

# SEMI-NONPARAMETRIC IV ESTIMATION OF SHAPE-INVARIANT ENGEL CURVES

---

*Richard Blundell*  
*Xiaohong Chen*  
*Dennis Kristensen*

THE INSTITUTE FOR FISCAL STUDIES  
DEPARTMENT OF ECONOMICS, UCL  
cemmap working paper CWP15/03

# Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves\*

Richard Blundell                      Xiaohong Chen  
University College London          New York University

Dennis Kristensen  
London School of Economics

First version: November 2001, Current version: October 2003

## Abstract

This paper concerns the identification and estimation of a shape-invariant Engel curve system with endogenous total expenditure. The shape-invariant specification involves a common shift parameter for each demographic group in a pooled system of Engel curves. Our focus is on the identification and estimation of both the nonparametric shape of the Engel curve and the parametric specification of the demographic scaling parameters. We present a new identification condition, closely related to the concept of bounded completeness in statistics. The estimation procedure applies the sieve minimum distance estimation of conditional moment restrictions allowing for endogeneity. We establish a new root mean squared convergence rate for the nonparametric IV regression when the endogenous regressor has unbounded support. Root- $n$  asymptotic normality and semiparametric efficiency of the parametric components are also given under a set of ‘low-level’ sufficient conditions. Monte Carlo simulations shed lights on the choice of smoothing parameters and demonstrate that the sieve IV estimator performs well. An application is made to the estimation of Engel curves using the UK Family Expenditure Survey and shows the importance of adjusting for endogeneity in terms of both the curvature and demographic parameters of systems of Engel curves.

**Keywords:** Consumer demands, nonparametric IV, bounded completeness, sieve minimum distance, measure of ill-posedness, nonparametric convergence rate, root- $n$  semiparametric efficiency.

---

\*We are thankful to C. Ai, I. Crawford, B. Honore, J. Horowitz, A. Lewbel, W. Newey, J. Powell and J-M. Robin for helpful comments. The first version was presented in 2001 North American Winter Meetings of the Econometric Society in New Orleans, and the seminars at CREST, Stanford, UPenn and Harvard/MIT. This study is part of the program of research of the ESRC Centre for the Microeconomic Analysis of Fiscal Policy at IFS. The financial support of the ESRC is gratefully acknowledged. The usual disclaimer applies.

## 1. Introduction

Consumer demand presents an important area for application of semi- and nonparametric methods. Nonparametric analysis of the Engel curve relationship is now common place (see, for example, Bierens and Pott-Butter (1990), Härdle and Jerison (1988) and Blundell and Duncan (1998)). There are many reasons why it is important to recover an accurate and flexible specification of the Engel curve relationship. Most obviously so as to guide the appropriate form of preferences to use in modelling consumer responses to policy reforms and in the nonparametric measurement of the welfare impact of such reforms. The Engel curve plays a key role in the analysis of revealed preference on micro data (see, for example, Blundell, Browning and Crawford (2003)). In estimating the impact of demographic change and equivalence scales, the shape of Engel curves is also critical (see Blackorby and Donaldson (1994), for example).

When pooling across households of different demographic specifications it is typical to adopt some semiparametric specification - nonparametric with regard to total expenditure and parametric with regard to demographic variables. Restrictions from consumer theory are not innocuous on the form in which demographics enter. In a nonparametric budget share specification, demographics cannot in general enter additively into each Engel curve equation while retaining consistency with optimisation theory. Blundell, Browning and Crawford (2003) show that they must also enter so as to scale the total expenditure variable inside the nonparametric Engel curve for each commodity. Consequently the simple partially linear semiparametric specification of Robinson (1988) has to be generalized. The generalisation we adopt corresponds to the 'base - independent' (or 'equivalence scale exactness') method of introducing demographics in demand analysis (see Blackorby and Donaldson (1994)). Interestingly this has the same form considered in the pooling of 'shape-similar' or 'shape-invariant' nonparametric regression curves of Härdle and Marron (1990) and Pinkse and Robinson (1995), explored further in the context of the demographic adjustment of Engel curves by Blundell, Duncan and Pendakur (1998).

The aim of the present paper is to extend this work to the case of endogenous regressors, in particular when the total expenditure variable is considered endogenous for individual commodity demands. The focus of attention is on semi-nonparametric estimation, that is on both the nonparametric estimation of the Engel curve shape and the estimation of the parametric specification of the demographic variables. The

attraction of the shape-invariant specification is that demographics simply shift and scale the function without altering its overall shape. If total expenditure is endogenous for individual commodity demands, then the conditional mean estimated by nonparametric least squares (LS) regression will not identify the economically meaningful ‘structural’ Engel curve relationship. That is the ‘statistical’ Engel curve will not recover the shape necessary for the analysis of consumer preferences, equivalences scales or expansion paths. In a semi-nonparametric regression framework of the type adopted here, there are two alternative approaches to estimation under endogeneity - the ‘instrumental variables’ (IV) and ‘control function’ approaches. Our aim in this paper is to develop the IV approach for this semi-nonparametric Engel curve case.<sup>1</sup>

The instrumental variable approach is investigated in Newey and Powell (2003), Darolles, Florens and Renault (2002) and Hall and Horowitz (2003) for the purely nonparametric regression model. Ai and Chen (2003) have considered the IV approach in the context of semiparametric efficient estimation of models with conditional moment restrictions containing unknown functions. In this paper we apply the sieve IV estimation methods of Newey and Powell (2003) and Ai and Chen (2003) to the semi-/nonparametric Engel curve application. We focus on the sieve IV method because we argue that it is an attractive alternative to kernel based methods for the semi-nonparametric IV case that we consider. Most existing papers on Engel curve models assume exogenous total expenditure and consider kernel-based methods. In this paper, given the shape-invariant semi-nonparametric specification, we argue that the sieve-based methods are easier to implement numerically. Moreover, this semi-nonparametric form is common in economic applications and will therefore be more generally applicable.

It is well-known that a purely nonparametric IV regression is a difficult ill-posed inverse problem, and has not been implemented in empirical research prior to the study reported in this paper. Although this paper applies the general sieve IV estimation method of Newey and Powell (2003) and Ai and Chen (2003), our theoretical justification is highly nontrivial. While Newey and Powell (2003) provide consistency of the nonparametric sieve IV estimators, and Ai and Chen (2003) obtain root-n asymptotic

---

<sup>1</sup>Blundell, Duncan and Pendakur (1998) allow for endogeneity of total expenditure using a parametric additive control function approach within the context of a kernel regression framework. See Newey, Powell and Vella (1999) for the development of the control function approach and also the reviews by Blundell and Powell (2003), Florens (2003) and Florens, Heckman, Meghir and Vytlačil (2002).

normality and semiparametric efficiency of estimators of the parametric components, their results are established under sets of relatively “high-level” sufficient conditions since they aim at very general models of conditional moment restrictions containing unknown functions. In our application to the shape-invariant semi-/nonparametric Engel curve models, we first provide identification under a “bounded completeness” condition, which is natural since Engel curves are all bounded between zero and one, and which is also much weaker than the “completeness” condition stated in Newey and Powell (2003) and Darolles, Florens and Renault (2002). Moreover, we are able to provide a set of “low-level” sufficient conditions for consistency of the sieve IV estimator of Engel curves, and for the root-n asymptotic normality and efficiency of the estimator of the parametric demographic effects. In addition, we obtain the nonparametric convergence rate in root mean squared metric when the endogenous log-total expenditure has unbounded support, which is a new contribution even in the literature on ill-posed inverse problems. The only other results on convergence rates of nonparametric IV regression are those by Darolles, Florens and Renault (2002) and Hall and Horowitz (2003).<sup>2</sup> We note that their estimation procedures and sufficient conditions for their convergence rate results are different from ours. In particular they assume that the endogenous regressor has bounded support, while we allow that the endogenous regressor to have unbounded support, which is natural in the shape-invariant Engel curve application with endogenous total expenditure.<sup>3</sup> In our convergence rate study, we introduce a “sieve measure of ill-posedness”, which directly affects the variance part hence the mean squared convergence rate of the sieve nonparametric IV estimator. The “sieve measure of ill-posedness” is identically one for the standard nonparametric LS regression, but is always greater than one and increases with sample size for the nonparametric IV regression. The greater the “sieve measure of ill-posedness” is, the bigger is the variance and the slower is the mean squared convergence rate of the sieve IV estimator.

The application of the sieve IV system estimator is to the estimation of a system of Engel curves describing the allocation of total non-durable consumption expenditure

---

<sup>2</sup>In mathematics and statistics literature, there are plenty results on convergence rates for linear ill-posed inverse problems of the form  $Th = g$ , where  $T$  is a known compact operator and  $g$  is known up to a small additive perturbation  $\delta$ , see e.g. Kress (1999) and Korostelev and Tsybakov (1993). However, the nonparametric IV regression in econometrics corresponds to an ill-posed inverse problem with both the conditional expectation operator  $T$  and the  $g$  being unknown.

<sup>3</sup>Although one could transform an endogenous regressor with unbounded support into another one with bounded support in a purely nonparametric IV regression, it is difficult to do so in the semi-nonparametric engel curve application without destroying the shape invariant specification.

across eight groups of non-durables and services for a sample of families with and without children in the UK Family Expenditure Survey (FES). The FES records detailed information on expenditures, incomes and family composition and has been a central data source for many applications of consumer behavior. In the application we select only working age families in which the head is in employment. Total expenditure is allowed to be endogenous and we use the gross earnings of the household head as an instrument for total expenditure. The idea is that consumption and savings are jointly endogenous, so that total expenditure may be endogenous for the individual consumption shares. We find the estimated curves and demographic parameters to be plausible and we document a significant impact of accounting for the endogeneity of total expenditure. Adjusting for endogeneity increases the common demographic shift parameter and produces a much more plausible estimate of the income equivalence scale.

A Monte Carlo study is included to assess the performance of the sieve IV estimator, where the simulation is designed to mimic the subsample of household without children from the FES data. The estimated “sieve measure of ill-posedness” is relatively large for the subsample of couple without children, which translates into a slow mean squared convergence rate of the sieve IV estimator given a typical sample size and given a finite smoothness of the true unknown Engel curve function. Nevertheless, the Monte Carlo results indicate that the slow convergence rate is mainly due to the large variance and not due to the bias. In contrast, the inconsistent sieve LS estimator has a small variance but large bias. We find that there are choices of smoothing parameters which reduce the variance hence make the mean squared errors of the sieve IV estimators small, while there is no choice of smoothing parameters which can reduce the large bias of the inconsistent sieve LS estimator in the simulation. All these findings are consistent with our theoretical result on the convergence rate of sieve IV estimator.

The structure of the remaining paper is as follows. Section 2 sets up the semi-/nonparametric Engel curve model specification, and discusses the issue of endogenous total non-durable expenditure. Section 3 considers the identification and estimation of the system of shape invariant semi-/nonparametric Engel curves. Section 4 provides consistency and nonparametric convergence rates of the sieve IV estimators of Engel curves, and Section 5 obtains root-n asymptotic normality and efficiency of the estimators of the parametric parts. Section 6 discusses the actual implementation of the sieve estimation procedure for the system of shape invariant Engel curves, and presents a small Monte Carlo study to assess the performance of the nonparametric IV estima-

tor. Section 7 reports the empirical performance of the sieve semi-nonparametric IV estimators. Section 8 briefly concludes. All proofs and some lemmas are collected into the Appendix.

## 2. Model Specification

As a baseline Engel curve specification for our empirical application we adopt the standard Working-Leser (Working (1943) and Leser (1963)) or Piglog specification in which budget shares are expressed in terms of log total expenditure. Indeed, the empirical investigations by Working and Leser established the Piglog form, in which the budget share for each commodity is a linear function of log total budget, as a reasonable specification for certain goods - in particular the budget share of food. This form for the Engel curve relationship also underlies the popular Almost Ideal and Translog demand models of Deaton and Muellbauer (1980) and Jorgenson, Lau and Stoker (1982).

Subsequent models of consumer behaviour have typically followed this specification for Engel curves, although many have pointed out the restrictive nature of this specification for many commonly used commodities (see Banks, Blundell and Lewbel (1997), for example). Recent attention has focused on Engel curves which have more variety of curvature than is permitted by the Piglog. This reflects growing evidence from a series of empirical studies that suggest higher order logarithmic expenditure terms are required for certain expenditure share equations (see, for example, Hausman, Newey, Ichimura and Powell (1991), Lewbel (1991) and Banks, Blundell and Lewbel (1997).)

Let  $\{(Y_{2i}, Y_{1il})\}_{i=1}^n$  represent a sequence of  $n$  household observations on the log of total expenditure  $Y_{2i}$  and on the  $l$ -th budget share  $Y_{1il}$ , for each household  $i$  facing the same relative prices. For each commodity  $l$ , budget shares and total outlay are related by the stochastic Engel curve

$$Y_{1il} = g_l(Y_{2i}) + \varepsilon_{il} \quad (2.1)$$

where we assume that, for each household  $i$ , the unobservable term  $\varepsilon_{il}$  satisfies

$$E(\varepsilon_{il}|Y_{2i}) = 0 \text{ and } Var(\varepsilon_{il}|Y_{2i}) = \sigma_l^2(Y_{2i}) \forall \text{ goods } l = 1, \dots, L + 1 \quad (2.2)$$

so that the nonparametric regression of budget shares on log total expenditure estimates  $g_l(Y_{2i})$ . This model generalises standard Engel curve specifications implied by popular demand models.

## 2.1. Demographic Composition and Semiparametric Specification

Household expenditures typically display a large variation with demographic composition. Let  $X_1$  represent a vector of household composition variables. A general Engel curve specification might have the form

$$Y_{1il} = G_l(Y_{2i}, X_{1i}) + \varepsilon_{il} \quad (2.3)$$

with

$$E(\varepsilon_{il}|Y_{2i}, X_{1i}) = 0 \text{ and } Var(\varepsilon_{il}|Y_{2i}, X_{1i}) = \sigma_l^2(Y_{2i}, X_{1i}). \quad (2.4)$$

When  $X_1$  is discrete, one approach to estimation would be to stratify by each distinct discrete outcome of  $X_1$  and estimate by nonparametric regression within each cell. Alternatively we may wish to pool Engel curves across household demographic types and allow the  $X_1$  to enter semiparametrically in each Engel curve. A simple semiparametric specification would be to assume additivity of  $G_l$

$$Y_{1il} = G_l(Y_{2i}, X_{1i}) + \varepsilon_{il} = h_l(Y_{2i}) + X'_{1i}\gamma_l + \varepsilon_{il} \quad (2.5)$$

in which  $\gamma_l$  represents a finite parameter vector of household composition effects for commodity  $l$  and  $h_l(Y_{2i})$  is some unknown function as in (2.1). This is the partially linear model of Robinson (1988). Although the partially linear model (2.5) motivates the approach taken in this paper, consideration of the integrability conditions indicates that some modification is required. Blundell, Browning and Crawford (2003) show that the additive structure underlying (2.5) together with the Slutsky symmetry conditions requires that  $h_l(Y_2)$  is linear in  $Y_2$  for all  $l$ .

An alternative specification for pooling across demographic types, and one that we adopt, is the following extension of the partially linear model

$$Y_{1il} = h_l(Y_{2i} - \phi(X'_{1i}\theta_1)) + X'_{1i}\theta_{2l} + \varepsilon_{il} \quad (2.6)$$

in which  $\phi(X'_{1i}\theta_1)$  is some known function up to a finite set of unknown parameters  $\theta_1$  and can be interpreted as the log of a general equivalence scale for household  $i$ , see e.g. Pendakur (1998).<sup>4</sup> For example, we may choose  $\phi(X'_{1i}\theta_1) = X'_{1i}\theta_1$  where  $X_{1i}$  is a vector of demographic variables representing different household types and  $\theta_1$  is

---

<sup>4</sup>This is nested within the fully nonparametric specification (2.3). Blundell, Duncan and Pendekar (1998) compare the specification used here with this more general alternative and find that it provides a good representation of demand behavior for households in the FES.

the vector of corresponding equivalence scales. Interestingly, the extended partially linear model (2.6) is precisely the shape invariant restricted specification considered in the work on pooling nonparametric regression curves by Härdle and Marron (1990) and Pinkse and Robinson (1995). Notice that (2.6) reduces to an additive form for functions of the demographic variables  $X_{1i}$  only when  $h_l$  is linear. This corresponds to the Translog or Almost Ideal model. For nonlinear specifications of  $h_l$ , including the QUAIDS specification, the theoretical consistency result tells us that the demographic terms must also enter in the function  $h_l$  as is the case for (2.6).

Two important assumptions will be required in estimation. First, the density of  $Y_{2i} - \phi(X'_{1i}\theta_1)$  is bounded away from zero at the true parameter value for  $\theta_1$ . The second critical assumption is that there is at least one good for which  $h$  is nonlinear.<sup>5</sup>

## 2.2. Endogeneity of Total Expenditure

There are both theoretical and empirical reasons why the total expenditure is likely to be endogenous in the sense that  $E[\varepsilon_{il}|Y_{2i}] \neq 0$ . Notice that the log of total expenditure  $Y_{2i}$  reflects savings and other consumption decisions made at the same time as the budget shares  $Y_{1il}$  are chosen. In fact the system of budget shares can be thought of as the second stage in a two-stage budgeting model in which total expenditure and savings are first determined and then, conditional on total expenditure, individual commodity shares are chosen at the second stage (see Blundell (1988), for example). In our application we use gross earnings of the head of household as an instrument  $X_{2i}$ . The gross earnings of the household head will be exogenous for consumption expenditures under the assumption that heterogeneity in earnings is not correlated with households' preferences over consumption.

A central objective of this paper is to relax the exogeneity assumption on  $Y_{2i}$  in the estimation of the semi-nonparametric budget share system (2.6). Blundell, Duncan and Pendakur (1998) have analyzed the parametric control function approach. In this paper, we consider the alternative nonparametric instrumental variables approach to solve the endogeneity problem. In particular, we consider semi-nonparametric IV estimation where  $h_l(\cdot)$  is a unknown function and  $\theta_1, \theta_{2l}$  are unknown finite-dimensional parameters.

---

<sup>5</sup>Nonlinear behavior in the Engel curve relationship is commonplace for many goods - see Banks, Blundell and Lewbel (1997), for example.

Functions of  $X_{2i}$  are then used as instrumental variables. More precisely we shall assume:

$$E[\varepsilon_{il}|X_{1i}, X_{2i}] = 0, \quad l = 1, \dots, L. \quad (2.7)$$

Next we examine the identification and estimation under this condition (2.7). This is followed by sections on the large sample properties of the estimator.

### 3. Identification and Estimation

#### 3.1. Notation

Here we first lay out the notations which will be adopted throughout the remaining discussion. Let  $\{(Y_{1il}, Y_{2i}, X_{1i}, X_{2i})\}_{i=1}^n$  represent an i.i.d. sequence of  $n$  household observations on the budget share  $Y_{1il}$  of good  $l = 1, \dots, L \geq 1$ ,<sup>6</sup> the log of total expenditure  $Y_{2i}$ , a vector of discrete household composition variables  $X_{1i}$ , and a vector of continuous instrumental variables  $X_{2i}$ . Denote  $Y_{1i} = (Y_{1i1}, \dots, Y_{1iL})' \in \mathcal{R}^L$ ,  $X_i = (X'_{1i}, X'_{2i})' \in \mathcal{X}$  with  $\dim(X_1), \dim(X_2) \geq 1$  and  $Z_i = (Y'_{1i}, Y_{2i}, X'_i)'$ . Also denote  $\rho = (\rho_1, \dots, \rho_L)' \in \mathcal{R}^L$ , where for  $l = 1, \dots, L$ ,

$$\rho_l(Z_i, \theta_1, \theta_{2l}, h_l) \equiv Y_{1il} - h_l(Y_{2i} - \phi(X'_{1i}\theta_1)) - X'_{1i}\theta_{2l}.$$

For each household  $i$  facing the same relative prices and for goods  $l = 1, \dots, L$ , the Engel curve model satisfies (2.6) and (2.7) which we rewrite as:

$$E[\rho_l(Z_i, \theta_{o1}, \theta_{o2l}, h_{ol})|X_i] = 0, \quad l = 1, \dots, L \quad (3.1)$$

where  $\phi$  is a known function,  $h_{ol}(\cdot), l = 1, \dots, L$  are true unknown real-valued functions,  $\theta_{o1}, \theta_{o2l}, l = 1, \dots, L$  are true unknown vector-valued finite-dimensional parameters. For policy analysis, we would like to estimate  $\theta_{o1}, \theta_{o2l}$ , the Engel curve function  $h_{ol}$  and other smooth functional of  $h_{ol}(\cdot)$  such as the average derivative  $E[\nabla h_{ol}(Y_{2i} - \phi(X'_{1i}\theta_{o1}))]$ .

Finally we denote  $\alpha = (\theta, h)$ ,  $\theta = (\theta'_1, \theta'_{21}, \dots, \theta'_{2L})' \in \Theta$ , a compact subset of  $\mathcal{R}^b$  with  $b = (1 + L)\dim(X_1)$ , and  $h = (h_1, \dots, h_L)' \in \mathcal{H}$ , a subset of space of functions that are square integrable against the probability measure associated with  $Y_2$  (to be specified later). Let  $\alpha_o = (\theta_o, h_o) \in \Theta \times \mathcal{H}$  denote the true parameters of interest. Then we can rewrite (3.1) as:

$$E[\rho(Z_i, \alpha_o)|X_i] = 0 \quad (3.2)$$

---

<sup>6</sup>Since budget shares should add up to one, total number of goods is actually  $L + 1$ . Provided the same basis functions are chosen to approximate  $h_l(Y_{2i} - \phi(X'_{1i}\theta_1)), l = 1, \dots, L$ , the estimators we derive will be invariant to the commodity omitted.

### 3.2. Identification

The first assumption is about identification of  $\alpha_o = (\theta_o, h_o)$ .

**Condition I (Identification):**

$$E[Y_{1il} - h_l(Y_{2i} - \phi(X'_{1i}\theta_1)) - X'_{1i}\theta_{2l}|X_{1i}, X_{2i}] = 0 \quad \text{for } l = 1, \dots, L,$$

$$\text{implies } \theta_1 = \theta_{o1}, \theta_{2l} = \theta_{o2l} \text{ and } h_l = h_{ol} \text{ a.s. for } l = 1, \dots, L.$$

We provide the following set of sufficient conditions, which might not be a minimal set of conditions but appear quite sensible for our Engel curve system application:

**Theorem 1.** *Suppose (3.1) and the followings hold: (1) for all bounded measurable functions  $\delta(Y_2, X_1)$ ,  $E[\delta(Y_2, X_1)|X_1, X_2] = 0$  implies  $\delta(Y_2, X_1) \equiv 0$  almost surely; (2) the conditional distribution of  $Y_2$  given  $(X_1, X_2)$  is absolutely continuous with respect to the Lebesgue measure on  $(-\infty, +\infty)$ ; (3)  $h_l$ ,  $l = 1, \dots, L$ , and  $\phi$  are bounded, differentiable, and cannot be simultaneously linear; (4)  $X_1$  is a vector of linearly independent, discrete random variables which only takes finite many values and does not contain constant one; (5) if  $X_1$  is a scalar dummy variable, then at least one  $h_l$  is not linear and  $\phi$  is not periodic. Then Condition I is satisfied.*

**Remark 1:** Condition (1) is equivalent to the *bounded completeness* in  $X_2$  of the conditional distribution of  $Y_2$  given  $X = (X_1, X_2)$ . Note that this is a weaker concept than the *completeness* in  $X_2$  of the conditional distribution of  $Y_2$  given  $X = (X_1, X_2)$  [which is equivalent to *for all measurable functions  $\delta(Y_2, X_1)$  with finite expectations,  $E[\delta(Y_2, X_1)|X_1, X_2] = 0$  implies  $\delta(Y_2, X_1) \equiv 0$  almost surely*]. By definition, completeness automatically implies bounded completeness. However there are examples in statistics literature showing that bounded completeness does not imply completeness, see e.g. Lehmann (1986, page 173), Hoeffding (1977) and Mattner (1993). There are not many known families of distributions beyond the exponential family are complete, while there are larger families of distributions are bounded complete. For instance, within the location family of absolutely continuous distributions (with respect to Lebesgue measure), they are bounded complete iff the characteristic functions are zero-free; while within its subclass of very thin tailed densities, the only complete class is either a Gaussian or a Dirac measure, see Mattner (1993). As an example, a family of nontrivial finite scale mixtures of the standard Gaussian  $N(0,1)$ -distribution is bounded complete but

not complete. For the identification of  $h$  in a purely nonparametric IV regression model  $E[Y_1 - h(Y_2, X_1)|X_1, X_2] = 0$ , Newey and Powell (2003) and Darolles, Florens and Renault (2002) impose the “completeness” condition. Here, since Engel curves should be all bounded below by zero and above by one, it suffices to impose the weaker “bounded completeness” condition for the Engel curve application.

