)	.66 (.02)	.59 (.03)			4.52	
2SLS	.82 (.05)	.78 (.04)	.74 (.06)			.47	
SM0	.82 (.04)	.78 (.04)	.74 (.06)	.61 (.08)		3.67	6.36
SM1	.84 (.20)	.78 (.09)	.71 (.36)	.31 (1.49)	-.01 (.05)	.04	6.08

## Clothing

	25th	50th	75th	$\sigma$	$\gamma$	t-stat	Q
LS	1.21 (.05)	1.08 (.04)	.97 (.05)			18.66	
2SLS	1.61 (.12)	1.42 (.09)	1.30 (.10)			2.25	
SM0	1.63 (.20)	1.40 (.10)	1.26 (.18)	.02 (.38)		.56	6.11
SM1	1.62 (.46)	1.28 (.30)	1.07 (.56)	-.0009 (.0018)	.15 (.11)	.20	1.99

## Transportation

	25th	50th	75th	$\sigma$	$\gamma$	t-stat	Q
LS	1.28 (.07)	1.44 (.06)	1.50 (.07)			11.19	
2SLS	.99 (.08)	1.06 (.08)	1.12 (.12)			1.00	
SM0	1.02 (.07)	1.01 (.06)	1.01 (.06)	1.71 (2.01)		.04	11.28
SM1	1.40 (.27)	.98 (.07)	.63 (.18)	.10 (.07)	.028 (.018)	3.10	7.54

## Recreation

	25th	50th	75th	$\sigma$	$\gamma$	t-stat	Q
LS	1.40 (.07)	1.20 (.06)	1.06 (.07)			16.59	
2SLS	1.70 (.15)	1.31 (.12)	1.07 (.15)			11.45	
SM0	2.97 (.63)	1.33 (.12)	.39 (.48)	.02 (.34)		6.48	10.80
SM1	6.98 (4.40)	2.32 (.36)	1.28 (.12)	.71 (.18)	.024 (.002)	5.58	.09

Table Four: Elasticity Estimates for Level Equations

Food							
	25th	50th	75th	$\sigma$	$\gamma$	t-stat	Q
LS	.68 (.04)	.63 (.03)	.57 (.03)			.31	
2SLS	.90 (.08)	.80 (.05)	.70 (.05)			3.34	
SM0	.98 (.08)	.81 (.04)	.63 (.05)	.35 (.02)		12.4	18.16
SM1	1.21 (.21)	.83 (.06)	.46 (.13)	.24 (.06)	.008 (.005)	5.09	16.42

Clothing							
	25th	50th	75th	$\sigma$	$\gamma$	t-stat	Q
LS	1.40 (.11)	1.11 (.06)	.89 (.04)			53.95	
2SLS	2.04 (.26)	1.50 (.12)	1.21 (.13)			4.92	
SM0	2.07 (.21)	1.36 (.08)	.96 (.07)	.34 (.02)		39.04	17.15
SM1	2.14 (.58)	1.37 (.11)	.93 (.18)	.33 (.09)	.001 (.009)	3.19	17.13

Transportation							
	25th	50th	75th	$\sigma$	$\gamma$	t-stat	Q
LS	3.14 (1.06)	1.95 (.19)	1.48 (.09)			1.68	
2SLS	.23 (.65)	.93 (.12)	1.53 (.53)			2.04	
SM0	1.16 (.08)	.94 (.05)	.76 (.04)	.64 (.04)		38.27	10.63
SM1	1.25 (.59)	.97 (.20)	.74 (.09)	.58 (.27)	.004 (.021)	.69	10.52



Recreation.							
	25th	50th	75th	$\sigma$	$\gamma$	t-stat	Q
LS	1.74 (.19)	1.26 (.09)	.95 (.05)			42.87	
2SLS	1.84 (.28)	1.32 (.16)	1.01 (.15)			14.85	
SM0	2.35 (.26)	1.44 (.09)	.95 (.07)	.36 (.03)		48.62	33.23
SM1	7.85 (4.76)	1.93 (.34)	.33 (.15)	.25 (.02)	.024 (.005)	15.88	21.83

Table Five: Elasticity Estimates for Level Equations with Covariates

Food							
	25th	50th	75th	$\sigma$	$\gamma$	t-stat	Q
LS	.72 (.05)	.66 (.03)	.58 (.03)			1.90	
2SLS	.97 (.09)	.85 (.06)	.74 (.06)			4.40	
SM0	1.00 (.08)	.86 (.05)	.73 (.05)	.33 (.02)		7.14	43.51
SM1	1.25 (.29)	.90 (.07)	.58 (.15)	.21 (.07)	.009 (.007)	2.47	41.59

Clothing							
	25th	50th	75th	$\sigma$	$\gamma$	t-stat	Q
LS	1.35 (.11)	1.08 (.06)	.88 (.04)			32.76	
2SLS	2.01 (.31)	1.50 (.14)	1.22 (.15)			4.22	
SM0	1.94 (.22)	1.31 (.09)	.94 (.08)	.33 (.02)		29.32	43.23
SM1	2.01 (.73)	1.33 (.14)	.92 (.23)	.31 (.11)	.002 (.011)	1.59	42.95

# Transportation

	25th	50th	75th	$\sigma$	$\gamma$	t-stat	Q
LS	2.54 (.63)	1.83 (.14)	1.50 (.08)			1.30	
2SLS	.05 (.55)	.70 (.16)	1.38 (.45)			3.09	
SM0	1.07 (.09)	.87 (.06)	.69 (.04)	.61 (.04)		28.01	30.12
SM1	1.86 (.87)	1.13 (.26)	.61 (.07)	.40 (.08)	.016 (.009)	2.61	28.92

# Recreation

	25th	50th	75th	$\sigma$	$\gamma$	t-stat	Q
LS	1.73 (.19)	1.25 (.09)	.95 (.05)			35.96	
2SLS	1.78 (.30)	1.31 (.17)	1.01 (.16)			10.69	
SM0	2.40 (.32)	1.41 (.11)	.88 (.08)	.31 (.03)		37.86	54.23
SM1	8.38 5.45	1.91 (.32)	.18 (.23)	.21 (.03)	.021 (.005)	10.91	44.88

7076 010









Date Due

Lib-26-67

MIT LIBRARIES



3 9080 01972 0975