In the empirical application (Section 7), we take  $X_{1i} = 1$  or 0 to indicate if the  $i$ -th family has kids or not, while  $X_{2i}$  is a one-to-one transformation of log gross earnings of the household head. In our data set and in many other empirical Engel curve analyses, the estimated joint density of log-total expenditure and log-gross earnings is approximately bivariate normal with high correlation coefficient.

### 3.3. Estimation Procedure

Our estimation method is similar to that in Newey and Powell (2003) for nonparametric IV regression, and in Ai and Chen (2003) for semiparametric conditional moment restrictions. First we approximate the unknown functions  $h \in \mathcal{H} \equiv \mathcal{H}_1 \times \cdots \times \mathcal{H}_L$  by  $h_n \in \mathcal{H}_n$ , where  $\mathcal{H}_n \equiv \mathcal{H}_{1,n} \times \cdots \times \mathcal{H}_{L,n}$  is some sieve space, that is, some finite-dimensional approximation spaces (e.g. Fourier series, orthogonal polynomials, splines, power series, wavelets, etc.) which become dense in  $\mathcal{H}$  as sample size  $n \rightarrow \infty$ . Then for arbitrarily fixed candidate value  $\alpha = (\theta, h_n) \in \mathcal{A}_n \equiv \Theta \times \mathcal{H}_n$ , we estimate the population conditional moment function  $m(x, \alpha) \equiv E[\rho(Z_i, \alpha)|X_i = x]$  nonparametrically by  $\hat{m}(x, \alpha)$ . Finally we estimate the  $\theta$  and the unknown sieve coefficients jointly by a generalized version of minimal distance estimation procedure:

$$\min_{\alpha \in \mathcal{A}_n} \frac{1}{n} \sum_{i=1}^n \hat{m}(X_i, \alpha)' [\hat{\Sigma}(X_i)]^{-1} \hat{m}(X_i, \alpha), \quad (3.3)$$

where  $\hat{\Sigma}$  is some consistent estimator of some positive definite weighting matrix  $\Sigma$ . We denote the resulting estimator as  $\hat{\alpha}_n = (\hat{\theta}_n, \hat{h}_n) \in \mathcal{A}_n$ .

To obtain semiparametric efficient estimator of  $\theta_o$ , we may follow a three-step procedure proposed in the first version of Ai and Chen (2003):

**Step 1.** Obtain an initial consistent estimator  $\hat{\alpha}_n$  as the solution to:

$$\min_{\alpha \in \mathcal{A}_n} \frac{1}{n} \sum_{i=1}^n \hat{m}(X_i, \alpha)' \hat{m}(X_i, \alpha), \quad (3.4)$$

where  $\widehat{m}(X_i, \alpha)$  is a nonparametric consistent estimator of  $m(X_i, \alpha)$  uniformly over  $(X_i, \alpha) \in \mathcal{X} \times \mathcal{A}_n$ .

**Step 2.** Obtain a consistent estimator  $\widehat{\Sigma}_o(X)$  of the optimal weighting matrix  $\Sigma_o(X) \equiv \text{Var}[\rho(Z, \alpha_o)|X]$  using  $\widehat{\alpha}_n$  and any nonparametric regression procedures (such as kernel, nearest-neighbor or linear sieves).

**Step 3.** Obtain the optimally weighted estimator  $\widetilde{\alpha}_n = (\widetilde{\theta}_n, \widetilde{h}_n)$  by solving

$$\min_{\alpha \in \mathcal{A}_n} \frac{1}{n} \sum_{i=1}^n \widehat{m}(X_i, \alpha)' [\widehat{\Sigma}_o(X_i)]^{-1} \widehat{m}(X_i, \alpha). \quad (3.5)$$

**Remark 2:** If total expenditure  $Y_2$  is assumed to be exogenous, then  $Y_2$  should be the “perfect IV” and we have  $E[\rho(Z_i, \theta_o, h_o)|X_{1i}, Y_{2i}] = 0$ . In this case, we do not need to estimate  $\alpha_o = (\theta_o, h_o)$  via (3.3). Instead a simpler estimation method will be the following sieve generalized Least Squares (LS) method:

$$\min_{\alpha \in \mathcal{A}_n} \frac{1}{n} \sum_{i=1}^n \rho(Z_i, \alpha)' \widehat{\Sigma}(X_i)^{-1} \rho(Z_i, \alpha). \quad (3.6)$$

Again the semiparametric efficient estimator of  $\theta_o$  can be obtained by the above three-step procedure except with  $\widehat{m}(X_i, \alpha)$  replaced by  $\rho(Z_i, \alpha)$ .

**Remark 3:** An alternative estimation procedure is the sieve profile approach. For each fixed  $\theta \in \Theta$ , we denote  $h^*(\theta; \cdot)$  as the solution to

$$\min_{h \in \mathcal{H}} E [m(X, \theta, h)' \Sigma(X)^{-1} m(X, \theta, h)].$$

Clearly  $h_o(\cdot) = h^*(\theta_o; \cdot)$ . We can apply the following profile estimation procedure: First for each fixed  $\theta \in \Theta$ , we estimate  $h^*(\theta; \cdot)$  by  $\widehat{h}_n(\theta; \cdot)$ , the solution to

$$\min_{h \in \mathcal{H}_n} \frac{1}{n} \sum_{i=1}^n \widehat{m}(X_i, \theta, h)' [\widehat{\Sigma}(X_i)]^{-1} \widehat{m}(X_i, \theta, h). \quad (3.7)$$

Second we estimate  $\theta_o$  by  $\widehat{\theta}_n$  which solves

$$\min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n \widehat{m}(X_i, \theta, \widehat{h}_n(\theta; \cdot))' [\widehat{\Sigma}(X_i)]^{-1} \widehat{m}(X_i, \theta, \widehat{h}_n(\theta; \cdot)), \quad (3.8)$$

and estimate  $h_o(\cdot) = h^*(\theta_o; \cdot)$  by  $\widehat{h}_n = \widehat{h}_n(\widehat{\theta}_n; \cdot)$ .

If total expenditure  $Y_2$  is assumed to be exogenous, then we can again estimate  $\alpha_o = (\theta_o, h_o)$  via the following profile sieve generalized Least Squares (LS):

$$\begin{aligned}\widehat{h}_n(\theta; \cdot) &= \arg \min_{h \in \mathcal{H}_n} \frac{1}{n} \sum_{i=1}^n \rho(Z_i, \theta, h)' \widehat{\Sigma}(X_i)^{-1} \rho(Z_i, \theta, h), \\ \widehat{\theta}_n &= \arg \min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n \rho(Z_i, \theta, \widehat{h}_n(\theta; \cdot))' \widehat{\Sigma}(X_i)^{-1} \rho(Z_i, \theta, \widehat{h}_n(\theta; \cdot)).\end{aligned}$$

### 3.4. Possible Sieve Bases for $h$

There are many sieve spaces which can approximate  $\mathcal{H}$  well. Since  $h_{ol}$ ,  $l = 1, \dots, L$  have the same argument  $Y_2 - \phi(X_1' \theta_1)$  and similar smoothness, they may be approximated by the same kind of sieve bases. In our empirical application  $Y_2$  is log total expenditure, and a simple nonparametric estimation of the density of  $Y_2$  using our data set shows that it could be closely approximated by a normal density. Therefore we assume that the support of  $Y_2 - \phi(X_1' \theta_1)$  is the entire real line  $\mathcal{R}$ . Then the choice of sieve bases are partially suggested by what kind of smoothness we want to impose on  $h_o \in \mathcal{H}$ .

Since consumer demand theory and many empirical studies suggest that  $h_{ol}$ ,  $l = 1, \dots, L$  are sup-norm bounded (actually  $0 \leq h_{ol} \leq 1$  and  $0 \leq \sum_{l=1}^L h_{ol} \leq 1$ ), and are reasonably smooth curves, one possible assumption is that  $h_{ol} \in \mathcal{H}_l$  for all  $l = 1, \dots, L$ , where

$$\mathcal{H}_l = \left\{ h_{ol} \in W_\infty^r(\mathcal{R}) : 0 \leq h_{ol} \leq 1, \|h_{ol}\|_{W_\infty^r} \leq c < \infty \right\}, \quad (3.9)$$

where  $c$  is a known constant (say 1.5),  $W_\infty^r(\mathcal{R})$  denotes the Sobolev space of smoothness  $r \geq 1$ . When  $r$  is an integer, the norm  $\|g\|_{W_\infty^r}$  is equivalent to  $\|g\|_\infty + \|\nabla^r g\|_\infty$ , here  $\|\cdot\|_\infty$  denotes the sup-norm  $\|g\|_\infty \equiv \sup_y |g(y)|$  and  $\nabla^r g$  denotes the  $r$ -th derivative of  $g$ . An attractive property of  $W_p^r(\mathcal{R})$ ,  $1 \leq p \leq \infty$  is that any function  $g$  in this space can be represented as linear combinations of a frame basis  $\{\psi_{kj}\}$  such that the norm  $\|g\|_{W_p^r}$  is equivalent to the sum of  $\ell_p$ -norms on the level and on the  $r$ -th order difference of the coefficients sequences, (see e.g. Meyer, 1990).

For the space (3.9), we may consider the following sieve:

$$\mathcal{H}_{l,n} = \left\{ \begin{array}{l} h_{l,n}(Y_2 - \phi(X_1' \theta_1)) = \sum_{k=0}^{K_n} \sum_{j \in \mathcal{K}_n} \pi_{kj}^l \psi_{kj}(Y_2 - \phi(X_1' \theta_1)) : \\ 0 \leq h_{l,n}(\cdot) \leq 1, \|\nabla^r h_{l,n}\|_\infty \leq c \end{array} \right\} \quad (3.10)$$

where

$$\psi_{kj}(Y_2 - \phi(X_1' \theta_1)) = 2^{k/2} B_\gamma(2^k [Y_2 - \phi(X_1' \theta_1)] - j),$$

and  $B_\gamma(\cdot)$  denotes the cardinal B-spline of order  $\gamma = [r] + 1$ :<sup>7</sup>

$$B_\gamma(y) = \frac{1}{(\gamma - 1)!} \sum_{i=0}^{\gamma} (-1)^i \binom{\gamma}{i} [\max(0, y - i)]^{\gamma-1}, \quad (3.11)$$

which is a piecewise polynomial of highest degree  $\gamma - 1$ . Obviously  $B_\gamma(\cdot)$  is  $\gamma - 1$  times differentiable, has support on  $[0, \gamma]$  and is symmetric about the center of its support. In addition, it satisfies

$$B_\gamma(\cdot) \geq 0, \quad \sum_{j=-\infty}^{+\infty} B_\gamma(y - j) = 1 \text{ for all } y, \quad (3.12)$$

which is crucial to preserve the shape of the approximated unknown functions, see Chui (1992, Chapter 4) and Anastassiou and Yu (1992) for details. Moreover, the compact support  $[0, \gamma]$  of  $B_\gamma(\cdot)$  makes the summation  $\sum_{k=0}^{K_n} \sum_{j \in \mathcal{K}_n}$  in (3.10) becomes effectively summation over finite many terms for any fixed  $K_n$ .<sup>8</sup> The approximation accuracy of this sieve can be found in Chen, Hansen and Scheinkman (1997) when the support of the function is the entire real line.

### 3.5. Possible Sieve Bases for $m$

There are many nonparametric procedures such as kernel, local linear regression, nearest neighbor and various sieve methods that can be used to estimate  $m(x, \alpha)$  and  $\Sigma_o(x)$ . Here we again consider the sieve estimator as illustration. For each fixed  $(X_i, \alpha)$ , we approximate  $E[\rho(Z_i, \alpha) | X_i] = m(X_i, \alpha)$  by  $m(X_i, \alpha) \approx \sum_{j \in \mathcal{J}_n} a_j(\alpha) p_{oj}(X_i)$ , where  $p_{oj}$  some known fixed basis functions, and  $J_n \equiv \#(\mathcal{J}_n) \rightarrow \infty$  slowly as  $n \rightarrow \infty$ . We then estimate the sieve coefficients  $\{a_j\}$  simply by LS regression:

$$\min_{\{a_j\}} \frac{1}{n} \sum_{i=1}^n [\rho(Z_i, \alpha) - \sum_{j \in \mathcal{J}_n} a_j p_{oj}(X_i)]' [\rho(Z_i, \alpha) - \sum_{j \in \mathcal{J}_n} a_j p_{oj}(X_i)],$$

and the resulting estimator is denoted as:  $\hat{m}(X, \alpha) = \sum_{j \in \mathcal{J}_n} \hat{a}_j(\alpha) p_{oj}(X)$ . In the following we denote:  $p^{J_n}(x) = (p_{o1}(x), \dots, p_{oJ_n}(x))'$  and  $P = (p^{J_n}(X_1), \dots, p^{J_n}(X_n))'$ , then:

$$\hat{m}(X, \alpha) = \sum_{i=1}^n \rho(Z_i, \alpha) p^{J_n}(X_i)' (P'P)^{-1} p^{J_n}(X). \quad (3.13)$$

<sup>7</sup>The notation  $[r]$  means the integer part of  $r$ , where  $r \geq 1$  denotes the smoothness of the engel curve function  $h_i$ .

<sup>8</sup>See Chen and Conley (2001) for an application of this sieve to estimate a spatial panel time series model.

Again many known sieve bases could be used as  $\{p_{oj}\}$ . In our empirical application to Engel curve estimation, we take  $X_1$  to be the set of discrete random variables which has finite support  $\mathcal{X}_1$ , while  $X_2$  is the normal transformation of the log of gross earnings<sup>9</sup>  $X_2 = \Phi(\log\text{-gross earnings}) \in [0, 1]$ . We have taken  $p^{J_n}(x)$  to be Fourier series and B-splines. The empirical findings are not sensitive to the different choice of sieve bases.

#### 4. Consistency and Convergence Rate of Nonparametric Parts

In this section we focus on the properties of the sieve IV estimators of the unknown Engel curve  $h_{ol}, l = 1, \dots, L$ . We first provide the consistency of  $\hat{h}_l$  under a strong norm, then establish the convergence rate of  $\hat{h}_l$  under the mean squared error metric:  $\|\hat{h} - h_o\|_{Y_2}^2 = \max_{l=1, \dots, L} \|\hat{h}_l - h_{ol}\|_{Y_2}^2$  with

$$\|\hat{h}_l - h_{ol}\|_{Y_2}^2 = E \left[ \left\{ \hat{h}_l(Y_2 - \phi(X_1' \theta_{o1})) - h_{ol}(Y_2 - \phi(X_1' \theta_{o1})) \right\}^2 \right].$$

In this section we take  $\hat{\Sigma}(X) = I_L$  without loss of generality. We denote  $k_n \equiv 2^{K_n} = \dim(\mathcal{H}_{l,n})$ , and assume:

**Assumption 1.** (i) The data  $\{Z_i = (Y_i', X_i')' : i = 1, 2, \dots, n\}$  are i.i.d.; (ii)  $0 \leq Y_{1il} \leq 1$  for  $l = 1, \dots, L$ ; (iii) conditions of theorem 1 hold.

**Assumption 2.** (i)  $\theta_o = (\theta'_{o1}, \theta'_{o21}, \dots, \theta'_{o2L})' \in \Theta$ , a compact subset of  $\mathcal{R}^b$  with  $b = (1 + L) \dim(X_1)$ ; (ii)  $h_o = (h_{o1}, \dots, h_{oL})' \in \mathcal{H}$  with  $h_{ol} \in W_\infty^r(\mathcal{R})$ ,  $r > 1$ , and  $0 \leq h_{ol} \leq 1$  and  $0 \leq \sum_{l=1}^L h_{ol} \leq 1$ ,  $\sum_{l=1}^L \|h_{ol}\|_{W_\infty^r} \leq c < \infty$ .

**Assumption 3.** For any  $x_1$  in the support of  $X_1$ , (i)  $p^{J_n}(x) = (p_{o1}(x), \dots, p_{oJ_n}(x))'$  is either a Fourier series or a B-spline basis of order  $[r_m] + 1$  for functions in  $W_\infty^{r_m}(\mathcal{X}_2)$ ,  $r_m > 1/2$ ; (ii) the smallest and the largest eigenvalues of  $E\{p^{J_n}(X)p^{J_n}(X)'\}$  are bounded and bounded away from zero for each  $J_n$ ; (iii) the density of  $X_2$  is bounded and bounded away from zero over its support  $\mathcal{X}_2$ , which is a compact interval with non-empty interior.

**Assumption 4.** (i) For any  $x_1$  in the support of  $X_1$ ,  $E[Y_{1l}|X_1 = x_1, X_2]$ ,  $l = 1, \dots, L$ , and  $E[\psi_{kj}(Y_2 - \phi(X_1' \theta_1))|X_1 = x_1, X_2]$ ,  $k = 0, \dots, K_n$ ,  $j \in \mathcal{K}_n$  all belong to  $W_\infty^{r_m}(\mathcal{X}_2)$ ,  $r_m > 1/2$ .

**Assumption 5.** (i)  $k_n \equiv 2^{K_n} \rightarrow \infty$ ,  $J_n/n \rightarrow 0$ ,  $J_n \geq (1 + L^{-1}) \dim(X_1) + k_n$ .

---

<sup>9</sup>Of course any bounded monotone one-to-one transformation will be fine. However, the empirical distribution of the log-gross wage is known to be well approximated by a normal cdf  $\Phi$ , we take this transformation so that the resulting  $X_{2i}$  is approximatedly uniformly distributed over  $[0, 1]$ .

The following consistency result is obtained by applying Theorem 4.1 in Newey and Powell (2003).

**Proposition 1.** *Let  $Y_2$  be endogenous and  $E\{|Y_2|^{2a} | X\}$  be bounded for some  $a > 0$ . Suppose Assumptions 1 - 3, 4(i) and 5(i) are satisfied. Let  $\hat{\alpha}$  be the sieve MD estimator (3.4) with the sieve space given by (3.10). Then:  $\|(\hat{h} - h_o)(1 + y^2)^{-a/2}\|_{W_\infty^{r_1}(\mathcal{R})} = o_p(1)$  for any  $r_1 \in [0, r)$ , hence  $\|\hat{h} - h_o\|_{Y_2} = o_p(1)$ .*

In order to obtain the rate of convergence for  $\|\hat{h} - h_o\|_{Y_2}$ , we impose the following additional conditions:

**Assumption 4.** (ii) There is a finite constant  $c$  such that for any  $x_1$  in the support of  $X_1$ ,  $\|E[Y_{1l}|X_1 = x_1, X_2 = \cdot]\|_{W_\infty^{r_m}} \leq c$  for  $l = 1, \dots, L$ , and  $\|E[\psi_{kj}(Y_2 - \phi(X_1'\theta_1))|X_1 = x_1, X_2 = \cdot]\|_{W_\infty^{r_m}} \leq c$  for  $k = 0, \dots, K_n, j \in \mathcal{K}_n$ .

**Assumption 5.** (ii)  $\lim_{n \rightarrow \infty} (J_n/k_n) = c_0 \geq 1$ ;  $\lim_{n \rightarrow \infty} (k_n^2/n) = 0$ ;

**Assumption 6.** (i) Conditioning on  $X_1 = 0$ , the joint probability measure of  $(X_2, Y_2)$  is absolutely continuous with respect to the product probability measure of  $X_2$  and  $Y_2$ ;

We define  $\tau_n$  as a *sieve measure of ill-posedness*:

$$\tau_n \equiv \sup_{h_n \in \mathcal{H}_n: h_n \neq 0} \frac{\sqrt{E\{h_n(Y_2)\}^2}}{\sqrt{E\{E[h_n(Y_2)|X_1 = 0, X_2]\}^2}}, \quad (4.1)$$

which is well defined under the conditions for identification. Obviously  $\tau_n \geq 1$ , and  $\tau_n = 1$  if and only if  $Y_2$  is measurable with respect to the sigma-field generated by  $\{X_1 = 0, X_2\}$  (then  $E[h_n(Y_2)|X_1 = 0, X_2] = h_n(Y_2)$  for all  $h_n \in \mathcal{H}_n$ ). For example  $\tau_n = 1$  when  $Y_2$  is exogenous (and we take  $Y_2 = X_2$ ). We note that the  $\tau_n$  measure of ill-posedness, as given in (4.1), depends on the choice of sieve space  $\mathcal{H}_n$ . This is why we call it a ‘‘sieve measure of ill-posedness’’.<sup>10</sup>

Under Assumption 6(i),  $\tau_n$  is closely related to the  $k_n$ -th singular number associated with the conditional expectation operator  $(Th)(X_2) \equiv E[h(Y_2)|X_2, X_1 = 0]$ .

<sup>10</sup>The sieve measure of ill-posedness,  $\tau_n$ , can be estimated from the data by

$$\hat{\tau}_n \equiv \sup_{h_n \in \mathcal{H}_n: h_n \neq 0} \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n \{h_n(Y_{2i})\}^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n \{\hat{E}[h_n(Y_2)|X_{1i} = 0, X_{2i}]\}^2}}, \quad (4.2)$$

where for any fixed  $h_n \in \mathcal{H}_n$ ,  $\hat{E}[h_n(Y_2)|X_1 = 0, X_2]$  is a nonparametric estimate of the conditional expectation  $E[h_n(Y_2)|X_1 = 0, X_2]$  such as a sieve LS estimator using the sieve basis function  $p^{J_n}(X_2)$ , see Section 6 for more details.

In the following we let  $f_{0,X_2,Y_2}$ ,  $f_{0,X_2}$ ,  $f_{0,Y_2}$  denote respectively the joint density of  $(X_2, Y_2)$ , marginal densities of  $X_2$  and  $Y_2$ , all conditioning on  $X_1 = 0$ . Let  $T : W_\infty^r(\mathcal{R}) \cap L_2(\mathcal{R}, f_{0,Y_2}) \rightarrow W_\infty^{r_m}(\mathcal{X}_2) \cap L_2(\mathcal{X}_2, f_{0,X_2})$  denote the conditional expectation operator that maps a square integrable function of  $Y_2$  in  $W_\infty^r(\mathcal{R})$  into a square integrable function of  $X_2$  in  $W_\infty^{r_m}(\mathcal{X}_2)$ , and let  $T^* : W_\infty^{r_m}(\mathcal{X}_2) \cap L_2(\mathcal{X}_2, f_{0,X_2}) \rightarrow W_\infty^r(\mathcal{R}) \cap L_2(\mathcal{R}, f_{0,Y_2})$  denote the adjoint operator of  $T$ , which is the conditional expectation operator mapping a square integrable function of  $X_2$  into a square integrable function of  $Y_2$ , i.e.,  $(T^*g)(Y_2) \equiv E[g(X_2)|Y_2, X_1 = 0]$ .

Under Assumption 6(i), it is well-known that the self-adjoint operators  $T^*T$  and  $TT^*$  share the same eigenvalue sequence  $\{\mu_k^2\}$  with  $\mu_1^2 = 1 \geq \mu_2^2 \geq \mu_3^2 \geq \dots$ . The non-negative value of  $\sqrt{\mu_k^2}$  is called the  $k$ -th singular number of  $T$  and  $T^*$ , which will be denoted as  $\mu_k$  to save notation in this paper. Moreover, the corresponding eigenfunctions of the operators  $T^*T$  and  $TT^*$  are also orthonormal bases  $\{\phi_{1k} : k = 1, \dots, \infty\}$  (for  $W_\infty^r(\mathcal{R})$  and  $L_2(\mathcal{R}, f_{0,Y_2})$ ),  $\{\phi_{0k} : k = 1, \dots, \infty\}$  (for  $W_\infty^{r_m}(\mathcal{X}_2)$  and  $L_2(\mathcal{X}_2, f_{0,X_2})$ ), such that

$$\frac{f_{0,X_2,Y_2}(x_2, y_2)}{f_{0,X_2}(x_2)f_{0,Y_2}(y_2)} = \sum_{k=1}^{\infty} \mu_k \phi_{0k}(x_2) \phi_{1k}(y_2), \quad \sum_{k=1}^{\infty} \mu_k^2 < \infty,$$

and for all  $k \geq 1$ ,

$$T\phi_{1k} = \mu_k \phi_{0k}, \quad T^*\phi_{0k} = \mu_k \phi_{1k}; \quad T^*T\phi_{1k} = \mu_k^2 \phi_{1k}, \quad TT^*\phi_{0k} = \mu_k^2 \phi_{0k}.$$

Denote  $\|g\|_{X_2} \equiv \sqrt{E\{g(X_2)\}^2}$  and  $\|h\|_{Y_2} \equiv \sqrt{E\{h(Y_2)\}^2}$ . In the Appendix A, we establish that

$$\tau_n \equiv \sup_{h_n \in \mathcal{H}_n: h_n \neq 0} \frac{\|h_n\|_{Y_2}}{\|Th_n\|_{X_2}} \leq \frac{1}{\mu_{k_n}}. \quad (4.3)$$

**Theorem 2.** *Let  $Y_2$  be endogenous with  $E[|Y_2|^{2a}|X] < \infty$ ,  $E[|Y_2|^{pa}] < \infty$  for some  $a > r$  and  $p \geq 4$ . Suppose Assumptions 1 - 4, 5(i)(ii) and 6(i) are satisfied. Let  $\hat{\alpha}$  be the sieve MD estimator (3.4) with the sieve space given by (3.10). Then*

$$\|\hat{h}_{nl} - h_{ol}\|_{Y_2} = O_p \left\{ (k_n)^{-r} + \tau_n \times \left[ \sqrt{\frac{k_n}{n}} + (k_n)^{-r_m} \right] \right\} \quad \text{for all } l = 1, \dots, L;$$

Under  $k_n = O(n^{1/(2r_m+1)})$ , we have

$$\|\hat{h}_{nl} - h_{ol}\|_{Y_2} = O_p \left\{ (k_n)^{-r} + \tau_n \sqrt{\frac{k_n}{n}} \right\} = O_p \left\{ [(k_n)^{r_m-r} + \tau_n] \times n^{-\frac{r_m}{2r_m+1}} \right\}.$$

**Remark 4:** (i) For exogenous total expenditure  $Y_2$ , under Assumptions 1 and 2 and some mild additional ones, we have  $\|\widehat{h}_{nl} - h_{ol}\|_{Y_2} = O_p \left\{ (k_n)^{-r} + \sqrt{\frac{k_n}{n}} \right\}$ .<sup>11</sup> Comparing this rate to that in Theorem 2 (the endogenous  $Y_2$  case), we note that the bias part  $(k_n)^{-r}$  is of the same order, however, the standard deviation part blows up from  $\sqrt{\frac{k_n}{n}}$  in the exogenous case to  $\tau_n \sqrt{\frac{k_n}{n}}$  in the endogenous case.

(ii) Under Assumptions 1 - 4, 5(ii) and  $J_n = O(n^{1/(2r_m+1)})$ , Claim 2 in Appendix A shows that

$$\sqrt{E\{\widehat{E}[h_{ol}|X_2] - E[h_{ol}|X_2]\}^2} = O_p \left\{ n^{-\frac{r_m}{2r_m+1}} \right\}.$$

Therefore we have

$$\|\widehat{h}_{nl} - h_{ol}\|_{Y_2} = O_p \left\{ ((k_n)^{r_m-r} + \tau_n) \sqrt{E\{\widehat{E}[h_{ol}|X_2] - E[h_{ol}|X_2]\}^2} \right\}.$$

Recall that  $h_{ol}$  has  $r$ -smoothness and  $E[h_{ol}(Y_2 - \phi(0))|X_1 = 0, X_2] = E[Y_{1l}|X_1 = 0, X_2]$  has  $r_m$ -smoothness, and that  $r_m > r$ ,  $\tau_n > 1$  for endogenous  $Y_2$ . The convergence rate of  $\|\widehat{h}_{nl} - h_{ol}\|_{Y_2}$  is always slower than that of  $\sqrt{E\{\widehat{E}[h_{ol}|X_2] - E[h_{ol}|X_2]\}^2}$  by a factor  $(k_n)^{r_m-r} + \tau_n$ . Hence the difference “ $r_m - r$ ” can be regarded as a measure of the *exponent degree of ill-posedness*. Obviously  $\|\widehat{h}_{nl} - h_{ol}\|_{Y_2}$  will go to zero very slowly when  $r_m - r$  is very big.

The inequality (4.3) is very useful, since there exist many approximation results on the behavior of singular value  $\mu_{k_n}$  in terms of smoothness and integrability of the kernel function  $\frac{f_{0,X_2,Y_2}(x_2,y_2)}{f_{0,X_2}(x_2)f_{0,Y_2}(y_2)}$  of the conditional expectation operator  $T$ , see e.g. Birman and Solomyak (1977) and Chen, Hansen and Scheinkman (2000). We now make one of such sufficient conditions.

**Assumption 6.** (ii) for each fixed  $y_2$ ,  $\frac{f_{0,X_2,Y_2}(\cdot,y_2)}{f_{0,X_2}(\cdot)f_{0,Y_2}(y_2)}$  belongs to  $W_\infty^{\gamma_0}(\mathcal{X}_2)$ ,  $0 < \gamma_0 = r_m - r - \frac{1}{2}$ , and

$$\int_{\mathcal{R}} \left[ \left\| \frac{f_{0,X_2,Y_2}(\cdot,y_2)}{f_{0,X_2}(\cdot)f_{0,Y_2}(y_2)} \right\|_{W_\infty^{\gamma_0}(\mathcal{X}_2)} \right]^2 f_{0,Y_2}(y_2) dy_2 < \infty.$$

Assumption 6(ii) implies

$$c_1 k^{-(\gamma_0 + \frac{1}{2})} [L(k)]^{-1} \leq \mu_k \leq c k^{-(\gamma_0 + \frac{1}{2})} \equiv c k^{-(r_m - r)},$$

---

<sup>11</sup>The convergence rate under exogenous  $Y_2$  can be obtained from Theorem 1 in Newey (1997) or Theorem 1 in Chen and Shen (1998).

for some slow varying function  $L(k)$  that goes to  $\infty$  slower than any polynomial order, see e.g. Birman and Solomjak (1977). Assumption 6, the inequality (4.3) and Theorem 2 imply the following corollary.

**Corollary 1.** *Suppose all the conditions of Theorem 2 and Assumption 6(ii) hold. Let  $\widehat{\alpha}$  be the sieve MD estimator (3.4) with the sieve space given by (3.10). Let  $k_n = O(n^{1/(2r_m+1)})$ . Then for all  $l = 1, \dots, L$ ,*

$$\|\widehat{h}_l - h_{ol}\|_{Y_2} = O_p(n^{-r/(2r_m+1)} L(n^{1/(2r_m+1)}))$$

for some function  $L(k)$  goes to  $\infty$  slowly such that  $L(k)/k^\epsilon \rightarrow 0$  for all  $\epsilon > 0$ .

**Remark 5:** If we assume that  $Y_2$  has bounded interval support say  $[0, 1]$ , then conditions very similar to Assumption 6(ii) will provide a precise order of singular value decay rate. For example, suppose Assumption 6(ii) is replaced by Assumption 6(ii)':

**Assumption 6(ii)'**:  $Y_2$  has support  $[0, 1]$ , and for each fixed  $y_2 \in [0, 1]$ ,  $\frac{f_{0, X_2, Y_2}(\cdot, y_2)}{f_{0, X_2}(\cdot) f_{0, Y_2}(y_2)}$  belongs to  $C^{\gamma_0}(\mathcal{X}_2)$ ,  $\gamma_0 = r_m - r - \frac{1}{2} \in (0, 1]$ , and

$$\sup_{y_2 \in [0, 1]} \left\| \frac{f_{0, X_2, Y_2}(\cdot, y_2)}{f_{0, X_2}(\cdot) f_{0, Y_2}(y_2)} \right\|_{C^{\gamma_0}(\mathcal{X}_2)} < \infty.$$

Then we have:

$$c_1 k^{-(\gamma_0 + \frac{1}{2})} \leq \mu_k \leq c k^{-(\gamma_0 + \frac{1}{2})} \equiv c k^{-(r_m - r)},$$

see e.g. Heinrich and Kuhn (1985, corollary). This, the inequality (4.3) and Theorem 2 together imply

$$\|\widehat{h}_l - h_{ol}\|_{Y_2} = O_p(n^{-r/(2r_m+1)})$$

which coincides with the optimal rate established in Hall and Horowitz (2003) when  $Y_2$  has support  $[0, 1]$ .

## 5. $\sqrt{n}$ Asymptotic Normality and Efficiency of $\widehat{\theta}$

This section specializes the general theory of Ai and Chen (2003) to our system of shape invariant Engel curve models by providing relatively low-level sufficient conditions. We first establish a faster than  $n^{-1/4}$  rate of convergence under a weaker metric  $\|\alpha - \alpha_o\|$ , which is crucial to establish  $\sqrt{n}$ -asymptotic normality of estimates of  $\theta_o$ . We then provide sufficient conditions for  $\sqrt{n}$ -asymptotic normality and efficiency of  $\widehat{\theta}$ .

### 5.1. Nonparametric Convergence Rate under a Weaker Metric

We define the following metric:

$$\|\alpha - \alpha_o\|^2 = E \left[ E \left\{ \frac{d\rho(Z, \alpha_o)}{d\alpha} [\alpha - \alpha_o] | X \right\}' \Sigma(X)^{-1} E \left\{ \frac{d\rho(Z, \alpha_o)}{d\alpha} [\alpha - \alpha_o] | X \right\} \right],$$

where  $\frac{d\rho(Z, \alpha)}{d\alpha} [v]$  denotes the directional derivatives of  $\rho$  at the direction  $v = (v_\theta, v_h)$  evaluated at  $\alpha$ :

$$\frac{d\rho(Z, \alpha)}{d\alpha} [v] = \left( \frac{d\rho_1(Z, \alpha)}{d\alpha} [v], \dots, \frac{d\rho_L(Z, \alpha)}{d\alpha} [v] \right)' = \frac{d\rho(Z, \alpha)}{d\theta'} (v_\theta) + \frac{d\rho(Z, \alpha)}{dh} [v_h],$$

here  $\frac{d\rho(Z, \alpha)}{d\theta'}$  is the ordinary derivative with respect to  $\theta' = (\theta'_1, \theta'_{21}, \dots, \theta'_{2L})$  evaluated at  $\alpha$ ,  $\frac{d\rho(Z, \alpha)}{dh} [v_h]$  is the directional derivative with respect to  $h$  at direction  $v_h$  evaluated at  $\alpha$ . Notice that in this paper, for all  $l = 1, \dots, L$ ,

$$\frac{d\rho_l(Z, \alpha_o)}{d\alpha} [v] = \nabla h_{ol}(Y_2 - \phi(X_1' \theta_{o1})) \times \nabla \phi(X_1' \theta_{o1}) \times X_1' [v_{\theta_1}] - X_1' [v_{\theta_{2l}}] - [v_{h_l}(Y_2 - \phi(X_1' \theta_{o1}))].$$

We impose the following conditions:

**Assumption 5.** (iii)  $\frac{(k_n)^2 \ln(n)}{\sqrt{n}} \rightarrow 0$ ; (iv)  $(k_n)^{-r} = O(n^{-1/4})$ ;  $(k_n)^{-rm} = o(n^{-1/4})$ .

**Assumption 7.** uniformly over  $X \in \mathcal{X}$ , (i)  $\widehat{\Sigma}(X) = \Sigma(X) + o_p(n^{-1/4})$ ; (ii)  $\Sigma(X)$  is finite positive definite.

The next result can be obtained by applying Theorem 3.1 in Ai and Chen (2003).

**Proposition 2.** Let  $Y_2$  be endogenous with  $E[|Y_2|^{4a}] < \infty$  for some  $a > r$ . Suppose Assumptions 1 - 5, 6(i) and 7 are satisfied. Let  $\widehat{\alpha}$  be the sieve MD estimator (3.3) with the sieve space given by (3.10). Then  $\|\widehat{\alpha} - \alpha_o\| = o_p(n^{-1/4})$ .

### 5.2. $\sqrt{n}$ Asymptotic Normality

Before we state the result on asymptotic normality, we need to introduce additional notations. Let

$$D_{w^1}(X, \alpha_o) = E \left( \nabla h_o(Y_2 - \phi(X_1' \theta_{o1})) \nabla \phi(X_1' \theta_{o1}) X_1' + w^1(Y_2 - \phi(X_1' \theta_{o1})) \mid X \right)$$

where

$$\nabla h_o(Y_2 - \phi(X_1' \theta_{o1})) \equiv \begin{bmatrix} \nabla h_{o1}(Y_2 - \phi(X_1' \theta_{o1})) \\ \dots \\ \nabla h_{oL}(Y_2 - \phi(X_1' \theta_{o1})) \end{bmatrix} \quad L \times 1 \text{ vector};$$

for  $l = 1, \dots, L$ ,

$$D_{w^{2l}}(X, \alpha_o) = E \left( -e_l X_1' + w^{2l}(Y_2 - \phi(X_1' \theta_{o1})) \mid X \right),$$

where  $e_l$  denotes the  $L \times 1$ -vector with 0's everywhere except 1 in the  $l$ -th element. And denote

$$D_w(X, \alpha_o) = (D_{w^1}(X, \alpha_o), D_{w^{21}}(X, \alpha_o), \dots, D_{w^{2L}}(X, \alpha_o)).$$

Since  $\theta \in \Theta \subset \mathcal{R}^b$  with  $b \equiv (1+L) \dim(X_1)$ , obviously  $D_w(X, \alpha)$  is a  $L \times b$ -matrix valued function, where  $D_{w^j}(X, \alpha_o)$  is a  $L \times \dim(X_1)$ -matrix valued function for  $j = 1, 21, \dots, 2L$ .

Define the spaces

$$L_2(\mathcal{X}, \Sigma) = \{f : \mathcal{X} \rightarrow \mathcal{R}^{L \times \dim(X_1)} : E[f(X)' \Sigma(X)^{-1} f(X)] < \infty\}$$

$$\mathcal{S} = \left\{ \begin{array}{l} s \in L_2(\mathcal{X}, \Sigma) : s(X) = E[g(Y_2 - \phi(X_1' \theta_{o1})) \mid X] \\ \text{for some } g \text{ such that } E[|g(Y_2 - \phi(X_1' \theta_{o1}))| \mid X] < \infty \end{array} \right\}$$

and  $\overline{\mathcal{S}}$  the closure of  $\mathcal{S}$  in  $L_2(\mathcal{X}, \Sigma)$ . Let  $\text{Proj}_{\overline{\mathcal{S}}}(\cdot)$  denote the orthogonal projection onto  $\overline{\mathcal{S}}$ , and  $\overline{\mathcal{S}}^\perp$  denote the subspace of  $L_2(\mathcal{X}, \Sigma)$  that is orthogonal to  $\overline{\mathcal{S}}$ .

We also denote  $E[w^{*j}(Y_2 - \phi(X_1' \theta_{o1})) \mid X]$ ,  $j = 1, 21, \dots, 2l, \dots, 2L$ , as the solution to

$$\inf_{E[w^j(\cdot) \mid X] \in \overline{\mathcal{S}}} E \left[ D_{w^j}(X, \alpha_o)' \Sigma(X)^{-1} D_{w^j}(X, \alpha_o) \right], \quad (5.1)$$

or equivalently,

$$E \left[ D_{w^{*j}}(X, \alpha_o)' \Sigma(X)^{-1} \{E[w^j(Y_2 - \phi(X_1' \theta_{o1})) - w^{*j}(Y_2 - \phi(X_1' \theta_{o1})) \mid X]\} \right] = 0$$

$$\text{for all } E[w^j(Y_2 - \phi(X_1' \theta_{o1})) \mid X] \in \overline{\mathcal{S}}, \quad j = 1, 21, \dots, 2l, \dots, 2L.$$

Let  $E[w^*(\cdot) \mid X] = E[(w^{*1}(\cdot), w^{*21}(\cdot), \dots, w^{*2L}(\cdot)) \mid X]$ , then  $E[D_{w^*}(X, \alpha_o)' \Sigma(X)^{-1} D_{w^*}(X, \alpha_o)]$  is simply a  $b \times b$ -variance-covariance matrix with the  $(j, k)$ -th element given by  $E[D_{w^{*j}}(X, \alpha_o)' \Sigma(X)^{-1} D_{w^{*k}}(X, \alpha_o)]$  for  $j, k = 1, 21, \dots, 2L$ . Lemma 1 in the Appendix shows that  $E[D_{w^*}(X, \alpha_o)' \Sigma(X)^{-1} D_{w^*}(X, \alpha_o)]$  is finite, positive-definite under the following Assumption N1:

- Assumption N1.** (i)  $E[\nabla h_o(Y_2 - \phi(X_1' \theta_{o1})) \mid X] \nabla \phi(X_1' \theta_{o1})$  is not a constant;  
(ii)  $E[\nabla h_o(Y_2 - \phi(X_1' \theta_{o1})) \mid X] \nabla \phi(X_1' \theta_{o1}) X_1' a - \text{Proj}_{\overline{\mathcal{S}}}(E[\nabla h_o(Y_2 - \phi(X_1' \theta_{o1})) \mid X] \nabla \phi(X_1' \theta_{o1}) X_1' a) = 0$  a.s. if and only if  $a \equiv 0$ ;  
(iii)  $X_1' a - \text{Proj}_{\overline{\mathcal{S}}}(X_1' a) = 0$  a.s. if and only if  $a \equiv 0$ .

**Assumption N2.** (i)  $\theta_o \in \text{int}(\Theta)$ ; (ii)  $\Sigma_o(X) \equiv \text{Var}[\rho(Z, \alpha_o)|X]$  is positive definite for all  $X \in \mathcal{X}$ .

**Proposition 3.** *Suppose all conditions of Proposition 2, Assumptions N1-N2 and N3 (in the Appendix) are satisfied. Then:  $\sqrt{n}(\widehat{\theta}_n - \theta_o) \implies \mathcal{N}(0, V^{-1})$ , where*

$$V \equiv \begin{bmatrix} E\{D_{w^*}(X, \alpha_o)' \Sigma(X)^{-1} D_{w^*}(X, \alpha_o)\} \times \\ (E\{D_{w^*}(X, \alpha_o)' \Sigma(X)^{-1} \Sigma_o(X) \Sigma(X)^{-1} D_{w^*}(X, \alpha_o)\})^{-1} \times \\ E\{D_{w^*}(X, \alpha_o)' \Sigma(X)^{-1} D_{w^*}(X, \alpha_o)\} \end{bmatrix}. \quad (5.2)$$

Lemma 2 in the Appendix provides simpler expression of  $D_{w^*}(X, \alpha_o)$ ; and shows that Assumption N1 is implied by the following Assumption N1':

**Assumption N1'.** the conditional distribution of  $Y_2$  given  $X$  depends on  $X$  only through  $X_2 - \phi(X_1' \theta_{o1})$ ; further,

- (i)  $E[\nabla h_o(Y_2 - \phi(X_1' \theta_{o1})) | X_2 - \phi(X_1' \theta_{o1})] \nabla \phi(X_1' \theta_{o1})$  is not a constant;
- (ii)  $E[\nabla h_o(Y_2 - \phi(X_1' \theta_{o1})) | X_2 - \phi(X_1' \theta_{o1})] \{\nabla \phi(X_1' \theta_{o1}) X_1' a - E[\nabla \phi(X_1' \theta_{o1}) X_1' a | X_2 - \phi(X_1' \theta_{o1})]\} = 0$  a.s. iff  $a \equiv 0$ ;
- (iii)  $X_1' a - E[X_1' a | X_2 - \phi(X_1' \theta_{o1})] = 0$  a.s. iff  $a \equiv 0$ .

**Remark 6:** (i) Assumption N1' includes the exogenous  $Y_2 = X_2$  as a special case.

(ii) When  $X_1 \in \{0, 1\}$  and  $Y_2, X_2$  are scalar continuous random variables, the assumption that the conditional density of  $Y_2$  given  $X$  depends on  $X$  only through  $X_2 - X_1' \theta_{o1}$  is reasonable. In particular, we need such an assumption in the identification theorem 1.

### 5.3. Semiparametric Efficiency

To obtain semiparametric efficient estimator of  $\theta_o$ , we may follow the three-step procedure described in subsection 3.3. In particular we can estimate  $\Sigma_o(X)$  in **Step 2** by regressing  $\rho(Z, \widehat{\alpha}_n) \rho(Z, \widehat{\alpha}_n)'$  on  $p^{J_n}(X)$ . Let  $\Sigma_o(X) = [\sigma_{olk}(X)]_{l,k=1,\dots,L}$ , where  $\sigma_{olk}(X) = E[\rho_l(Z, h_{ol}, \theta_1, \theta_{o2l}) \rho_k(Z, h_{ok}, \theta_1, \theta_{o2k}) | X]$  denotes the  $(l, k)$ -th element. Then  $\Sigma_o(X)$  is estimated by  $\widehat{\Sigma}_o(X, \widehat{\alpha}_n)$ , where  $\widehat{\Sigma}_o(X, \alpha)$  denotes the  $L \times L$ -matrix with its  $(l, k)$ -th element given by  $\widehat{\sigma}_{olk}(X, \alpha)$ :

$$\widehat{\sigma}_{olk}(X, \alpha) \equiv p^{J_n}(X)' (P' P)^{-1} \sum_{i=1}^n p^{J_n}(X_i) \rho_l(Z_i, \alpha) \rho_k(Z_i, \alpha).$$

**Proposition 4.** *Let  $\widetilde{\alpha}_n = (\widetilde{\theta}_n, \widetilde{h}_n)$  be the three-step estimator (3.5). Suppose all conditions of Proposition 3 are satisfied with  $\Sigma = \Sigma_o$ . Then  $\widetilde{\theta}_n$  is asymptotic efficient,*

and  $\sqrt{n}(\tilde{\theta}_n - \theta_o) \implies \mathcal{N}(0, V_o^{-1})$ , where

$$V_o = \inf_{E[w^j(\cdot)|X] \in \bar{\mathcal{S}}_o, j=1, 21, \dots, 2L} E [D_w(X, \alpha_o)' [\Sigma_o(X)]^{-1} D_w(X, \alpha_o)],$$

and  $\bar{\mathcal{S}}_o$  is defined the same way as  $\bar{\mathcal{S}}$  except with  $\Sigma = \Sigma_o$ .

#### 5.4. Covariance Estimator

To conduct any statistical inference using the semiparametric efficient estimator  $\tilde{\theta}$ , we need a consistent estimator  $\tilde{V}_o$  of  $V_o$ . Let  $\hat{D}_w(x, \tilde{\alpha}) = (\hat{D}_{w^1}(x, \tilde{\alpha}), \hat{D}_{w^{21}}(x, \tilde{\alpha}), \dots, \hat{D}_{w^{2L}}(x, \tilde{\alpha}))$  denote a consistent estimator of  $D_w(x, \alpha_o) = (D_{w^1}(x, \alpha_o), D_{w^{21}}(x, \alpha_o), \dots, D_{w^{2L}}(x, \alpha_o))$ . Then  $\tilde{V}_o$  can be computed as

$$\tilde{V}_o = \inf_{w^j(\cdot) \in \bar{\mathcal{W}}_n, j=1, 21, \dots, 2L} \frac{1}{n} \sum_{i=1}^n \hat{D}_w(X_i, \tilde{\alpha})' [\hat{\Sigma}_o(X_i)]^{-1} \hat{D}_w(X_i, \tilde{\alpha}),$$

where  $\bar{\mathcal{W}}_n$  denote the linear completion of  $\mathcal{H}_n - \{\Pi_n h_o\}$ , and  $\mathcal{H}_n$  could be the same sieve space (3.10) used to compute  $h$ . That is, each  $w^j(Y_{2i} - \phi(X'_{1i} \tilde{\theta}_1)), j = 1, 21, \dots, 2L$  can be approximated by the same basis for  $h$  before:

$$w^j(Y_{2i} - \phi(X'_{1i} \theta_1)) \approx \sum_{k=0}^{K_n} \sum_{t \in \mathcal{K}_n} \pi_{kt}^j 2^{k/2} B_\gamma(2^k [Y_{2i} - \phi(X'_{1i} \theta_1)] - t).$$

In this paper we use

$$\hat{D}_{w^j}(x, \tilde{\alpha}) = p^{J_n}(x)' (P' P)^{-1} \sum_{i=1}^n p^{J_n}(X_i) \left( \frac{d\rho(Z_i, \tilde{\alpha})}{d\theta_j} - \frac{d\rho(Z_i, \tilde{\alpha})}{dh} [w^j(Y_{2i} - \phi(X'_{1i} \tilde{\theta}_1))] \right),$$

and

$$\begin{aligned} & \frac{d\rho(Z_i, \tilde{\alpha})}{d\theta_1} - \frac{d\rho(Z_i, \tilde{\alpha})}{dh} [w^1(Y_{2i} - \phi(X'_{1i} \tilde{\theta}_1))] \\ &= \nabla \tilde{h}(Y_{2i} - \phi(X'_{1i} \tilde{\theta}_1)) \nabla \phi(X'_{1i} \tilde{\theta}_1) X'_{1i} + w^1(Y_{2i} - \phi(X'_{1i} \tilde{\theta}_1)) \end{aligned}$$

where

$$\nabla \tilde{h}(Y_{2i} - \phi(X'_{1i} \tilde{\theta}_1)) = \begin{bmatrix} \nabla \tilde{h}_1(Y_{2i} - \phi(X'_{1i} \tilde{\theta}_1)) \\ \dots \\ \nabla \tilde{h}_L(Y_{2i} - \phi(X'_{1i} \tilde{\theta}_1)) \end{bmatrix} \quad L \times 1 \text{ vector,}$$

and for  $l = 1, \dots, L$ ,

$$\frac{d\rho(Z_i, \tilde{\alpha})}{d\theta_{2l}} - \frac{d\rho(Z_i, \tilde{\alpha})}{dh} [w^{2l}(Y_{2i} - \phi(X'_{1i}\tilde{\theta}_1))] = -e_l X'_{1i} + w^{2l}(Y_{2i} - \phi(X'_{1i}\tilde{\theta}_1)).$$

That is, our estimate  $\tilde{V}_o$  is the simple weighted LS residuals:

$$\tilde{V}_o = \inf_{\{\pi_{kt}^j\}, j=1, 21, \dots, 2L} \frac{1}{n} \sum_{i=1}^n \hat{D}_w(X_i, \tilde{\alpha})' [\hat{\Sigma}_o(X_i)]^{-1} \hat{D}_w(X_i, \tilde{\alpha})$$

**Proposition 5.** *Under the same conditions as that for Proposition 4, we have:  $\tilde{V}_o = V_o + o_p(1)$ .*

## 6. Implementation and Simulation

In this section, we first describe how the general sieve minimum distance procedure can be easily implemented in our application to a system of shape invariant Engel curves. We then present a Monte Carlo study to evaluate the performance of nonparametric IV regression.

### 6.1. Implementation

We only discuss the implementation of the sieve minimum distance for the case when  $Y_2$  is endogenous; the exogenous case follows along the same lines.

In the empirical application we have  $\phi(X'_1\theta_1) = X'_1\theta_1$ . The unknown function  $h_l$  is approximated by  $h_{nl}(Y_2 - X'_1\theta_1) = B^{k_n}(Y_2 - X'_1\theta_1)' \Pi^l$ , where  $B^{k_n}(y) = (B_1(y), \dots, B_{k_n}(y))'$  is the set of basis functions and  $k_n$  is the number of unknown sieve coefficient in approximating  $h_l$ .<sup>12</sup> In the empirical section we have implemented two kinds of sieve basis functions for  $h_l$ : (1) the cardinal B-spline sieve (3.10) of order  $\gamma$  (with  $\gamma \geq 3$ ),

$$B^{k_n}(Y_2 - X'_1\theta_1)' \Pi^l = \sum_{k=0}^{K_n} \sum_{j \in \mathcal{K}_n} \pi_{kj}^l 2^{k/2} B_\gamma(2^k[Y_2 - X'_1\theta_1] - j);$$

and (2) the polynomial spline sieve of order  $q_n$  with  $r_n$  number of knots,

$$B^{k_n}(Y_2 - X'_1\theta_1)' \Pi^l = \sum_{j=0}^{q_n} \pi_j^l (Y_2 - X'_1\theta_1)^j + \sum_{k=1}^{r_n} \pi_{q_n+k}^l (Y_2 - X'_1\theta_1 - \nu_k)_+^{q_n},$$

---

<sup>12</sup>Recall that we use the same basis functions and the same number of sieve terms for all goods  $l = 1, \dots, L$ .

where  $(y - \nu)_+^q = \max\{(y - \nu)^q, 0\}$  while  $\{\nu_k\}_{k=1, \dots, r_n}$  are the knots. For any given value of  $r_n$ , the knots  $\{\nu_k\}$  are simply chosen as the empirical quantiles of  $Y_2$ , i.e.,  $\nu_k = \frac{k}{r_n+1}$ th quantile of  $Y_2$ . In this case we have  $k_n = q_n + r_n + 1$ .

In the empirical section, the unknown conditional mean function  $m_l(X, \alpha) = E[Y_{1l} - B^{k_n}(Y_2 - X'_1\theta_1)' \Pi^l - X'_1\theta_{2l} \mid X]$  is approximated by  $m_{nl}(X, \alpha) = p^{J_n}(X)' A^l$ , where  $p^{J_n}(x) = (p_{o1}(x), \dots, p_{oJ_n}(x))'$  is the set of basis functions and  $J_n$  is the number of unknown sieve coefficient in approximating  $m_l$ .<sup>13</sup> Let  $P = (p^{J_n}(X_1), \dots, p^{J_n}(X_n))'$ , then the linear sieve estimator  $\hat{m}_l$  given in (3.13) is now

$$\hat{m}_l(X, \alpha) = \sum_{i=1}^n [Y_{1il} - X'_{1i}\theta_{2l} - B^{k_n}(Y_{2i} - X'_{1i}\theta_1)' \Pi^l] p^{J_n}(X_i)' (P'P)^{-1} p^{J_n}(X).$$

We can easily compute the sieve profile estimator (c.f. Remark 3) for  $\alpha = (\theta, \Pi)$ . In the following we ignore the budget share constraints  $0 \leq h_{ol} \leq 1$  and  $0 \leq \sum_{l=1}^L h_{ol} \leq 1$ , and we first compute the sieve profile estimator for  $h_l$  with  $\hat{\Sigma}(X_i) = I_L$ . For any fixed  $\theta \in \Theta$ , we estimate  $\Pi^l$  by minimizing  $\sum_{i=1}^n \{\hat{m}_l(X_i, \alpha)\}^2$  subject to the smoothness constraints imposed on the function  $h_l$ :

$$\min_{\Pi^l: \int |\nabla^r h_{nl}(y)|^2 dy \leq D} [\mathcal{Y}_{1l}(\theta_{2l}) - \mathcal{B}(\theta_1)\Pi^l]' P (P'P)^{-1} P' [\mathcal{Y}_{1l}(\theta_{2l}) - \mathcal{B}(\theta_1)\Pi^l] \quad (6.1)$$

for some given upper bound  $D > 0$ , where

$$\mathcal{Y}_{1l}(\theta_{2l}) = (Y_{11l} - X'_{11}\theta_{2l}, \dots, Y_{1nl} - X'_{1n}\theta_{2l})', \quad (n \times 1) \quad (6.2)$$

$$\mathcal{B}(\theta_1) = (B^{k_n}(Y_{21} - X'_{11}\theta_1), \dots, B^{k_n}(Y_{2n} - X'_{1n}\theta_1))' \quad (n \times k_n). \quad (6.3)$$

Let  $C = (C_{kj})$  be the  $k_n \times k_n$ -matrix given by  $C_{kj} = \int [\nabla^r B_k(y)] [\nabla^r B_j(y)] dy$ ,<sup>14</sup> and  $\lambda \geq 0$  the Lagrange multiplier. Then an equivalent representation of (6.1) for a given known  $D$  is:

$$\min_{\Pi^l, \lambda} \left( \mathcal{Y}_{1l}(\theta_{2l}) - \mathcal{B}(\theta_1)\Pi^l \right)' P (P'P)^{-1} P' \left( \mathcal{Y}_{1l}(\theta_{2l}) - \mathcal{B}(\theta_1)\Pi^l \right) + \lambda \{ \Pi^l C \Pi^l - D \}. \quad (6.4)$$

In practice we often take  $C$  to be either first or second order derivative matrix, although we know that Engel curve  $h_l$  itself is bounded between zero and one, but we do not know the bound  $D$  on its derivatives. Hence we could solve (6.4) by either letting the bound

<sup>13</sup>We also use the same basis functions and the same number of sieve terms for all conditional means  $m_l$ ,  $l = 1, \dots, L$ .

<sup>14</sup>One may also choose  $C$  as the self-adjoint difference (Gram) matrix, c.f. Schumaker (1993, p. 203).

$D$  grow slowly with the sample size (say  $D = \log \log n$  or  $\log n$ ),<sup>15</sup> or simply letting the bound  $D$  take a few possible values (such as 1.5, 2, 3).<sup>16</sup> The latter method is equivalent to solve the following problem for a few possible values of  $\lambda \geq 0$ :

$$\min_{\Pi^l} \left( \mathcal{Y}_{1l}(\theta_{2l}) - \mathcal{B}(\theta_1)\Pi^l \right)' P(P'P)^{-1}P' \left( \mathcal{Y}_{1l}(\theta_{2l}) - \mathcal{B}(\theta_1)\Pi^l \right) + \lambda \Pi^{l'} C \Pi^l. \quad (6.5)$$

The minimization problem (6.5) has a simple close form solution:

$$\widehat{\Pi}_\lambda^l(\theta) = \left( \mathcal{B}(\theta_1)' P(P'P)^{-1} P' \mathcal{B}(\theta_1) + \lambda C \right)^{-1} \mathcal{B}(\theta_1)' P(P'P)^{-1} P' \mathcal{Y}_{1l}(\theta_{2l}),$$

and the corresponding sieve profile estimator of  $h_l$  for any given value of  $\theta$  is  $\widehat{h}_{nl}(\theta; \cdot) = B_n(\cdot)' \widehat{\Pi}_\lambda^l(\theta)$ .<sup>17</sup> We note that when  $\lambda = 0$  (i.e. without smoothness constraints), the above solution is simply the well known IV/2SLS estimator.

Next, we plug  $\widehat{h}_n(\theta; \cdot) = (\widehat{h}_{n1}(\theta; \cdot), \dots, \widehat{h}_{nL}(\theta; \cdot))'$  into the problem (3.8) with  $\widehat{\Sigma}(X_i) = I_L$ , and end up with a low-dimensional optimization problem for  $\theta$ :

$$\min_{\theta \in \Theta} \left( \sum_{i=1}^n \rho(Z_i, \theta, \widehat{h}_n(\theta; \cdot)) \otimes p^{J_n}(X_i) \right)' \left( I \otimes (P'P) \right)^{-1} \left( \sum_{i=1}^n \rho(Z_i, \theta, \widehat{h}_n(\theta; \cdot)) \otimes p^{J_n}(X_i) \right). \quad (6.6)$$

The solution  $\widehat{\theta}$  to the problem (6.6) will be a root- $n$  consistent estimator for  $\theta_o$ , and the corresponding sieve estimator for  $h_{ol}$  is  $\widehat{h}_{nl}(\widehat{\theta}; \cdot) = B_n(\cdot)' \widehat{\Pi}_\lambda^l(\widehat{\theta})$ ,  $l = 1, \dots, L$ . To solve the last problem (6.6), one needs to run a numerical routine since  $\theta$  enters non-linearly, but it is relatively easily performed compared to optimizing over both  $\Pi$  and  $\theta$  simultaneously<sup>18</sup>.

The above sieve profile estimate of  $\alpha = (\theta, \Pi)$  with  $\widehat{\Sigma}(X_i) = I_L$  can then be used as starting point for the numerical routine required to run the optimally weighted estimation procedure where we optimize over both  $\Pi$  and  $\theta$  simultaneously. This extra step saves one a considerable amount of computation time, since the profile estimator in general is close to the simultaneous one.

In our application we shall also present fully nonparametric estimates of the model of the Engel curves as given in (2.3):  $Y_{1l} = G_l(Y_2, X_1) + \varepsilon_l$  for  $l = 1, \dots, L$ ; this is

<sup>15</sup>This is the approach taken in Chen and Shen (1998) and in the old version of Ai and Chen (2003).

<sup>16</sup>This is the approach taken in Newey and Powell (2003) in their Monte Carlo simulations.

<sup>17</sup>In a similar fashion, one can obtain closed form (albeit more complicated) expressions for profile estimates of  $\Pi_\lambda^l$  in the case where  $\widehat{\Sigma}(X_i) = \widehat{\Sigma}_o(X_i)$  is a consistent estimator of the conditional covariance matrix.

<sup>18</sup>In fact, since  $\theta_2$  also enters linearly a closed form expression for the estimator of this can also be derived such that only  $\widehat{\theta}_1$  has to be found numerically.

done both for  $Y_2$  assumed exogenous  $E[\varepsilon_l|Y_2, X_1] = 0$ , and for  $Y_2$  assumed endogenous  $E[\varepsilon_l|X_2, X_1] = 0$ . We estimate  $G_l$  using the same sieve minimum distance procedure as for the semiparametric model. First, approximate  $G_l(\cdot)$  by  $G_{nl}(\cdot) = B^{k_n}(\cdot)' \Pi^l$ , and then obtain an estimator of  $\Pi^l$  by solving

$$\begin{aligned} & \min_{\Pi^l} \sum_{i=1}^n \widehat{m}_l(X_i, G_{nl}) \widehat{m}_l(X_i, G_{nl}) + \lambda \left( \Pi^l \right)' C \Pi^l \\ & = \min_{\Pi^l} [\mathcal{Y}_{1l} - \mathcal{B} \Pi^l]' P (P' P)^{-1} P' [\mathcal{Y}_{1l} - \mathcal{B} \Pi^l] + \lambda \left( \Pi^l \right)' C \Pi^l, \end{aligned} \quad (6.7)$$

where  $\mathcal{Y}_{1l} = (Y_{11l}, \dots, Y_{1nl})'$ ,  $\mathcal{B} = (B^{k_n}(Y_{21}, X_{11}), \dots, B^{k_n}(Y_{2n}, X_{1n}))'$ ,  $P$  is given as before, and the smoothness penalization matrix  $C$  is as before except that the  $r$ -th partial derivatives of  $B_j(y_2, x_1)$  are only with respect to the  $y_2$  argument. Again the minimization problem (6.7) has a simple closed form solution

$$\widehat{\Pi}_\lambda^l = (\mathcal{B}' P (P' P)^{-1} P' \mathcal{B} + \lambda C)^{-1} \mathcal{B}' P (P' P)^{-1} P' \mathcal{Y}_{1l}, \quad (6.8)$$

such that  $\widehat{G}_l(y_2, x_1) = B^{k_n}(y_2, x_1)' \widehat{\Pi}_\lambda^l$  for  $l = 1, \dots, L$ .

In the actual implementation of the above procedures, one has to specify  $\{p_{oj}(\cdot)\}_{j \geq 1}$ ,  $\{B_k(\cdot)\}_{k \geq 1}$ ,  $J_n$ ,  $k_n$ ,  $\lambda$  and  $r$ . We set the smoothness  $r = 2$  which includes the popular QUAIDS Engle curve specification as a special case. There is a wide range of various basis functions which one can choose for  $p_{oj}(\cdot)$  and  $B_k(\cdot)$ , e.g. Hermite polynomials, wavelet cardinal B-splines, polynomial splines and Fourier series. Our empirical findings below are not sensitive to the choice of basis functions. The theoretical results obtained in Sections 4 and 5 give us certain guidelines about how to choose  $J_n$  and  $k_n$ . For purely nonparametric IV regression, one should choose  $J_n \geq k_n$  to ensure identification and the requirement of  $\lim_n(J_n/k_n) = c_0 \geq 1$  in Theorem 2, but then the choice of  $J_n$  will be mainly related to the invertibility of the matrix  $P'P$ , and the “quality” of the instruments  $p^{J_n}(X)$  for the endogenous regressors  $B^{k_n}(Y_2)$ , see the empirical section for details. In our application we have found that  $J_n = c_0 k_n$ ,  $c_0 \approx 2$  or  $3$  works fine. When one imposes smoothness restrictions,  $\lambda$  also has to be chosen. There is a certain interdependence between  $k_n$  and  $\lambda$ ; a high number  $k_n$  could potentially lead to overfitting (i.e. the estimated Engel curve  $h_l, l = 1, \dots, L$  becomes wiggly and the variance gets big), but this can be controlled for with a slightly bigger penalisation weight  $\lambda$ . For a given choice of  $k_n$ , one may try out different values of  $\lambda$  and choose the one which appears most plausible; an alternative method would be to use a data-driven procedure such as the generalized cross-validation (GCV) to choose  $\lambda \in [0, 1]$ ,

see e.g. Eubank (1988) for a discussion of this procedure in a standard nonparametric least squares regression setting. However, there is no theoretical justification for such a procedure in the endogenous case.<sup>19</sup> In fact, how to optimally choose  $k_n$  and  $\lambda$  simultaneously is still an open problem even in the standard sieve nonparametric LS regressions. Finally, for the semiparametric efficient estimation of  $\theta$ , one can choose slightly bigger  $J_n, k_n$  (or smaller  $\lambda$ ) than those values for the purely nonparametric IV estimation of  $h$ , however, the choices of  $J_n, k_n$  (with  $J_n = c_0 k_n$ ,  $c_0 > 1$ ) have to satisfy Assumption 5 to ensure the  $\sqrt{n}$ -normality of  $\theta$ .

## 6.2. A Monte Carlo Study

Before applying the sieve minimum distance estimators of the shape-invariant Engel curves to the British FES data set, we assess the performance of the purely nonparametric IV estimator in a small simulation study, where the Monte Carlo design will mimic the specific FES data set. In particular we are concerned with the quality of our chosen instrument  $X_2$  (gross earnings) for the endogenous variable  $Y_2$  (total expenditure), as well as the impacts of the choice of sieve basis functions and various smoothing parameters. The simulation results suggest that our chosen instrument  $X_2$  is a reasonable one in the sense that our sieve IV estimator performs well for the FES data set in consideration. Moreover, our sieve IV estimator is found to be relatively insensitive to the choice of sieve basis functions, while many different combinations of smoothing parameters  $J_n, k_n$  and  $\lambda$  will lead to similar estimated functions which are all consistent estimates of the true unknown function.

For a detailed description of the data set, we refer the readers to the empirical section. All we need to know here is that the data set consists of two sub-samples of households: one consisting of families with no children and one of families with 1-2 children, hence  $X_1 \in \{0, 1\}$ . In this simulation study, we shall only use the data from the group of households with no children (i.e.  $X_1 = 0$ ), which has sample size  $n = 628$ . For each household in this group, we observe an endogenous regressor  $Y_2$  (log-total expenditure), and an instrument  $X_2$  (normal transformation of log-gross earnings, which takes values in  $[0, 1]$ ). We may then estimate the joint density of  $(Y_2, X_2)$  using kernel methods, and denote the resulting nonparametric estimator as  $\hat{f}(y_2, x_2)$ , from which the data will be

---

<sup>19</sup>Another approach would be to specify  $D$  (the constraint parameter in equation (6.5)), and then leave  $\lambda$  as a free variable. We would then have to estimate  $\alpha$  and  $\lambda$  simultaneously; see Newey and Powell (2003) for the this approach.

drawn in our simulation study. The model we simulate is given by

$$Y_1 = h_o(Y_2) + \varepsilon, \quad \varepsilon = E[h_o(Y_2) | X_2] - h(Y_2) + v, \quad (6.9)$$

where  $v \sim N(0, 0.01)$  and is independent of  $(Y_2, X_2) \sim \hat{f}(y_2, x_2)$ . We draw an i.i.d. sample from  $(v, Y_2, X_2)$  with sample size  $n = 628$ , and use these to calculate  $Y_1$  via (6.9) for two choices of  $h_o$ , one is linear and the other is nonlinear:

$$(1) \text{ linear} \quad h_o(y_2) = -0.1095y_2 + 0.7229,$$

which closely mimics the estimate obtained for food-in expenditure in the empirical application;

$$(2) \text{ nonlinear} \quad h_o(y_2) = \Phi((y_2 - 5.5)/0.3),$$

where  $\Phi$  denotes the standard normal cdf. and where the mean (5.5) and variance  $(0.3^2)$  have been chosen such that  $h_o(y_2) \approx 0$  for  $y_2 = \min_i \{Y_{2i}\}$  and  $h_o(y_2) \approx 1$  for  $y_2 = \max_i \{Y_{2i}\}$ .

For each choice of  $h_o$ , we simulate 100 data sets  $\{(Y_{2i}, X_{2i}, Y_{1i})\}_{i=1}^{n=628}$ , and for each simulated data set we estimate  $h_o$  using the sieve nonparametric IV-estimator. We tried various basis functions  $\{p_{oj}(X_2)\}_{j=1}^{J_n}$  and  $\{B_k(Y_2)\}_{k=1}^{k_n}$  for the conditional mean  $m(X_2, h) = E[Y_1 - h(Y_2) | X_2]$  and  $h$  respectively, all yielding similar results as long as the sieve approximating terms  $J_n$  and  $k_n$  and the penalization weight  $\lambda$  are similar. However, due to the length of the paper, here we only report the simulation results for a few combinations:  $h$  is approximated by either a 3rd order polynomial spline with  $k_n = 4, 5, 6, 7, 8, 9$ , or a 3rd order B-spline with  $k_n = 9, 14$ ;  $m$  is approximated by either a cos-sin basis with  $J_n \approx 3k_n, 27$ , or a 4th order B-spline with  $J_n = 15, 25$ . To check for the robustness of the sieve IV estimators towards the choice of penalisation, we also present the results for penalizing the integrated squared norm of the second order derivatives of  $h$  with different weights  $\lambda = 0.8, 0.1, 0.01, 0.001$  and  $0.0$ .

Tables B.1 and B.2 report the integrated squared bias, variance, and MSE based on the 100 simulations for the sieve IV estimators of nonlinear  $h$ ,<sup>20</sup> where  $h$  was estimated

---

<sup>20</sup>Let  $\hat{h}_i$  be the estimate of  $h_o$  from the  $i$ th simulated data set, and  $\bar{h}(y) = \sum_{i=1}^{100} \hat{h}_i(y)/100$  be the pointwise average across 100 simulations. We calculate the pointwise squared bias as  $[\bar{h}(y) - h_o(y)]^2$ , and the pointwise variance as  $100^{-1} \sum_{i=1}^{100} [\hat{h}_i(y) - \bar{h}(y)]^2$ . The integrated squared bias is calculated by numerically integrating the pointwise squared bias from  $\underline{y}$  to  $\bar{y}$  which are respectively the 2.5th and 97.5th empirical percentiles of  $Y_2$  from the no-kids subsample of the FES data set our simulations are based on; The integrated variance and the integrated MSE are computed in a similar way.

using different sieves with different smoothing parameters  $k_n, J_n$  and  $\lambda$ . In all the cases, the sieve IV estimator behaves well for  $\lambda \geq 0.01$ ; the integrated bias of the sieve IV estimators is relatively small and is quite insensitive towards the choice of the penalisation weight  $\lambda$ ; however the variance increases as  $\lambda$  decreases; and  $\lambda = 0.8$  yields the best performance in terms of the integrated MSE for large  $k_n \geq 8$ . We obtain similar results for the integrated squared bias, variance and MSE of sieve IV estimators of linear  $h$ , but we do not tabulate them here due to the length of the paper. The sieve IV estimates for the linear and the nonlinear  $h$  are plotted in Figures C.1 and C.2 for  $k_n = 9, J_n = 25, \lambda = 0.8$  and  $0.0$ , where the - - lines denote the true  $h$  and the — lines denote the estimates. Figures C.3 and C.4 present the corresponding plots for  $k_n = 5, J_n = 15$ . These plots are similar to the  $k_n = 9$  case and suggest little sensitivity to these choices for  $k_n$  and  $J_n$ . Since the  $k_n = 5$  case satisfies Assumption 5 for root-n normality and efficiency of  $\theta$ , the similarity of the plots is important for our empirical application in the next section.

From these results it is apparent that imposing smoothness constraints (i.e.,  $\lambda > 0$ ) improves the quality of the sieve IV estimators, both in terms of the variance and the smoothness of the estimated functions. However, the overall shape of the estimated functions and their relative positions to the true  $h$  are not strongly affected by the choice of  $\lambda$ , which again indicates that for a given value  $k_n$  of sieve terms in approximating  $h$ , the penalization weight  $\lambda$  does not have a great influence on the bias of the sieve IV estimator.

In the empirical section below we note that for the group of families with no children, the Stock-Yogo (2002) test for weak instruments in the parametric linear 2SLS regression problem, suggests the presence of weak instruments under the specification  $\lambda = 0.0, J_n = 15$  or  $25$ , and  $k_n \geq 4$ . For the sample with children and for the pooled sample this turns out not to be the case. So again our focus here on the sample without children is relevant. However, a consequence of weak instruments is that if one wrongly treats the sieve IV estimation as a parametric 2SLS regression, then each estimated sieve coefficients will be heavily biased towards their LS estimates. Hence, the corresponding sieve IV curve should be biased towards the inconsistent sieve LS estimator of  $h$ . Figures C.1 and C.2 show no indication of any bias towards LS. This is also confirmed by Table B.2 where with  $\lambda = 0.0$ , the bias of sieve IV generally decreases as  $J_n$  increases from 15 to 25. Finally Tables B.3 and B.4 report the integrated squared bias, variance and MSE of the sieve IV and the inconsistent sieve LS estimators of both the linear and nonlinear

$h$  with  $k_n = 6, 9$ . These tables show that the sieve LS estimator is not sensitive toward the choice of  $\lambda$ , while consistently its variance is small but its bias is big compared to the sieve IV estimator.

To summarize, we have the following Monte Carlo findings for the sieve estimators of the linear and the nonlinear  $h$ :

(1) The choices of basis functions for  $h$  [3rd order poly-spline vs 3rd order B-spline], and  $m$  [cos-sin vs 4th order B-spline] are not very important.

(2) For any fixed  $\lambda \in [0, 1]$ , the choice  $J_n = c_0 k_n$  with  $c_0 \approx 2, 3$  works well for sieve IV estimator.

(3) For any fixed  $\lambda \in [0, 1]$ , increasing  $k_n$  will slightly reduce squared bias but increase variance of the sieve IV estimator; In particular for fixed small  $\lambda \in [0.0, 0.001]$ ,  $k_n$  has to be small (4 or 5) to get a small variance (hence a small MSE) of the sieve IV estimator.

(4) For any fixed  $k_n$ , increasing  $\lambda$  towards 1 reduces variance hence makes the MSE of a sieve IV small. In particular, a large  $k_n$  (7, 8, 9, 14) can be balanced by a high  $\lambda \in [0.01, 0.8]$  that still keeps the variance and the MSE of a sieve IV estimator small.

(5) There are many combinations of smoothing parameters  $J_n, k_n, \lambda$  which can reduce the variance part and lead to a small MSE of the consistent sieve IV estimator.

(6) There is no combination of smoothing parameters  $J_n, k_n, \lambda$  which can reduce the bias part of the inconsistent sieve LS estimator, hence the inconsistent sieve LS estimator has a big MSE.

(7) For any fixed  $\lambda \in [0.1, 1]$  and fixed  $k_n$ , increasing  $J_n = c_0 k_n$  with a bigger  $c_0 \geq 2$  still leads to small MSE of sieve IV estimator, and the sieve IV estimator is not biased towards the inconsistent LS estimator.

We note that the findings (1) - (3) are consistent with our theoretical results in Sections 4 and 5. In the empirical application with sample size  $n = 1655$ , the set of smoothing parameters  $k_n = 4, 5, 6$ ,  $J_n \approx 3k_n$  and  $\lambda \in [0.0, 0.001]$  will satisfy Assumption 5 (with  $r = 2$  say) for  $\sqrt{n}$ -normality and efficiency of  $\theta$  estimates. The findings (4) - (5) should be related to the smoothing spline literature, although there is no theoretical justification yet. The finding (6) is not too surprising since, given the Monte Carlo design, the sieve LS estimators of the linear and the nonlinear  $h$  are inconsistent. The finding (7) seems in contradiction to the results in the parametric weak IV literature. It could be interesting to study the relation between the sieve IV estimation and parametric weak IV regression in the future.

Before we conclude this Monte Carlo section, we want to make sure that the Monte Carlo design does mimic the actual FES data well, so that the insights we learnt about the smoothing parameters  $k_n, J_n, \lambda$  from the Monte Carlo can be applied to the empirical estimation in the next section. From the discussion of the sieve measure of ill-posedness and from the proof of Theorem 2, we know that what is crucial for nonparametric IV regression is the behavior of the singular values  $\{\mu_k\}$  associated with the conditional expectation operator  $T(h)(\cdot) = E[h(Y_2) | X_2 = \cdot]$ . Therefore, to make sure the simulated data sets do mimic the real FES data set with no children subsample we estimate the singular values  $\{\mu_k\}$  associated with the conditional expectation operator. We restrict  $h$  to  $h(Y_2) = \Pi' B^{k_n}(Y_2)$  and impose smoothness constraints on it via the aforementioned penalisation matrix  $C$  and Lagrange multiplier  $\lambda$ , while the operator  $T$  is approximated using  $p^{J_n}$ . Then for each simulated data set, we estimate  $1 = \mu_1^2 \geq \mu_2^2 \geq \dots \geq \mu_{k_n}^2$  by the solutions to the eigenvalue problem

$$\left| \mu^2 (\mathcal{B}\mathcal{B}' + \lambda C) - \mathcal{B}P'(PP')^{-1}P\mathcal{B}' \right| = 0,$$

where  $B^{k_n}(Y_2)$  is 3rd order B-spline basis with dimension  $k_n = 14$  and  $p^{J_n}(X_2)$  is the 4th order B-spline basis with dimension  $J_n = 25$ . We repeat this 100 times. The average estimated  $k_n = 14$  singular values are:  $\hat{\mu}_1 = 0.9999$ ,  $\hat{\mu}_2 = 0.5391$ ,  $\hat{\mu}_3 = 0.3943$ ,  $\hat{\mu}_4 = 0.2890$ ,  $\hat{\mu}_5 = 0.1691$ ,  $\hat{\mu}_6 = 0.1278$ ,  $\hat{\mu}_7 = 0.0836$ ,  $\hat{\mu}_8 = 0.0412$ ,  $\hat{\mu}_9 = 0.0197$ ,  $\hat{\mu}_{10} = 0.0099$ ,  $\hat{\mu}_{11} = 0.0024$ ,  $\hat{\mu}_{12} = 0.0002$ ,  $\hat{\mu}_{13} = 1.0584e-005$ ,  $\hat{\mu}_{14} = 1.3881e-006$ . These estimates using simulated data match well with the ones estimated using the real FES data of no kids subsample, where the estimated singular values are:  $\hat{\mu}_1 = 0.9999$ ,  $\hat{\mu}_2 = 0.5586$ ,  $\hat{\mu}_3 = 0.4141$ ,  $\hat{\mu}_4 = 0.2406$ ,  $\hat{\mu}_5 = 0.1711$ ,  $\hat{\mu}_6 = 0.1143$ ,  $\hat{\mu}_7 = 0.0680$ ,  $\hat{\mu}_8 = 0.0315$ ,  $\hat{\mu}_9 = 0.0128$ ,  $\hat{\mu}_{10} = 0.0079$ ,  $\hat{\mu}_{11} = 0.0013$ ,  $\hat{\mu}_{12} = 0.0004$ ,  $\hat{\mu}_{13} = 4.1161e-006$ ,  $\hat{\mu}_{14} = 6.4036e-007$ . It is interesting to note that the corresponding singular values for the pooled sample decay less rapidly: the first 14 take the values  $\hat{\mu}_1 = 1.0000$ ,  $\hat{\mu}_2 = 0.5560$ ,  $\hat{\mu}_3 = 0.4073$ ,  $\hat{\mu}_4 = 0.3373$ ,  $\hat{\mu}_5 = 0.2595$ ,  $\hat{\mu}_6 = 0.1713$ ,  $\hat{\mu}_7 = 0.1478$ ,  $\hat{\mu}_8 = 0.1095$ ,  $\hat{\mu}_9 = 0.0910$ ,  $\hat{\mu}_{10} = 0.0385$ ,  $\hat{\mu}_{11} = 0.0235$ ,  $\hat{\mu}_{12} = 0.0047$ ,  $\hat{\mu}_{13} = 0.0028$ ,  $\hat{\mu}_{14} = 7.3967e-006$ . Consequently we expect our estimator to be better behaved on the pooled sample.

## 7. An Empirical Investigation

In this section we apply the model and the estimation procedure developed in the previous sections. We start out with a data description and examining the quality of the

gross earnings,  $X_2$ , as an instrument. We then present semi-nonparametric estimates of Engel curves under the two following assumptions: (i)  $Y_2$  is exogenous and (ii)  $Y_2$  is endogenous. We assess the importance of allowing for endogeneity both on the non-parametric shape of the Engel curve and on the parametric components which represent the demographic parameters.

### 7.1. Data Descriptives

In our application we consider  $L = 7$  broad categories of non-durables and services: alcohol, fares, food-in, food-out, fuel, leisure goods and services, and travel (motor). The data set is drawn from the British Family Expenditure Survey (FES). For the purposes of this discussion we select a single year (1995).<sup>21</sup> In order to preserve a degree of demographic homogeneity, we select from the FES a subset of married or cohabiting couples with and without children. We select those where the head of household is aged between 20 and 55 and we exclude all those with three or more children. So our demographic variable,  $X_1$ , will simply be a binary dummy variable reflecting whether the couple have 1-2 children ( $X_1 = 1$ ) or no children ( $X_1 = 0$ ) and we may write  $\phi(X_1' \theta_1)$  as  $X_1 \theta_1$ . The log of total expenditure on non-durables and services is our measure of the continuous endogenous explanatory variable  $Y_2$ . We exclude households where the head of the household is unemployed to be able to use log gross earnings as the instrumental variable,  $X_2$ , for  $Y_2$ . The earnings variable is the amount that the male of the household earned in the chosen year before taxes. This leaves us with 1655 observations, including 1027 couples with one or two children.

Table 7.1 gives some brief descriptive statistics for the main variables used in the empirical analysis. It shows the smaller share of alcohol, fares, food-out, leisure goods and travel expenditure for households with children while on the other hand the comparably larger expenditure share of food-in and fuel. This indicates strong differences in the spending patterns between the two demographic groups, and we should expect the parameter  $\theta$  in our semiparametric model to reflect these.

Figure C.5 presents the kernel density estimate of the log transformation of total expenditure. Total consumers expenditure is often supposed to have a lognormal cross section distribution which seems plausible judging from the figure. Figure C.6 shows the kernel density estimate for log-earnings. From this we see that the log-earnings

---

<sup>21</sup>It is worth noting that we have applied our IV-estimator on data from the survey for other years as well; which gave slightly different numerical results, but the qualitative conclusions were the same.

Table 7.1: Data Descriptives

	couples w/o children		couples w/ children	
	mean	std.	mean	std.
<i>budget shares:</i>				
alcohol	0.0712	0.0719	0.0496	0.0543
fares	0.0216	0.0499	0.0137	0.0399
food-in	0.1776	0.0950	0.2256	0.0938
food-out	0.0829	0.0591	0.0790	0.0555
fuel	0.0612	0.0385	0.0675	0.0364
leisure goods	0.1357	0.1456	0.1261	0.1268
travel	0.1488	0.0985	0.1324	0.0857
<i>expenditure and income:</i>				
log non-durable expenditure	5.3744	0.4864	5.4503	0.4229
log gross earnings	5.7712	0.5389	5.9112	0.5309
sample size	628		1027	

variable is not as close to a normal distribution but still reasonably close. The two variables show a strong positive correlation; for the sample with children the correlation is 0.5095 and for those without children 0.5111. Figure C.7-C.8 present plots of the bivariate kernel density estimates for these two variables together with a series estimator of  $E[Y_2|X_2 = x_2]$ . We see that the joint distribution is also smooth, and the shape of it together with the conditional mean confirms our beliefs that the gross earnings variable should be a good choice for our instrumental variable.<sup>22</sup>

## 7.2. Quality of the Instrument

Since our sieve IV estimator is similar to the parametric linear IV regression once after the basis functions  $\{p_{oj}(X)\}_{j=1}^{J_n}$  and  $\{B_k(Y_2)\}_{k=1}^{k_n}$ , and the smoothing parameters  $J_n, k_n$  and  $\lambda$  are chosen, and since the presence of weak instruments can ruin the finite and large sample properties of the classical parametric IV estimators,<sup>23</sup> we further examine the quality of our instrument by considering the following regression,

$$B^{k_n-1}(Y_2) = Ap^{J_n}(X) + e, \quad E[e|X] = 0, \quad (7.1)$$

<sup>22</sup>We have also tried disposable income as an instrument. Disposable income is arguably less likely to be exogenous as it includes savings and is measured net of taxes and benefits. However, it gave similar results and the estimated joint density of log-total expenditure and log-disposable income is closer to a bi-variate normal density.

<sup>23</sup>see e.g. Stock et al (2002).

where  $B^{k_n-1}(Y_2)$  is a vector of endogenous regressors of dimension  $k_n - 1$  (excluding the constant one from the original sieve basis  $B^{k_n}(Y_2)$ ) and  $p^{J_n}(X)$  is a vector of instrumental variables with dimension  $J_n$ . Here we take a 3rd order B-spline basis as  $B^{k_n}(Y_2)$  with  $k_n = 9$  (and  $\lambda = 0.4$ ), and let  $p^{J_n}(X)' = [B^{J_n}(X_2)', X_1 B^{J_n}(X_2)']$  where  $B^{J_n}(X_2)$  is a 4th order B-spline basis for functions of  $X_2$  with the number of sieve terms  $J_n = 15$ , hence the length of the vector  $p^{J_n}(X)$  is actually  $2J_n = 30$  for the full sample including both types of households. Throughout the empirical application,  $p^{J_n}(X_2) = B^{J_n}(X_2)$  will denote the vector of the sieve basis for conditional mean function for each subgroup of the households. We then performed two informal tests for the quality of instruments.

First, we test the hypothesis  $H_0 : \text{rank}(A) = r^*$  by applying the result in Robin and Smith (2000). We easily reject  $H_0$  for  $r^* = 0, \dots, 8$  as can be seen in Table 7.2. We note that the Robin-Smith (2000) test is based on the assumptions that the model (7.1) is correctly specified with both  $k_n$  and  $J_n$  being fixed and finite known numbers, and that  $A$  will be estimated root-n consistently with asymptotically normal distribution. However, our basic setup of unknown  $h(\cdot)$  implies that the model (7.1) with finite fixed  $k_n$  and  $J_n$  are misspecified, and that  $A$  could only be estimated at a slower than root-n rate. Nevertheless, this test could be seen as a parametric approximation of the test for  $E[h_o(Y_2) | X] = 0$ .

Table 7.2: Test of instrument

$r^*$	0	2	4	6	8
$T$	$2.2 \times 10^{14}$	$6.1 \times 10^{12}$	$6.1 \times 10^{11}$	$9.3 \times 10^{11}$	$9.4 \times 10^8$
Critical value	137.70	119.87	101.88	83.68	65.17

Second, we performed the Stock and Yogo's (2002) test of the null hypothesis  $H_0 : p^{J_n}(X_2)$  are weak instruments for  $B^{k_n-1}(Y_2)$ . This is done with  $B^{k_n-1}(Y_2) = (Y_2, Y_2^2, Y_2^3)$  for simplicity since they only report critical values for number of endogenous regressors less than or equal to 3.<sup>24</sup> The test statistic was 4.5647 for households without children, and 10.9535 for those with children. With number of instruments  $J_n = 15$ , the 5% critical values are given by 10.33 and 4.37 (for a 10% and 30 % max-

<sup>24</sup>M. Yogo has been kind enough to send us their Gauss program which could generate critical values for number of endogenous regressors greater than 3. However, we suspect that the conclusions will be similar. The test statistics reported are computed under the additive form for the Stock-Yogo test on the pooled sample.

imal bias relative to OLS respectively). When pooling the two household groups with number of instruments becoming  $2J_n = 30$ , the test statistic equals 14.0615 while the 5% critical value is 10.77 (for a 10% maximal bias relative to OLS). This indicates that the basis  $p^{J_n}(X_2)$  might not be a terribly good instrument for families with no children, but that for the pooled data and for the subsample of households with kids, it does a good job.

We note that the Stock and Yogo's (2002) test is conservative, and is based on the parametric 2SLS estimator under the assumption that the model (7.1) is correctly specified with  $k_n$  being fixed and finite known numbers. Again our basic setup of unknown  $h(\cdot)$  implies that the model (7.1) with finite fixed  $k_n$  is misspecified, and our sieve IV estimator is a penalized version of 2SLS. Therefore this parametric test result should be interpreted with great care. For example, considering the sample of no kids, and the vector of endogenous regressors  $B^{k_n-1}(Y_2)$  with fixed  $k_n$ , the result based on the Stock and Yogo (2002) test statistic will indicate that the instruments  $p^{J_n}(X_2)$  with  $J_n = 15$  are all weak and the 2SLS estimate of  $\Pi$  for the model:

$$Y_1 = B^{k_n}(Y_2)' \Pi + \varepsilon, \quad B^{k_n-1}(Y_2) = Ap^{J_n}(X_2) + e, \quad E[e|X_2] = 0,$$

will be heavily biased towards its OLS estimated values. This was not what we found in the Monte Carlo section on the sieve IV estimate of unknown  $h(\cdot)$ . There the unknown  $h(Y_2)$  is approximated by  $B^{k_n}(Y_2)' \Pi$  while the unknown  $E[B^{k_n-1}(Y_2)' \Pi | X_2]$  is approximated by  $p^{J_n}(X_2)' A \Pi$ , and the sieve IV estimate  $\hat{h} = B^{k_n}(\cdot)' \hat{\Pi}_{iv}$  is much closer to the true  $h(\cdot)$  instead of bias towards the sieve OLS estimate, see Figures C.1, C.2, C.3 and C.4. Finally, recall that the estimated singular values were larger for the pooled sample.

Altogether these results suggest that we may wish to be cautious in our interpretation of the nonparametric IV results for the sub-sample of families without children. But the results should be reliable for the pooled sample which is used for the semi-nonparametric IV estimator of the shape invariant Engel curve model developed in this paper.

### 7.3. Estimation Results

In both the case where  $Y_2$  is assumed exogenous, and the case where it is treated as endogenous, we have approximated  $h$  by several different sieve bases such as Hermite polynomials, wavelet cardinal B-splines and polynomial splines. For the endogenous case, we have also approximated the conditional mean  $m$  by several sieve bases such as

power series, cos-sin series, wavelet cardinal B-splines (with  $X_2 =$  normal transformation of log-earnings), and Hermite polynomials (with  $X_2 =$  log earnings). As a general rule, the number of sieve terms  $J_n$  for  $m$  is chosen to be larger than the number of sieve terms  $k_n$  for  $h$  and such that the  $P'P$  in (3.13) is invertible. As discussed earlier this should ensure identification of our model. To examine the robustness of the sieve estimator of the conditional mean, we also employed kernel regression methods to estimate  $m$ ;<sup>25</sup> this gave very similar results, which therefore are not reported here.

The shapes of the estimated Engel curves based on different bases all look similar as long as the number of effective sieve terms in approximating  $h$  is  $k_n = 5$  to  $9$ , and the number of sieve terms in approximating  $m$  is  $J_n = 15$  to  $27$ , excluding  $X_1$ . The smoothness parameter  $\lambda$  should increase as  $k_n$  grows; we tried out different values of  $\lambda$  for each value of  $k_n$  in the range  $5$  to  $9$ ; the shape of the Engel curves proved to be fairly robust towards the choice of  $\lambda$  in the range  $0.05$  to  $0.8$  with only the level of smoothness changing as we also saw in the Monte Carlo study.<sup>26</sup> For the estimation of  $\theta$  alone, smaller values of  $\lambda$  might be preferable. The estimation of the system was performed as described in section 6: First we obtained a profile estimator of  $\alpha = (\theta, \Pi)$  using the closed form sieve IV solution of  $\hat{\Pi}(\theta)$ . We then used this estimator as a starting point for the numerical optimisation procedure employed to obtain simultaneous estimates of  $\theta$  and  $\Pi$ . In most cases however, the simultaneous estimates proved to be practically identical to the initial profile ones. Also observe that we did not restrict  $h$  to  $0 \leq h_l \leq 1$  and  $0 \leq \sum_{l=1}^L h_l \leq 1$  in the estimation procedure; as we shall see, imposing this restriction would have no effect on our estimates anyway, since the resulting unrestricted estimates all satisfy these constraints for  $y_2$  in the domain of our sample of  $Y_2$ .

In order to obtain efficient estimates of  $\theta$ , we ran the 3-step procedure described in Section 3.3. In the 2nd step, the conditional covariance matrix,  $\Sigma_o(X)$ , is estimated. One can either use the sieve estimator described in Section 3.5 or use standard kernel methods. The results reported here are based on kernel estimates, but  $\hat{\theta}$  proved to be fairly robust to the choice of estimator for  $\Sigma_o(X)$ . To improve the estimates we ran an iterative procedure, repeating Step 2 and 3 until  $\hat{\theta}$  converged towards a stable level.

<sup>25</sup> See Härdle and Linton (1994) for a review of the kernel method.

<sup>26</sup> In general, a smaller penalization was needed in the semiparametric estimation compared to the fully nonparametric one. This owes to the fact that in the semiparametric specification the same  $h$ -function is used for both household groups, while in the non-parametric estimation a different  $h$  is used for each group. This allows us to pool the two groups of households in the semiparametric estimation, while in the nonparametric setting we treat the two groups separately.

In our case, we required that  $\|\hat{\theta}^{(i+1)} - \hat{\theta}^{(i)}\| < 0.005$  where  $\hat{\theta}^{(i)}$  denotes the estimate obtained in the  $i$ th iteration; as a rule, the convergence criterion was satisfied after 7-10 iterations.

Together with the estimated Engel curves, we also report 95% pointwise confidence bands of these. The bands were obtained using the nonparametric bootstrap based on 1000 resamples. In each resample,  $n = 1655$  observations were drawn from the original data set with replacement, and then  $h$  was reestimated. We did this with  $\theta = \hat{\theta}$  fixed at its efficient estimated value, since  $\hat{\theta}$  is  $\sqrt{n}$ -consistent while  $\hat{h}$  has a slower than  $\sqrt{n}$ -convergence rate; hence, this will have no effect asymptotically. For simplicity, we used the same  $J_n$  and  $k_n$  in the estimation of  $h$  using the bootstrap sample, however, to control for the asymptotic bias in  $\hat{h} - h_o$ , we slightly decreased  $\lambda$  in the bootstrap sample so in effect we were overfitting (or undersmoothing in kernel literature),<sup>27</sup> see e.g. Hall (1992, Section 4.5) for theoretical justification of this undersmoothing procedure for kernel least squares regression. In the exogenous case, we know that  $\hat{h}(y_2)$  has a pointwise asymptotic normal distribution, see e.g. Theorem 2 in Newey (1997). So in this case the bootstrap yields consistent estimates of the true confidence bands, see e.g. Theorem 1.2.1 in Politis, Romano and Wolf (1999). In the endogenous case, we have no theoretical justification for the bootstrap since we have not derived a pointwise asymptotic distribution of  $\hat{h}(y_2)$ , but we conjecture that one exists. From Theorem 2 on the convergence rate of the nonparametric IV regression, we know that compared to the exogenous case, the endogenous estimates have similar asymptotic bias but a bigger variance, The reported confidence bands in the endogenous case are wider compared to those for the exogenous case, which is consistent with the theory.<sup>28</sup>

Figure C.9 to C.15 illustrate the estimation of our system of Engel curves. The plots offer a comparison of the fully nonparametric estimates vs. the semiparametric ones, and the endogenous case vs. the exogenous one. For these plots, we used a 3rd order B-spline sieve for  $h$  with number of sieve terms  $k_n = 9$ , and a 4th order B-spline

---

<sup>27</sup> Alternatively we could fix  $\lambda$  value but slightly increase the number of sieve terms  $k_n$  in the estimation of  $h$  using the bootstrap sample. We have tried this as well and the results are similar.

<sup>28</sup> Newey (1997) also supplies us with an estimator of the asymptotic variance of  $\hat{h}(y_2)$  in the exogenous case which can be used to construct alternative confidence bands. Since the endogenous case with identity weighting is simply a penalized 2SLS regression, we can easily compute an estimate of the asymptotic variance of  $\hat{h}(y_2)$  in the endogenous case, still assuming that it is asymptotically normal. This gave confidence bands very similar to the ones obtained by the bootstrap, both in the exogenous and endogenous case.

of dimension  $J_n = 15$  for  $m$ . We penalised both the level and the 2nd order derivative of  $h$ , each with penalisation weight  $\lambda = 0.4$ . All plots are with identity weighting,  $\widehat{\Sigma}(X_i) = I_L$ .<sup>29</sup> As noted earlier, the nonparametric IV estimates using the subsample of households without children should be interpreted with care. However, the plots of the estimated curves seem to be consistent with the Monte Carlo findings and appear reasonably well behaved, even for the subsample without children. Our main focus is on the lower rhs plot in each panel which represent the sieve IV estimates under the shape invariant restrictions. Several interesting features are present in the plots. As may be expected the estimated shares of alcohol and food-out for households with children are everywhere below those for households without children. As family size increases, for any given total outlay, the shares going to alcohol and food-out fall; at the same time, the share going to food-in increases. So there is a shift in expenditure shares from one set of non-durables to another when families have children. The curvature also changes significantly as we allow for endogeneity. So neglecting potential endogeneity in the estimation can lead to incorrect estimates of the Engel curve shape. The Engel curve for food-in, for example, showing a much more pronounced reverse ‘S’ shape under endogeneity with a more dramatic shift to the right in the curve resulting from the presence of children.

The semiparametric efficient estimates of  $\theta$  are given in Table 7.3. These estimates have been obtained using the same functional bases and the same  $J_n = 15$  and  $k_n = 9$  as used to obtain the Engel curves  $h$  estimates, except with a smaller  $\lambda$ . The estimates of  $\theta$  are plausibly signed in both the endogenous and the exogenous case. The differences can be assessed more formally. Let  $\widehat{\theta}_{LS}$  and  $\widehat{\theta}_{IV}$  denote the semiparametric efficient estimate of  $\theta$  under  $H_0: Y_2$  exogenous and  $H_1: Y_2$  endogenous respectively. Furthermore, let  $\widehat{V}_{LS}$  and  $\widehat{V}_{IV}$  denote the estimates of their respective variances. we then have that  $T \equiv n \left( \widehat{\theta}_{LS} - \widehat{\theta}_{IV} \right)' \left( \widehat{V}_{LS} - \widehat{V}_{IV} \right)^{-1} \left( \widehat{\theta}_{LS} - \widehat{\theta}_{IV} \right) \overset{asy.}{\sim} \chi_{L+1}^2$  under the null. This Hausman test for the exogeneity of  $Y_2$  produces a statistic of 880.06 with critical value  $\chi_8^2(95) = 15.5$ , and we reject the null-hypothesis. That is, the data supports the hypothesis that

---

<sup>29</sup>We also obtained estimates for  $h$  with  $\widehat{\Sigma}(X_i) = \widehat{\Sigma}_o(X_i)$ , these however were relatively wiggly since it was obtained with lower penalization  $\lambda$  value. Since the optimal weighted procedure is only theoretically justified for efficient estimation of  $\theta$ , we have no theoretical justification for these type of estimates of  $h$ , while economic theory suggests that engel curves  $h$  should be relatively smooth, we only present the estimated  $h$  with identity weighting. Furthermore, the identity weights allows for a fast implementation of the nonparametric bootstrap employed here.

$Y_2$  is endogenous.<sup>30</sup> The results show a strong impact on  $\theta_1$  of allowing for endogeneity. This parameter measures the general log equivalence scale for the presence of children with a couple normalised to unity. The LS estimate is implausibly low whereas the IV estimate is very plausible and represents an equivalence scale of about .45, normalised to unity for a couple without children. This is also seen in the more dramatic shift in the plotted curves between the two groups as commented on above. One can also give interpretations to the estimates of  $\theta_2$ ; e.g. the negative value of  $\theta_2$  for alcohol shows the decline in the overall alcohol budget share, given total equalised expenditure, that occurs for larger households.

Table 7.3: Efficient estimates of  $\theta$  in the exog. and endog. case

	Semiparametric IV		Semiparametric LS	
	coefficient	std. ( $10^{-3}\times$ )	coefficient	std. ( $10^{-3}\times$ )
$\theta_1$	0.3698	57.4712	0.1058	34.3810
$\theta_2$ - alcohol	-0.0216	4.5047	-0.0239	2.5322
$\theta_2$ - fares	-0.0023	2.5089	-0.0092	1.4027
$\theta_2$ - food-in	0.0213	6.5406	0.0461	4.8861
$\theta_2$ - food-out	0.0006	3.6744	-0.0046	2.4182
$\theta_2$ - fuel	-0.0035	2.7611	0.0054	1.9069
$\theta_2$ - leisure	0.0388	10.9148	-0.0016	6.2392
$\theta_2$ - travel	-0.0384	5.9912	-0.0226	3.9748

To check the robustness of our estimates with respect to the choices of sieve basis functions  $B^{k_n}$  and  $p^{J_n}$ , we also tried to approximate  $h$  with a 2nd and a 3rd order polynomial spline of dimension  $k_n = 5$  to 14, and to approximate  $m$  with Fourier series and 4th order B-splines with  $J_n = 15$  to 27. The estimates  $\hat{\theta}$  are very similar to the ones reported in Table 7.3, and are also stable as the number of spline sieve terms  $k_n$  increased in both the exogenous and the endogenous cases. Inspection of the associated plots for the Engel curves  $h$  also showed that the overall shape and turning points were maintained for these alternative sieve approximations. These findings are consistent with our Monte Carlo results. To conserve space, here we only report a small sensitivity check in terms of  $\theta$  estimates under the endogeneity. Although the  $\theta$  estimates reported

<sup>30</sup>Observe that the above test statistic is not optimal in the sense that it might have a higher than acceptable probability of accepting  $H_0$  when the hypothesis is not true; this owes to the fact that we only use  $\theta$  in the test statistics, not using the information contained in  $h$ . This however is not a problem here given that we reject  $H_0$ .

in Table 7.4 are obtained using smaller penalization  $\lambda$  than those reported in Table 7.3, the estimated values are virtually the same.

Table 7.4: Estimates of  $\theta$ : Sensitivity Analysis

	Semiparametric IV							
	$\theta$ estimates							
$k_n, J_n$	Alcohol	Fares	Food-in	Food-out	Fuel	Leisure	Travel	$\theta_1$
B-spl 9, B-spl 15	-0.0210	-0.0019	0.0207	0.0003	-0.0038	0.0422	-0.0393	0.3834
B-spl 9, B-spl 25	-0.0233	-0.0005	0.0171	-0.0005	-0.0027	0.0489	-0.0419	0.4113
B-spl 8, Cos 20	-0.0248	-0.0009	0.0204	-0.0005	-0.0029	0.0458	-0.0387	0.3989
P-spl 8, B-spl 15	-0.0222	-0.0004	0.0209	-0.0019	-0.0029	0.0429	-0.0359	0.3981
P-spl 5, B-spl 25	-0.0285	-0.0011	0.0191	0.0002	-0.0038	0.0496	-0.0399	0.4088

Finally, we note the results of a number of further comparisons that we carried out. The first implemented the control-function approach of Newey, Powell and Vella (1999). As one might expect, this gave estimates that lay between our sieve IV-estimates and the exogenous estimates. Second, we compared our semi-nonparametric model with a parametric quadratic model of the QUAIDS-class proposed in Banks, Blundell and Lewbel (1997). A test on  $\theta$  rejected the QUAIDS-model in favour of the semi-nonparametric model. It is not too surprising that the QUAIDS model does not fit data as well as our more flexible model. Finally, we implemented the sieve minimum distance procedure using three different years of FES data sets, and have tried both gross earnings and disposable income as the instrument. We have also considered different sieve basis functions and different number of sieve terms to approximate unknown Engel curves, the unknown conditional means and the optimal weighting matrix. The empirical findings in these final comparisons are surprisingly robust to those presented in the paper in the sense of being qualitatively similar and are available from the authors on request.

## 8. Conclusions

Endogeneity of explanatory variables in regression analysis is a central feature of economic relationships. With the development of semiparametric and nonparametric estimation methods there is an increasing demand for generalizations that allow for endogeneity. In this paper we have considered the sieve semi- /nonparametric IV estimation of the shape-invariant Engel curves with endogenous total expenditure. In this paper we have provided identification and established the nonparametric convergence rate and semiparametric efficiency properties of our estimators under relatively “low-level” suf-

ficient conditions. We have also presented Monte Carlo simulation results which shed some lights on the choice of smoothing parameters and the performance of the sieve nonparametric IV estimator. The simulation results indicate it is the variance part (not the bias part) that behaves poorly in the sieve nonparametric IV regression, which is on the contrary to the poor bias problem in the classical parametric IV regression under weak instruments.

In our application to the UK Family Expenditure Survey (FES) we have shown the importance of allowing for endogeneity and documented the relatively simple steps involved in implementing the sieve semi-nonparametric IV approach. The shape-invariant system of Engel curve specification, that pools across demographic groups, has enabled us to estimate the parametric effects of equivalence scales and the demographic impacts accurately and efficiently. We found the estimated curves and demographic parameters to be plausible and we have documented a significant impact of accounting for the endogeneity of total expenditure. Adjusting for endogeneity increases the common demographic shift parameter and produces a much more plausible estimate of the income equivalence scale. We have also contrasted our estimator with that which assumes exogenous total expenditure. It appears that the nonlinear behavior in the share Engel curve is systematically different under the exogeneity assumption. Our application illustrates the importance of utilizing the semi-nonparametric restrictions and suggests that it would be worthwhile to further investigate the imposition of restrictions derived from economic theory in identification and estimation of econometric models, see Matzkin (1994) for example.

## A. Proofs and Technical Lemmas

**Proof. (Theorem 1)** Without loss of generality and given assumption (4), we can assume that  $X_{1i}$  is a scalar dummy random variable (i.e.,  $X_{1i} \in \{0, 1\}$ ). First conditioning on  $X_{1i} = 0$ , we have:

$$E[Y_{1il} - h_l(Y_{2i} - \phi(0)) | X_{1i} = 0, X_{2i}] = 0 \quad \text{for } l = 1, \dots, L,$$

since  $\phi$  is known, assumptions (1) and (2) imply:

$$h_l = h_{ol} \text{ almost surely, for } l = 1, \dots, L.$$

Now  $h_{ol}(\cdot)$  is identified. This together with conditional moment restriction (3.2) and assumptions (3) and (4) identify  $\theta_{o1}$  and  $\theta_{o2l}$ . Since for all  $l = 1, \dots, L$ ,

$$E[Y_{1il} - h_{ol}(Y_{2i} - \phi(X_{1i}\theta_1)) - X_{1i}\theta_{2l} | X_{1i} = 1, X_{2i}] = 0,$$

this and (3.2) imply:

$$E[h_{ol}(Y_{2i} - \phi(\theta_{o1})) - h_{ol}(Y_{2i} - \phi(\theta_1)) + (\theta_{o2l} - \theta_{2l}) | X_{1i} = 1, X_{2i}] = 0 \quad (\text{A.1})$$

Since there is a  $l^*$  with  $h_{ol^*}(\cdot)$  nonlinear and differentiable, we have

$$E[\nabla h_{ol^*}(Y_{2i} - \phi(\bar{\theta}_{o1})) | X_{1i} = 1, X_{2i}] \times (\phi(\theta_1) - \phi(\theta_{o1})) + (\theta_{o2l^*} - \theta_{2l^*}) = 0$$

where  $\bar{\theta}_{o1}$  is some value between  $\theta_{o1}$  and  $\theta_1$ . Again by (1) and (2) and  $\nabla h_{ol^*}(Y_{2i} - \phi(\bar{\theta}_{o1})) \neq \text{const}$  (in particular  $\neq 0$ ) imply that  $E[\nabla h_{ol^*}(Y_{2i} - \phi(\bar{\theta}_{o1})) | X_{1i} = 1, X_{2i}] \neq \text{const}$  (in particular  $\neq 0$ ), hence

$$\phi(\theta_1) - \phi(\theta_{o1}) = 0 \text{ and } \theta_{o2l^*} - \theta_{2l^*} = 0$$

by (5) we have  $\theta_1 - \theta_{o1} = 0$ , this together with (A.1) implies  $\theta_{o2l} - \theta_{2l} = 0$  for  $l = 1, \dots, L$ .  $\blacksquare$

**Proof. (Proposition 1)** Notice that our estimation method (3.3) for endogenous expenditure ( $Y_2$ ) case is the same as that in Newey and Powell (2003) for nonparametric IV regression. We can apply their theorem 4.1 to establish the consistency of  $\hat{h}$  to  $h_o$  in other metrics, (for simplicity we set  $L = 1$ ,  $\phi(\cdot) = 0$  and  $\theta_{2l} = 0$  and consider the purely nonparametric IV regression model). Their assumption 1 is implied by our assumption 1(iii). Their assumption 2 is satisfied with our assumptions 3, 4(i) and 5(i). For their assumption 4 of compact parameter space  $\mathcal{H}$ , we notice that by the weighted compact embedding result in Chen, Hansen and Scheinkman (1997), a bounded ball in  $W_\infty^r(\mathcal{R})$  is compact under the norm

$$\|h\|_{W_\infty^{r_1, \omega}} = \|h\omega\|_{W_\infty^{r_1}} \text{ for } \omega(y) \equiv (1 + y^2)^{-a/2}, a > 0, r_1 < r.$$

For instance,

$$\|h\|_{W_{\infty,\omega}^0} = \sup_y \left| h(y)(1+y^2)^{-a/2} \right| \equiv \|h\|_{\infty,\omega}.$$

Hence their assumption 4 is satisfied given our assumption 2 and our parameter space  $\mathcal{H}$  in (3.9) with the metric  $\|\cdot\|_{W_{\infty,\omega}^{r_1}}$  for  $0 \leq r_1 < r$ . For their assumption 3 of  $\rho(Z, \alpha)$  being Holder continuous in  $h \in \mathcal{H}$ , noting that  $\rho_l(Z, \alpha) = Y_{1l} - h_l(Y_2 - \phi(X_1'\theta_1)) - X_1'\theta_{2l}$ , we have  $E[|\rho_l(Z, \alpha_o)|^2|X]$  is bounded given our assumptions 1(ii) and 2. Also for any fixed  $a > 0, 0 \leq r_1 < r$ , and for any  $h, \tilde{h} \in \mathcal{H}$ ,

$$\begin{aligned} |\rho_l(Z, h) - \rho_l(Z, \tilde{h})| &= |h_l(Y_2) - \tilde{h}_l(Y_2)| \\ &\leq [1 + (Y_2)^2]^{a/2} \times \sup_y \left| [h_l(y) - \tilde{h}_l(y)](1+y^2)^{-a/2} \right| \\ &\leq [1 + (Y_2)^2]^{a/2} \times \left\| h_l - \tilde{h}_l \right\|_{W_{\infty,\omega}^{r_1}}, \end{aligned}$$

assuming  $E\{[1 + (Y_2)^2]^a|X\}$  is bounded, then their assumption 3 is satisfied. Finally their assumption 5 is satisfied with our sieve space  $\mathcal{H}_n$  (3.10) and assumption 5(i). Now we can apply Newey and Powell's (2003) theorem 4.1 to obtain  $\left\| \hat{h}_l - h_{ol} \right\|_{W_{\infty,\omega}^{r_1}} = o_p(1)$  for any  $r_1 \in [0, r)$ . Finally, since

$$\left\| \hat{h}_l - h_{ol} \right\|_{Y_2} \leq \left[ \sup_y \left| [\hat{h}_l(y) - h_{ol}(y)](1+y^2)^{-a/2} \right| \right] \sqrt{\int (1+y^2)^a f_{Y_2}(y) dy},$$

hence  $\left\| \hat{h}_l - h_{ol} \right\|_{Y_2} = o_p(1)$ . ■

**Proof. (Theorem 2):** It suffices to establish the result for the purely nonparametric IV regression model  $E[h_{ol}(Y_2 - \phi(0))|X_1 = 0, X_2] = E[Y_{1l}|X_1 = 0, X_2]$  for an arbitrarily fixed  $l = 1, \dots, L$ . To simplify notations further, we assume  $\phi(0) = 0$  and suppress the conditioning variable  $X_1 = 0$  and drop the subscript  $l$ . We denote  $\hat{h} = \arg \min_{h \in \mathcal{H}_n} \frac{1}{n} \sum_{i=1}^n \{\hat{m}(X_{2i}, h)\}^2$  where

$$\mathcal{H}_n = \left\{ h(Y_2) = \sum_{k=0}^{K_n} \sum_{j \in \mathcal{K}_n} \pi_{kj} \psi_{kj}(Y_2) : 0 \leq h \leq 1, \|\nabla^r h\|_{\infty} \leq c \right\},$$

and

$$\begin{aligned}
\widehat{m}(X_{2i}, h) &= \sum_{t=1}^n \{Y_{1t} - h(Y_{2t})\} p^{J_n}(X_{2t})' (P'P)^{-1} p^{J_n}(X_{2i}) \\
&= \sum_{t=1}^n \{Y_{1t} - \sum_{k=0}^{K_n} \sum_{j \in \mathcal{K}_n} \pi_{kj} \psi_{kj}(Y_{2t})\} p^{J_n}(X_{2t})' (P'P)^{-1} p^{J_n}(X_{2i}) \\
&= \widehat{E}[Y_1 | X_{2i}] - \sum_{k=0}^{K_n} \sum_{j \in \mathcal{K}_n} \pi_{kj} \widehat{E}[\psi_{kj}(Y_2) | X_{2i}] = \widehat{E}[Y_1 | X_{2i}] - \widehat{E}[h(Y_2) | X_{2i}].
\end{aligned}$$

In the following we denote  $E_{n, X_2}\{f\} \equiv \frac{1}{n} \sum_{i=1}^n \{f(X_{2i})\}$ ,  $\langle g, f \rangle_{n, X_2} = E_{n, X_2}\{g(X_{2i})f(X_{2i})\}$ ,  $\|f\|_{n, X_2} = \sqrt{\langle f, f \rangle_{n, X_2}}$  and  $\|f\|_{X_2} = \sqrt{E\{f(X_2)\}^2}$ . Also let  $g_o(X_2) \equiv E[Y_1 | X_2]$ . Then  $Th_o = g_o$  by our theorem 1. Let  $\widehat{g}(X_{2i}) \equiv \widehat{E}[Y_1 | X_{2i}]$  and  $(\widehat{T}h)(X_{2i}) \equiv \widehat{E}[h(Y_2) | X_{2i}]$ . Then  $\widehat{h} = \arg \min_{h \in \mathcal{H}_n} \|\widehat{T}h - \widehat{g}\|_{n, X_2}^2$ , which is the solution to

$$\text{find } \widehat{h} \in \mathcal{H}_n \text{ such that } \langle \widehat{T}\widehat{h}, \widehat{T}h \rangle_{n, X_2} = \langle \widehat{g}, \widehat{T}h \rangle_{n, X_2} \quad \text{for all } h \in \mathcal{H}_n.$$

Let  $\mathcal{G} \equiv \{g \in W_\infty^{r_m}(\mathcal{X}_2) : \|g\|_{W_\infty^{r_m}} \leq \text{const}\}$ . Then by Assumptions 1(iii) and 4, we have  $Th_o = g_o \in \mathcal{G}$ . Also by the definition of  $\mathcal{H}_n$  and Assumption 4, we have  $Th \in \mathcal{G}$  for all  $h \in \mathcal{H}_n$ . We shall establish the following four claims later:

**Claim 1:** (i) under assumption 2(ii) and  $E[|Y_2|^{2a}] < \infty$  for some  $a > r > 0$ , and the sieve space  $\mathcal{H}_n$  (3.10), we have: there is a finite  $c > 0$  such that for any  $h \in \mathcal{H}$ , there is a  $\Pi_n h \in \mathcal{H}_n$  satisfying  $\|h - \Pi_n h\|_{Y_2} \leq c(k_n)^{-r}$ ;

(ii) under assumptions 3 and 4, we have: there is a finite  $c > 0$  such that for any  $g \in \mathcal{G}$ , there is a  $p^{J_n}(X_2)'A$  such that  $\|g - p^{J_n}(X_2)'A\|_{X_2} \leq c(J_n)^{-r_m}$ .

**Claim 2:** under assumptions 1, 2(ii), 3, 4 and 5(ii), we have:

$$\begin{aligned}
(i) \quad \|g_o - \widehat{g}\|_{X_2} &= O_p \left( J_n^{-r_m} + \sqrt{\frac{J_n}{n}} \right); \\
(ii) \quad \sup_{h \in \mathcal{H}_n} \|\{\widehat{T} - T\}h\|_{X_2} &= O_p \left( J_n^{-r_m} + \sqrt{\frac{J_n}{n}} \right); \\
(iii) \quad \sup_{h \in \mathcal{H}_n} \|\widehat{m}(\cdot, h) - m(\cdot, h)\|_{X_2} &= O_p \left( J_n^{-r_m} + \sqrt{\frac{J_n}{n}} \right).
\end{aligned}$$

**Claim 3:** under assumptions 1, 2(ii), 3, 4 and 5(ii), we have: there exist constants  $c_1, c_2 > 0$  such that

$$c_1 \|\widehat{m}(\cdot, h)\|_{X_2}^2 \leq \|\widehat{m}(\cdot, h)\|_{n, X_2}^2 \leq c_2 \|\widehat{m}(\cdot, h)\|_{X_2}^2,$$

uniformly over  $h \in \mathcal{H}_n$ , except on an event whose probability tends to zero as  $n \rightarrow \infty$ .

**Claim 4:** under assumptions 1, 2(ii) and 6(i) and the sieve space  $\mathcal{H}_n$  (3.10), we have:

- (i)  $\|T\{\Pi_n h_o - h_o\}\|_{X_2} \leq \text{const.} \times \mu_{k_n} \times \|\Pi_n h_o - h_o\|_{Y_2}$ ;
- (ii)  $\tau_n \leq 1/\mu_{k_n}$ .

We defer the proofs of these four claims later. Now denote  $\|h\|_{Y_2} \equiv \sqrt{E\{h(Y_2)\}^2}$ . Then by the triangular inequality and assumption 2,

$$\|\widehat{h} - h_o\|_{Y_2} \leq \|h_o - \Pi_n h_o\|_{Y_2} + \|\widehat{h} - \Pi_n h_o\|_{Y_2}.$$

Next by the definition of  $\tau_n$  and the triangular inequality,

$$\begin{aligned} \|\widehat{h} - \Pi_n h_o\|_{Y_2} &\leq \tau_n \times \|T\{\widehat{h} - \Pi_n h_o\}\|_{X_2} \\ &\leq \tau_n \times \{\|T\widehat{h} - \widehat{g}\|_{X_2} + \|\widehat{g} - g_o\|_{X_2} + \|Th_o - T\Pi_n h_o\|_{X_2}\} \\ &\leq \tau_n \times \{\|T - \widehat{T}\|\widehat{h}\|_{X_2} + \|\widehat{T}\widehat{h} - \widehat{g}\|_{X_2} + \|\widehat{g} - g_o\|_{X_2} + \|T\{h_o - \Pi_n h_o\}\|_{X_2}\}. \end{aligned}$$

Under Claim 3 and by the definition of  $\widehat{h}$ , we have:

$$\begin{aligned} \|\widehat{T}\widehat{h} - \widehat{g}\|_{X_2} \{1 + o_p(1)\} &= \|\widehat{T}\widehat{h} - \widehat{g}\|_{n, X_2} \\ &\leq \|\widehat{T}\Pi_n h_o - \widehat{g}\|_{n, X_2} = \|\widehat{T}\Pi_n h_o - \widehat{g}\|_{X_2} \{1 + o_p(1)\}. \end{aligned}$$

Now by the definitions of  $\widehat{m}$  and  $m$ , and the triangular inequality, we have:

$$\begin{aligned} \|\widehat{T}\Pi_n h_o - \widehat{g}\|_{X_2} &= \|\widehat{m}(\cdot, \Pi_n h_o)\|_{X_2} \\ &\leq \|\widehat{m}(\cdot, \Pi_n h_o) - m(\cdot, \Pi_n h_o)\|_{X_2} + \|m(\cdot, \Pi_n h_o) - m(\cdot, h_o)\|_{X_2} \\ &= \|\widehat{m}(\cdot, \Pi_n h_o) - m(\cdot, \Pi_n h_o)\|_{X_2} + \|T\{h_o - \Pi_n h_o\}\|_{X_2}. \end{aligned}$$

These and Claim 2 imply

$$\|\widehat{h} - \Pi_n h_o\|_{Y_2} \leq \tau_n \times \left\{ O_p \left( J_n^{-r_m} + \sqrt{\frac{J_n}{n}} \right) + \|T\{h_o - \Pi_n h_o\}\|_{X_2} \times \{2 + o_p(1)\} \right\}.$$

This and Claim 4 imply

$$\begin{aligned} \|\widehat{h} - h_o\|_{Y_2} &\leq \|h_o - \Pi_n h_o\|_{Y_2} + \tau_n \times O_p \left\{ J_n^{-r_m} + \sqrt{\frac{J_n}{n}} \right\} + \tau_n \times O \{ \mu_{k_n} \|h_o - \Pi_n h_o\|_{Y_2} \} \\ &\leq O \left\{ \|h_o - \Pi_n h_o\|_{Y_2} + \tau_n \times [J_n^{-r_m} + \sqrt{\frac{J_n}{n}}] \right\}. \end{aligned}$$

This, assumptions 2 and 5, and Claim 1(i) imply:

$$\|\widehat{h} - h_o\|_{Y_2} \leq \text{const.} (k_n)^{-r} + \tau_n \times O_p \left\{ \sqrt{\frac{k_n}{n}} + (k_n)^{-r_m} \right\}.$$

We now finish the proof of Theorem 2 by establishing the four claims.

**Proof of Claim 1:** (i) Under Assumption 2(ii) and given the sieve space  $\mathcal{H}_n$ , we have for any  $h \in \mathcal{H}$ , there exists  $\Pi_n h \in \mathcal{H}_n$  such that for any fixed  $a > r > 0$ , and  $c > 0$ ,

$$\sup_y \left| [h(y) - \Pi_n h(y)](1 + y^2)^{-a/2} \right| \leq c(k_n)^{-r},$$

see e.g. Chen, Hansen and Scheinkman (1997). Now let  $f_{Y_2}(y)$  denote the density of  $Y_2$ , then:

$$\begin{aligned} & \int [h(y) - \Pi_n h(y)]^2 f_{Y_2}(y) dy \\ &= \int [h(y) - \Pi_n h(y)]^2 (1 + y^2)^{-a} [(1 + y^2)^a f_{Y_2}(y)] dy \\ &\leq \left[ \sup_y \left| [h(y) - \Pi_n h(y)](1 + y^2)^{-a/2} \right| \right]^2 \int (1 + y^2)^a f_{Y_2}(y) dy \end{aligned}$$

hence

$$\begin{aligned} \|h - \Pi_n h\|_{Y_2} &\approx \sqrt{\int [h(y) - \Pi_n h(y)]^2 f_{Y_2}(y) dy} \\ &\leq C \sup_y \left| [h(y) - \Pi_n h(y)](1 + y^2)^{-a/2} \right| \leq C'(k_n)^{-r}. \end{aligned}$$

(ii) See Timan (1963) for Fourier series and Schumaker (1981) for spline sieve.

**Proof of Claim 2:** (i) By our assumption 1, 2 and 4,  $g_o = Th_o \in \mathcal{G}$ , this, assumptions 3 and 5(ii) imply that all the conditions of Theorem 1 in Newey (1997, p.150) are satisfied with his  $d = 0$ , his  $K =$  our  $J_n$ , his  $\zeta_0(K) =$  our  $\sqrt{J_n}$ , and his  $K^{-\alpha} =$  our  $J_n^{-r_m}$ , hence we obtain result (i).

(ii) By the definition of  $\mathcal{H}_n$  and Assumption 4, we have  $Th \in \mathcal{G}$  for all  $h \in \mathcal{H}_n$ . Moreover, since  $0 \leq h \leq 1$  for all  $h \in \mathcal{H}_n$ , we have that  $Var\{h(Y_2)|X_1 = 0, X_2\} \leq 1$  for all  $h \in \mathcal{H}_n$ . Note that  $\hat{T}h$  is simply the sieve LS regression of  $h(Y_2)$  on  $p^{J_n}(X_2)$ . We now go through the proof of Theorem 1 in Newey (1997, p.161-163), and see that Newey's result (with his  $d = 0$ ) actually holds uniformly over  $h \in \mathcal{H}_n$ , hence we obtain result (i).

(iii) Directly follows from (i) and (ii).

**Proof of Claim 3:** By the definition of  $\mathcal{H}_n$  and Assumptions 1(iii), 2(ii) and 4, we have  $m(\cdot, h) \in \mathcal{G}$  for all  $h \in \mathcal{H}_n$ . Moreover, since  $0 \leq Y_1 \leq 1$  and  $0 \leq h \leq 1$  for all  $h \in \mathcal{H}_n$ , we have that  $Var\{Y_1 - h(Y_2)|X_1 = 0, X_2\} \leq 1$  for all  $h \in \mathcal{H}_n$ . Note that  $\hat{m}(\cdot, h)$  is simply the sieve LS regression of  $Y_1 - h(Y_2)$  on  $p^{J_n}(X_2)$ , hence  $\hat{m}(\cdot, h)$  belongs to the closed linear span of  $p^{J_n}(X_2)$  with probability approaching to one. Now we go through the proof of Lemma 4 in Huang (1998) with his  $A_n =$  our  $\sqrt{J_n}$  and his  $N_n =$  our  $J_n$ .

Under our Assumptions 1, 2(ii), 3, 4 and 5(ii), we notice that Huang's result actually holds uniformly over  $h \in \mathcal{H}_n$ , hence we obtain Claim 3.

**Proof of Claim 4:** By assumption 6(i), there are orthonormal bases  $\{\phi_{0k} : k = 1, \dots, \infty\}$  (for  $W_\infty^{r_m}(\mathcal{X}_2)$  and  $L_2(\mathcal{X}_2, f_{0,X_2})$ ),  $\{\phi_{1k} : k = 1, \dots, \infty\}$  (for  $W_\infty^r(\mathcal{R})$  and  $L_2(\mathcal{R}, f_{0,Y_2})$ ), and corresponding singular numbers  $\{\mu_k\}$  with  $\mu_1 = 1 \geq \mu_2 \geq \mu_3 \geq \dots$  such that

$$\frac{f_{0,X_2,Y_2}(x_2, y_2)}{f_{0,X_2}(x_2)f_{0,Y_2}(y_2)} = \sum_{k=1}^{\infty} \mu_k \phi_{0k}(x_2) \phi_{1k}(y_2).$$

Next, under Assumption 2, the sieve space  $\mathcal{H}_n$  (3.10) for  $h$  is a Riesz (or a frame) basis for  $L_2(\mathcal{R}, f_{0,Y_2})$  i.e., which is equivalent to an orthogonal basis for  $L_2(\mathcal{R}, f_{0,Y_2})$ , see e.g. Chui (1992) and Meyer (1992). In particular, with  $k_n \equiv 2^{K_n}$ , the sieve space  $\mathcal{H}_n$  (3.10) is equivalent to the linear span of the orthonormal basis  $\{\phi_{1k} : k = 1, \dots, k_n\}$ . Thus we have the alternative representation:

$$h(y_2) = \sum_{k=1}^{k_n} \langle h, \phi_{1k} \rangle_{Y_2} \phi_{1k}(y_2) \quad \text{for any } h \in \mathcal{H}_n.$$

(i) Since

$$h_o(y_2) - \Pi_n h_o(y_2) = \sum_{k=k_n+1}^{\infty} \langle h_o, \phi_{1k} \rangle_{Y_2} \phi_{1k}(y_2).$$

Hence

$$\begin{aligned} T\{h_o - \Pi_n h_o\}(x_2) &\equiv E[h_o(Y_2) - \Pi_n h_o(Y_2) | X_2 = x_2] \\ &= \int (h_o(y_2) - \Pi_n h_o(y_2)) \frac{f_{0,X_2,Y_2}(x_2, y_2)}{f_{0,X_2}(x_2)f_{0,Y_2}(y_2)} f_{0,Y_2}(y_2) dy_2 \\ &= \sum_{k=k_n+1}^{\infty} \langle h_o, \phi_{1k} \rangle_{Y_2} \times \mu_k \phi_{0k}(x_2), \end{aligned}$$

and

$$\begin{aligned} \|T\{\Pi_n h_o - h_o\}\|_{X_2}^2 &= \sum_{k=k_n+1}^{\infty} \{\langle h_o, \phi_{1k} \rangle_{Y_2}\}^2 \times \mu_k^2 \\ &\leq \{\mu_{k_n}\}^2 \times \sum_{k=k_n+1}^{\infty} \{\langle h_o, \phi_{1k} \rangle_{Y_2}\}^2 \\ &= \{\mu_{k_n}\}^2 \times \|h_o - \Pi_n h_o\|_{Y_2}^2. \end{aligned}$$

Therefore,

$$\|T\{\Pi_n h_o - h_o\}\|_{X_2} \leq \mu_{k_n} \times \|h_o - \Pi_n h_o\|_{Y_2}.$$

(ii) For any  $h \in \mathcal{H}_n$  with  $h \neq 0$ ,

$$T\{h\}(X_2) = E\{h(Y_2)|X_2\} = \sum_{j=1}^{k_n} \mu_j \langle h, \phi_{1j} \rangle_{Y_2} \phi_{0j}(X_2)$$

$$\begin{aligned} \|Th\|_{X_2}^2 &= \sum_{j=1}^{k_n} \mu_j^2 \{\langle h, \phi_{1j} \rangle_{Y_2}\}^2 \\ &\geq (\mu_{k_n})^2 \sum_{j=1}^{k_n} \{\langle h, \phi_{1j} \rangle_{Y_2}\}^2 = (\mu_{k_n})^2 \|h\|_{Y_2}^2 \end{aligned}$$

Hence

$$\tau_n \equiv \sup_{h \in \mathcal{H}_n: h \neq 0} \frac{\|h\|_{Y_2}}{\|Th\|_{X_2}} \leq \frac{1}{\mu_{k_n}}.$$

■

**Proof. (Proposition 2):** Under Assumption 7(ii), it suffices to consider one good  $l$  and identity weighting matrix  $\Sigma = I$ . Moreover given our identification theorem 1, it suffices to study the convergence rate of  $\hat{h}_l()$  to  $h_{ol}()$  conditioning on  $X_1 = 0$ . We obtain the result by verifying assumptions 3.1 - 3.9 of theorem 3.1 in Ai and Chen (2003). Their assumptions 3.1, 3.3 and 3.4 are respectively satisfied by our Assumptions 1(i), 3(iii), 1(iii) and 7. Their assumption 3.2 is satisfied by our Assumptions 3, 4 and 5. Note that our Assumptions 2, 5(iv) and 6(i), the sieve space (3.10), the Claims 1(i) and 4(i) imply  $\|\alpha - \Pi_n \alpha\| \leq \mu_{k_n} \|\alpha - \Pi_n \alpha\|_{Y_2} = O\{\mu_{k_n} (k_n)^{-r}\} = o(n^{-1/4})$  since  $\mu_{k_n} = o(1)$  and  $k_n \equiv 2^{K_n}$ . hence Ai-Chen's assumption 3.5 is satisfied. For Ai-Chen's assumption 3.6(ii), since  $|\rho_l(Y, 0, X_2, \alpha)| \leq |Y_{1l}| + |h_l(Y_2 - \phi(0))| \leq C$  a.s., we have that  $\rho_l(Z_i, \alpha)$  satisfies the envelope condition 3.6(ii) with  $c_1(Z) = \text{const.}$  and  $p = \infty$ . Notice that for any  $h_{1l}, h_{2l} \in \mathcal{H}_{l,n}$  with  $\mathcal{H}_{l,n}$  given in (3.10),  $\omega(y) = (1 + y^2)^{-a/2}$  and  $\sup_y |h(y)\omega(y)| \equiv \|h\|_{\infty, \omega}$ ,

$$|h_{1l}(Y_2) - h_{2l}(Y_2)| \leq [\omega(Y_2)]^{-1} \|h_{1l} - h_{2l}\|_{\infty, \omega},$$

hence their Hölder condition 3.6(i) is satisfied. Next Ai-Chen's assumption 3.6(iii) is satisfied by our Assumption 4.

Ai-Chen's assumption 3.7 is satisfied with our Assumption 5 with their  $k_n$  = our  $J_n$ , their  $k_{1n}$  = our  $2^{K_n}$  and  $\xi_{0n} = \sqrt{J_n}$ . Ai-Chen's assumption 3.8 is satisfied with the sieve space (3.10), their  $\kappa = 1$ ,  $\|\cdot\|_s = \|\cdot\|_{\infty, \omega}$ , and  $\ln N(\varepsilon^{1/\kappa}, \mathcal{A}_n, \|\cdot\|_{\infty, \omega}) \leq \text{const} \times k_{1n} \ln(\varepsilon^{-1})$  and their  $k_{1n}$  = our  $2^{K_n}$ . Finally, conditioning on  $X_1 = 0$ , and under our Assumption 1(iii), we have:

$$m_l(0, X_2, \alpha) = E[h_{ol}(Y_2 - \phi(0)) - h_l(Y_2 - \phi(0))|X_1 = 0, X_2],$$

$$\|\alpha - \alpha_o\|^2 = E \left[ E \left[ \frac{d\rho_l(Z, \alpha_o)}{d\alpha} [\alpha - \alpha_o] | X_1 = 0, X_2 \right]' E \left[ \frac{d\rho_l(Z, \alpha_o)}{d\alpha} [\alpha - \alpha_o] | X_1 = 0, X_2 \right] \right],$$

$$E \left\{ \frac{d\rho_l(Z, \alpha_o)}{d\alpha} [\alpha - \alpha_o] | X_1 = 0, X_2 \right\} = -E [h_l(Y_2 - \phi(0)) - h_{ol}(Y_2 - \phi(0)) | X_1 = 0, X_2].$$

Hence  $\|\alpha - \alpha_o\|^2$  and  $E\{m_l(0, X_2, \alpha)'m_l(0, X_2, \alpha)\}$  are equivalent, and Ai-Chen's assumption 3.9 is satisfied. Now by theorem 3.1 of Ai and Chen (2003), we obtain the convergence rates  $\|\hat{\alpha} - \alpha_o\| = o_p(n^{-1/4})$ . ■

**Lemma 1.** *Suppose that  $\Sigma(X)$  is positive definite for all  $X$ ; that  $X'_1$  and  $\{E[\nabla h_o(Y_2 - \phi(X'_1\theta_{o1})) | X] \nabla \phi(X'_1\theta_{o1}) X'_1\}$  belong to  $L_2(\mathcal{X}, \Sigma)$ . Then: (i)*

$$E[w^{*1}(Y_2 - \phi(X'_1\theta_{o1})) | X] = -\text{Proj}_{\overline{\mathcal{S}}} (E[\nabla h_o(Y_2 - \phi(X'_1\theta_{o1})) | X] \nabla \phi(X'_1\theta_{o1}) X'_1), \quad (\text{A.2})$$

and

$$E[w^{*2l}(Y_2 - \phi(X'_1\theta_{o1})) | X] = \text{Proj}_{\overline{\mathcal{S}}} (e_l X'_1), \quad l = 1, \dots, L, \quad (\text{A.3})$$

belong to  $\overline{\mathcal{S}}$ , and solve the variational problem (5.1). Moreover,

$$D_{w^{*1}}(X, \alpha_o) = E[\nabla h_o(Y_2 - \phi(X'_1\theta_{o1})) | X] \nabla \phi(X'_1\theta_{o1}) X'_1 + E[w^{*1}(Y_2 - \phi(X'_1\theta_{o1})) | X], \quad (\text{A.4})$$

and

$$D_{w^{*2l}}(X, \alpha_o) = -e_l X'_1 + E[w^{*2l}(Y_2 - \phi(X'_1\theta_{o1})) | X], \quad l = 1, \dots, L, \quad (\text{A.5})$$

belong to  $\overline{\mathcal{S}}^\perp$ , and  $E[D_{w^*}(X, \alpha_o)' \Sigma(X)^{-1} D_{w^*}(X, \alpha_o)]$  is finite.

(ii) Suppose further Assumption N1 holds. Then  $E[D_{w^*}(X, \alpha_o)' \Sigma(X)^{-1} D_{w^*}(X, \alpha_o)]$  is positive-definite.

**Proof.** For (i), it is easy to verify that for all  $j = 1, 2, \dots, 2L$  and any  $w^j$  with  $E[w^j(\cdot) | X] \in \overline{\mathcal{S}}$ ,

$$\begin{aligned} & E [D_{w^j}(X, \alpha_o)' \Sigma(X)^{-1} D_{w^j}(X, \alpha_o)] - E[D_{w^{*j}}(X, \alpha_o)' \Sigma(X)^{-1} D_{w^{*j}}(X, \alpha_o)] \\ &= E [(D_{w^j}(X, \alpha_o) - D_{w^{*j}}(X, \alpha_o))' \Sigma(X)^{-1} (D_{w^j}(X, \alpha_o) - D_{w^{*j}}(X, \alpha_o))] \geq 0 \end{aligned}$$

This is because by definition,  $E[w^{*j}(Y_2 - \phi(X'_1\theta_{o1})) | X] \in \overline{\mathcal{S}}$  and  $D_{w^{*j}}(X, \alpha_o) \in \overline{\mathcal{S}}^\perp$  and

$$D_{w^j}(X, \alpha_o) - D_{w^{*j}}(X, \alpha_o) = E[w^j(Y_2 - \phi(X'_1\theta_{o1})) | X] - E[w^{*j}(Y_2 - \phi(X'_1\theta_{o1})) | X] \in \overline{\mathcal{S}}$$

hence

$$E [D_{w^{*j}}(X, \alpha_o)' \Sigma(X)^{-1} (D_{w^j}(X, \alpha_o) - D_{w^{*j}}(X, \alpha_o))] = 0.$$

The finiteness of  $E[D_{w^*}(X, \alpha_o)' \Sigma(X)^{-1} D_{w^*}(X, \alpha_o)]$  immediately follows from that  $X'_1$  and  $\{E[\nabla h_o(Y_2 - \phi(X'_1\theta_{o1})) | X] \nabla \phi(X'_1\theta_{o1}) X'_1\}$  belong to  $L_2(\mathcal{X}, \Sigma)$ , and that the expressions for  $D_{w^*}(X, \alpha_o)$  given in (A.4) and (A.5).

For (ii), since  $\Sigma(X)$  is finite, positive definite for all  $X$ , we have  $E[D_{w^*}(X, \alpha_o)' \Sigma(X)^{-1} D_{w^*}(X, \alpha_o)]$  is singular if and only if there is a non-zero  $\beta = (\beta'_1, \beta'_{21}, \dots, \beta'_{2L})' \in \mathcal{R}^b$  with

$$D_{w^*}(X, \alpha_o)\beta \equiv D_{w^{*1}}(X, \alpha_o)\beta_1 + D_{w^{*21}}(X, \alpha_o)\beta_{21} + \dots + D_{w^{*2L}}(X, \alpha_o)\beta_{2L} = 0 \quad \text{a.s.}$$

Given the expressions for  $D_{w^*}(X, \alpha_o)$  in (A.4) and (A.5), condition (1) implies that  $D_{w^{*1}}(X, \alpha_o)$  cannot be expressed as a linear combination of  $D_{w^{*2l}}(X, \alpha_o)$ ,  $l = 1, \dots, L$  almost surely; condition (2) implies that  $D_{w^{*1}}(X, \alpha_o)\beta_1 = 0$  a.s. only when  $\beta_1 \equiv 0$ ; condition (3) implies that  $D_{w^{*21}}(X, \alpha_o)\beta_{21} + \dots + D_{w^{*2L}}(X, \alpha_o)\beta_{2L} = 0$  a.s. only when  $\beta_{2l} \equiv 0, l = 1, \dots, L$ . Hence  $E[D_{w^*}(X, \alpha_o)' \Sigma(X)^{-1} D_{w^*}(X, \alpha_o)]$  is positive-definite. ■

**Lemma 2.** *Let the conditional distribution of  $Y_2$  given  $X$  depends on  $X$  only through  $X_2 - \phi(X'_1 \theta_{o1})$ . Suppose that  $\Sigma(X)$  is positive definite for all  $X$ ; and*

$$E[E[\nabla h_o(Y_2 - \phi(X'_1 \theta_{o1})) | X] \nabla \phi(X'_1 \theta_{o1}) X'_1 \Sigma(X)^{-1} E[\nabla h_o(Y_2 - \phi(X'_1 \theta_{o1})) | X] \nabla \phi(X'_1 \theta_{o1}) X'_1] < \infty,$$

$$E[X_1 \Sigma(X)^{-1} X'_1] < \infty.$$

Then (i)  $E[w^*(Y_2 - \phi(X'_1 \theta_{o1})) | X] = E[(w^{*1}(\cdot), w^{*21}(\cdot), \dots, w^{*2L}(\cdot)) | X]$  given in (A.2) and (A.3) can be solved from:

$$\begin{aligned} & E[\Sigma(X)^{-1} | X_2 - \phi(X'_1 \theta_{o1})] \times E[w^{*1}(Y_2 - \phi(X'_1 \theta_{o1})) | X] & (A.6) \\ = & -E[\Sigma(X)^{-1} E\{\nabla h_o(Y_2 - \phi(X'_1 \theta_{o1})) | X_2 - \phi(X'_1 \theta_{o1})\} \nabla \phi(X'_1 \theta_{o1}) X'_1 | X_2 - \phi(X'_1 \theta_{o1})] \end{aligned}$$

and

$$E[\Sigma(X)^{-1} | X_2 - \phi(X'_1 \theta_{o1})] \times E[w^{*2l}(Y_2 - \phi(X'_1 \theta_{o1})) | X] = E[\Sigma(X)^{-1} e_l X'_1 | X_2 - \phi(X'_1 \theta_{o1})] \quad (A.7)$$

(ii) If Assumption N1' (ii)-(iv) hold, then  $E[D_{w^*}(X, \alpha_o)' \Sigma(X)^{-1} D_{w^*}(X, \alpha_o)]$  is finite and positive definite.

**Proof.** Directly follows from Lemma 1. When the conditional distribution of  $Y_2$  given  $X$  depends on  $X$  only through  $X_2 - \phi(X'_1 \theta_{o1})$ , we have:

$$\overline{\mathcal{S}} = \{s \in \text{sigma-field}(X_2 - \phi(X'_1 \theta_{o1})) : E[s(X_2 - \phi(X'_1 \theta_{o1}))' \Sigma(X)^{-1} s(X_2 - \phi(X'_1 \theta_{o1}))] < \infty\}.$$

Hence  $w^{*1}(Y_2 - \phi(X'_1 \theta_{o1}))$  given in (A.6) and  $w^{*2l}(Y_2 - \phi(X'_1 \theta_{o1}))$  given in (A.7),  $l = 1, \dots, L$ , belong to  $\overline{\mathcal{S}}$ .

Moreover  $D_{w^{*j}}(X, \alpha_o) \in \overline{\mathcal{S}}^\perp$ , since for any  $s(X_2 - \phi(X'_1 \theta_{o1})) \in \overline{\mathcal{S}}$ ,

$$\begin{aligned} & E[D_{w^{*j}}(X, \alpha_o)' \Sigma(X)^{-1} s(X_2 - \phi(X'_1 \theta_{o1}))] \\ = & E[E\{D_{w^{*j}}(X, \alpha_o)' \Sigma(X)^{-1} | X_2 - \phi(X'_1 \theta_{o1})\} s(X_2 - \phi(X'_1 \theta_{o1}))] = 0 \end{aligned}$$

where the last equality is due to

$$\begin{aligned}
& E\{D_{w^{*1}}(X, \alpha_o)' \Sigma(X)^{-1} | X_2 - \phi(X_1' \theta_{o1})\} \\
&= E\{(E[\nabla h_o(Y_2 - \phi(X_1' \theta_{o1})) | X] \nabla \phi(X_1' \theta_{o1}) X_1' \\
&\quad + E[w^{*1}(Y_2 - \phi(X_1' \theta_{o1})) | X])' \Sigma(X)^{-1} | X_2 - \phi(X_1' \theta_{o1})\} \\
&= E\{(E[\nabla h_o(Y_2 - \phi(X_1' \theta_{o1})) | X_2 - \phi(X_1' \theta_{o1})] \nabla \phi(X_1' \theta_{o1}) X_1' \\
&\quad + E[w^{*1}(Y_2 - \phi(X_1' \theta_{o1})) | X_2 - \phi(X_1' \theta_{o1})])' E\{\Sigma(X)^{-1} | X_2 - \phi(X_1' \theta_{o1})\} \\
&= 0
\end{aligned}$$

and for all  $l = 1, \dots, L$ ,

$$\begin{aligned}
& E\{D_{w^{*2l}}(X, \alpha_o)' \Sigma(X)^{-1} | X_2 - \phi(X_1' \theta_{o1})\} \\
&= E\{(-e_l X_1' + E[w^{*2l}(Y_2 - \phi(X_1' \theta_{o1})) | X])' \Sigma(X)^{-1} | X_2 - \phi(X_1' \theta_{o1})\} \\
&= E\{(-e_l X_1')' \Sigma(X)^{-1} | X_2 - \phi(X_1' \theta_{o1})\} \\
&\quad + E[w^{*1}(Y_2 - \phi(X_1' \theta_{o1})) | X_2 - \phi(X_1' \theta_{o1})]' E\{\Sigma(X)^{-1} | X_2 - \phi(X_1' \theta_{o1})\} \\
&= 0
\end{aligned}$$

for  $w^*(Y_2 - \phi(X_1' \theta_{o1}))$  given in (A.6) and (A.7). ■

**Assumption N3.** There is a finite constant  $c$  such that for all  $\theta_1$  within  $o(n^{-1/4})$  shrinking neighborhood of  $\theta_{o1}$  and for  $x_1 = 0, 1$ , (i)  $\|E\{w_l^*(Y_2 - \phi(X_1' \theta_1)) | X_1 = x_1, X_2 = \cdot\}\|_{W_\infty^{rm}} \leq c$  for  $l = 1, \dots, L$ ; (ii)  $\nabla\{\Pi_n w_l^*(Y_2 - \phi(X_1' \theta_1))\}$  is continuous in  $\theta_1$  for  $l = 1, \dots, L$ ; (iii)  $\|E[\nabla \psi_{kj}(Y_2 - \phi(X_1' \theta_1)) | X_1 = x_1, X_2 = \cdot]\|_{W_\infty^{rm}} \leq c$  for  $k = 0, \dots, K_n, j \in \mathcal{K}_n$ .

**Proof. (Proposition 3):** We obtain the limiting distribution of  $\hat{\theta}$  by verifying that Assumptions 4.1 - 4.6 of Theorem 4.1 in Ai and Chen (2003) are satisfied. Their Assumption 4.1(i) is implied by our Lemma 1, their Assumption 4.1(ii)(iii) is our assumption N2. For their Assumptions 4.2 - 4.5, we recall that the directional derivative  $\frac{d\rho(Z, \alpha)}{d\alpha}[v]$  at the direction  $v = (v_\theta, v_h)$  is:

$$\frac{d\rho(Z, \alpha)}{d\alpha}[v] = \left( \frac{d\rho_1(Z, \alpha)}{d\alpha}[v], \dots, \frac{d\rho_L(Z, \alpha)}{d\alpha}[v] \right)' = \frac{d\rho(Z, \alpha)}{d\theta'}(v_\theta) + \frac{d\rho(Z, \alpha)}{dh}[v_h].$$

Notice that in this paper, for all  $l = 1, \dots, L$ ,

$$\frac{d\rho_l(Z, \alpha)}{d\alpha}[v] = \nabla h_l(Y_2 - \phi(X_1' \theta_1)) \times \nabla \phi(X_1' \theta_1) \times X_1'[v_{\theta_1}] - X_1'[v_{\theta_{2l}}] - [v_{h_l}(Y_2 - \phi(X_1' \theta_1))].$$

Denote  $\frac{dm(X, \alpha)}{d\alpha}[v] \equiv E\left[\frac{d\rho(Z, \alpha)}{d\alpha}[v] | X\right]$  then in this paper for all  $l = 1, \dots, L$ ,

$$\begin{aligned}
& \frac{dm_l(X, \alpha)}{d\alpha}[v] \\
&= E[\nabla h_l(Y_2 - \phi(X_1' \theta_1)) | X] \nabla \phi(X_1' \theta_1) X_1'[v_{\theta_1}] - X_1'[v_{\theta_{2l}}] - E[v_{h_l}(Y_2 - \phi(X_1' \theta_1)) | X].
\end{aligned}$$

Then

$$\begin{aligned}\|v\|^2 &= E \left[ \left\{ \frac{dm(X, \alpha_o)}{d\alpha} [v] \right\}' \Sigma(X)^{-1} \left\{ \frac{dm(X, \alpha_o)}{d\alpha} [v] \right\} \right] \\ &= v_\theta' E [D_w(X, \alpha_o)' \Sigma(X)^{-1} D_w(X, \alpha_o)] v_\theta \quad \text{when } v_\theta \neq 0,\end{aligned}$$

where  $v_h(\cdot) = -w(\cdot)v_\theta$  with  $w(\cdot) \equiv (w^1(\cdot), w^{21}(\cdot), \dots, w^{2L}(\cdot))$ . Also for any fixed  $\lambda \in \mathcal{R}^b$  with  $\lambda \neq 0$ , we have  $v^* \equiv (v_\theta^*, v_h^*)$  with

$$v_\theta^* = (E\{D_{w^*}(X, \alpha_o)' \Sigma(X)^{-1} D_{w^*}(X, \alpha_o)\})^{-1} \lambda, \quad v_h^* = -w^* \times v_\theta^*.$$

Hence given our Lemma 1 and Assumptions 5(iii)(iv), 7(ii) and N3(i),

$$\begin{aligned}& \|\Pi_n v^* - v^*\|^2 \\ &= v_\theta^{*'} E[E\{\Pi_n w^*(Y_2 - \phi(X_1' \theta_{o1})) - w^*(Y_2 - \phi(X_1' \theta_{o1})) | X\}' \Sigma(X)^{-1} E\{\Pi_n w^*(\cdot) - w^*(\cdot) | X\}] v_\theta^* \\ &= o(n^{-1/2}),\end{aligned}$$

and Ai-Chen's assumption 4.2 is satisfied.

Since for all  $l = 1, \dots, L$ ,

$$\frac{d\rho_l(Z, \alpha)}{d\alpha} [\Pi_n v^*] = \nabla h_l(Y_2 - \phi(X_1' \theta_1)) \nabla \phi(X_1' \theta_1) X_1' v_{\theta_1}^* - X_1' v_{\theta_{2l}}^* + \Pi_n w_l^*(Y_2 - \phi(X_1' \theta_1)) v_\theta^*$$

$$\begin{aligned}& \frac{dm_l(X, \alpha)}{d\alpha} [\Pi_n v^*] \\ &= E\{\nabla h_l(Y_2 - \phi(X_1' \theta_1)) | X\} \nabla \phi(X_1' \theta_1) X_1' v_{\theta_1}^* - X_1' v_{\theta_{2l}}^* + E\{\Pi_n w_l^*(Y_2 - \phi(X_1' \theta_1)) | X\} v_\theta^*\end{aligned}$$

and for all  $\alpha = \tau \alpha_o + (1 - \tau) \bar{\alpha}$  with  $\tau \in [0, 1]$ ,  $\bar{\alpha} \in \mathcal{A}_n$  within a shrinking  $o(n^{-1/4})$ -neighborhood of  $\alpha_o$ ,

$$E\{\nabla h_l(Y_2 - \phi(X_1' \theta_1)) | X\} = \tau E\{\nabla h_{ol}(Y_2 - \phi(X_1' \theta_1)) | X\} + (1 - \tau) E\{\nabla \bar{h}_l(Y_2 - \phi(X_1' \theta_1)) | X\}$$

$\bar{h}_l \in \mathcal{H}_{l,n}$  given in (3.10) implies

$$E\{\nabla \bar{h}_l(Y_2 - \phi(X_1' \theta_1)) | X\} = \sum_{k=0}^{K_n} \sum_{j \in \mathcal{K}_n} \pi_{kj}^l E\{\nabla \psi_{kj}(Y_2 - \phi(X_1' \theta_1)) | X\}$$

Hence Ai-Chen's assumption 4.3 is satisfied given our Assumptions 2 and N3.

For Ai-Chen's assumption 4.4, we note that for  $\alpha \in \mathcal{A}_n$  within a shrinking  $o(n^{-1/4})$ -neighborhood of  $\alpha_o$ ,

$$\begin{aligned}
& \frac{dm_l(X, \alpha)}{d\alpha} [\Pi_n v^*] - \frac{dm_l(X, \alpha_o)}{d\alpha} [\Pi_n v^*] \\
&= (E\{\nabla h_l(Y_2 - \phi(X'_1 \theta_1)) | X\} \nabla \phi(X'_1 \theta_1) - E\{\nabla h_{ol}(Y_2 - \phi(X'_1 \theta_{o1})) | X\} \nabla \phi(X'_1 \theta_{o1})) X'_1 v_{\theta_1}^* \\
&\quad + (E\{\Pi_n w_l^*(Y_2 - \phi(X'_1 \theta_1)) | X\} - E\{\Pi_n w_l^*(Y_2 - \phi(X'_1 \theta_{o1})) | X\}) v_{\theta}^* \\
&= (E\{\nabla h_l(Y_2 - \phi(X'_1 \theta_1)) | X\} - E\{\nabla h_{ol}(Y_2 - \phi(X'_1 \theta_1)) | X\}) \nabla \phi(X'_1 \theta_1) X'_1 v_{\theta_1}^* \\
&\quad + (E\{\nabla h_{ol}(Y_2 - \phi(X'_1 \theta_1)) | X\} - E\{\nabla h_{ol}(Y_2 - \phi(X'_1 \theta_{o1})) | X\}) \nabla \phi(X'_1 \theta_1) X'_1 v_{\theta_1}^* \\
&\quad + E\{\nabla h_{ol}(Y_2 - \phi(X'_1 \theta_{o1})) | X\} (\nabla \phi(X'_1 \theta_1) - \nabla \phi(X'_1 \theta_{o1})) X'_1 v_{\theta_1}^* \\
&\quad + (E\{\Pi_n w_l^*(Y_2 - \phi(X'_1 \theta_1)) | X\} - E\{\Pi_n w_l^*(Y_2 - \phi(X'_1 \theta_{o1})) | X\}) v_{\theta}^*
\end{aligned}$$

Under our assumption 2, we have

$$\nabla h_{ol}(Y_2 - \phi(X'_1 \theta_1)) = \sum_{k=0}^{\infty} \sum_{j \in \mathcal{K}_n} \pi_{kj}^{ol} \nabla \psi_{kj}(Y_2 - \phi(X'_1 \theta_1))$$

$h_l \in \mathcal{H}_{l,n}$  given in (3.10) implies

$$\nabla h_l(Y_2 - \phi(X'_1 \theta_1)) = \sum_{k=0}^{K_n} \sum_{j \in \mathcal{K}_n} \pi_{kj}^l \nabla \psi_{kj}(Y_2 - \phi(X'_1 \theta_1))$$

hence

$$\begin{aligned}
& \nabla h_l(Y_2 - \phi(X'_1 \theta_1)) - \nabla h_{ol}(Y_2 - \phi(X'_1 \theta_1)) \\
&= \sum_{k=0}^{K_n} \sum_{j \in \mathcal{K}_n} \{\pi_{kj}^l - \pi_{kj}^{ol}\} \nabla \psi_{kj}(Y_2 - \phi(X'_1 \theta_1)) + \{\Pi_n \nabla h_{ol}(Y_2 - \phi(X'_1 \theta_1)) - \nabla h_{ol}(Y_2 - \phi(X'_1 \theta_1))\}
\end{aligned}$$

Thus Ai-Chen's assumption 4.4 is satisfied under our assumptions 2, 5, 7(ii) and N3.

For Ai-Chen's assumption 4.5, we note that for  $\bar{\alpha} \in \mathcal{A}_n$  within a shrinking  $o(n^{-1/4})$ -neighborhood of  $\alpha_o$ , and for  $\alpha = \tau \alpha_o + (1 - \tau) \bar{\alpha}$  for all  $\tau \in [0, 1]$ ,

$$\begin{aligned}
& \frac{dm_l(X, \alpha)}{d\alpha} [\bar{\alpha} - \alpha_o] - \frac{dm_l(X, \alpha_o)}{d\alpha} [\bar{\alpha} - \alpha_o] \\
&= (E\{\nabla h_l(Y_2 - \phi(X'_1 \theta_1)) | X\} - E\{\nabla h_{ol}(Y_2 - \phi(X'_1 \theta_1)) | X\}) \nabla \phi(X'_1 \theta_1) X'_1 [\bar{\theta}_1 - \theta_{o1}] \\
&\quad + (E\{\nabla h_{ol}(Y_2 - \phi(X'_1 \theta_1)) | X\} - E\{\nabla h_{ol}(Y_2 - \phi(X'_1 \theta_{o1})) | X\}) \nabla \phi(X'_1 \theta_1) X'_1 [\bar{\theta}_1 - \theta_{o1}] \\
&\quad + E\{\nabla h_{ol}(Y_2 - \phi(X'_1 \theta_{o1})) | X\} (\nabla \phi(X'_1 \theta_1) - \nabla \phi(X'_1 \theta_{o1})) X'_1 [\bar{\theta}_1 - \theta_{o1}] \\
&\quad - (E\{\bar{h}_l(Y_2 - \phi(X'_1 \theta_1)) - h_{ol}(Y_2 - \phi(X'_1 \theta_1)) | X\}) \\
&\quad - (E\{\bar{h}_l(Y_2 - \phi(X'_1 \theta_{o1})) - h_{ol}(Y_2 - \phi(X'_1 \theta_{o1})) | X\})
\end{aligned}$$

Thus uniformly over  $\bar{\alpha} \in \mathcal{A}_n$  within a shrinking  $o(n^{-1/4})$ -neighborhood of  $\alpha_o$ , and for  $\alpha = \tau\alpha_o + (1 - \tau)\bar{\alpha}$  for all  $\tau \in [0, 1]$ ,

$$\begin{aligned} & E \left( \left\{ \frac{dm(X, \alpha)}{d\alpha} [\bar{\alpha} - \alpha_o] - \frac{dm(X, \alpha_o)}{d\alpha} [\bar{\alpha} - \alpha_o] \right\}' \left\{ \frac{dm(X, \alpha)}{d\alpha} [\bar{\alpha} - \alpha_o] - \frac{dm(X, \alpha_o)}{d\alpha} [\bar{\alpha} - \alpha_o] \right\} \right) \\ &= o(n^{-1/2}) \end{aligned}$$

this and our assumptions N1 and 7(ii) imply that Ai-Chen's assumption 4.5 is satisfied by applying Cauchy-Schwartz inequality.

For Ai-Chen's assumption 4.6, we note that for all  $\alpha \in \mathcal{A}_n$  within a shrinking  $o(n^{-1/4})$ -neighborhood of  $\alpha_o$ , and for all  $\tau \in [0, 1]$ , and for any  $l = 1, \dots, L$ ,

$$\begin{aligned} & \rho_l(Z, \alpha + \tau\Pi_n v^*) \\ &= Y_{1l} - [h_l(Y_2 - \phi(X'_1\{\theta_1 + \tau v_{\theta_1}^*\})) + \tau\Pi_n w_l^*(Y_2 - \phi(X'_1\{\theta_1 + \tau v_{\theta_1}^*\}))v_{\theta}^*] - X'_1\{\theta_{2l} + \tau v_{\theta_{2l}}^*\} \end{aligned}$$

hence

$$\begin{aligned} & \frac{d^2 \rho_l(Z, \alpha + \tau\Pi_n v^*)}{d\tau^2} \Big|_{\tau=0} \\ &= -\nabla^2 h_l(Y_2 - \phi(X'_1\theta_1))[\nabla\phi(X'_1\theta_1)X'_1 v_{\theta_1}^*]^2 + \nabla h_l(Y_2 - \phi(X'_1\theta_1))\nabla^2\phi(X'_1\theta_1)[X'_1 v_{\theta_1}^*]^2 \\ & \quad - \nabla\Pi_n w_l^*(Y_2 - \phi(X'_1\theta_1))v_{\theta}^*[\nabla\phi(X'_1\theta_1)X'_1 v_{\theta_1}^*] \end{aligned}$$

and Ai-Chen's assumption 4.6 is automatically satisfied under our assumptions 2 and N3(ii) and  $h_l \in \mathcal{H}_{l,n}$  given in (3.10). ■

**Proof. (Proposition 4):** Applying Theorems 6.1 and 6.2 of Ai and Chen (1999, 2003), the verification is very similar to that of Proposition 3, hence omitted. ■

**Proof. (Proposition 5):** Directly follows from Theorem 5.1 of Ai and Chen (2003). ■

## References

- [1] Ai, C. and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," forthcoming in *Econometrica*.
- [2] Anastassiou, G. and X.M. Yu (1992): "Monotone and Probabilistic Wavelet Approximation," *Stochastic Analysis & Applications*, 10, 251-264.
- [3] Banks, J., R.W. Blundell and A. Lewbel (1997): "Quadratic Engel Curves, Indirect Tax Reform and Welfare Measurement," *Review of Economics and Statistics*, LXXIX, 4, 527-539.
- [4] Bierens, H. and H. Pott-Butter (1990): "Specification of Household Engel Curves by Nonparametric Regression," *Econometric Reviews*, 9, 123-184.

- [5] Birman and Solomjak (1977): “Estimates of Singular Numbers of Integral Operators,” *Russian Math. Surveys*, 32, 15-89.
- [6] Blackorby, C. and D. Donaldson (1994): “Measuring the Cost of Children: a Theoretical Framework,” in *The Measurement of Household Welfare*, ed. by R. Blundell, I. Preston, and I. Walker, chp. 2, pp. 51-69. Cambridge University Press.
- [7] Blundell, R. (1998): “Consumer Behaviour: Theory and Empirical Evidence,” *Economic Journal*, March, 16-65.
- [8] Blundell, R., M. Browning and I. Crawford (2003): “Nonparametric Engel Curves and Revealed Preference,” *Econometrica*, 71, 205-240.
- [9] Blundell, R. and A. Duncan (1998): “Kernel Regression in Empirical Microeconomics,” *Journal of Human Resources*, 33, 62-87.
- [10] Blundell, R., A. Duncan and K. Pendakur (1998): “Semiparametric Estimation and Consumer Demand,” *Journal of Applied Econometrics*, 13, 435-461.
- [11] Blundell, R.W. and J.L. Powell (2003): “Endogeneity in Semiparametric and Nonparametric Regression Models,” In: *Advances in Economics and Econometrics: Theory and Applications*, Dewatripont, M., Hansen, L.P., and Turnovsky, S.J., eds, vol. 2, pp. 312-357. Cambridge, UK: Cambridge University Press.
- [12] Chen, X. and X. Shen (1998): “Sieve Extremum Estimates for Weakly Dependent Data,” *Econometrica*, 66, 289-314.
- [13] Chen, X., L.P. Hansen, and J. Scheinkman (1997): “Shape-preserving Estimation of Diffusions,” manuscript, University of Chicago, Dept. of Economics.
- [14] Chen, X. and T. Conley (2001): “A New Semiparametric Spatial Model for Panel Time Series,” *Journal of Econometrics*, 105, 59-83.
- [15] Chui (1992): *An Introduction to Wavelets*. Academic Press.
- [16] Darolles, S., J.-P. Florens and E. Renault (2002): “Nonparametric Instrumental Regression,” mimeo, GREMAQ, University of Toulouse.
- [17] Deaton A. S. and J. Muellbauer (1980): “An Almost Ideal Demand System,” *American Economic Review*, 70, 312-36.
- [18] Eubank, L. (1988): *Nonparametric Regression and Spline Smoothing*. Marcel Dekker.
- [19] Florens, J. (2003): “Inverse Problems and Structural Econometrics: the Example of Instrumental Variables,” In: *Advances in Economics and Econometrics: Theory and Applications*, Dewatripont, M., Hansen, L.P., and Turnovsky, S.J., eds, vol. 2, pp. 284-311. Cambridge, UK: Cambridge University Press.
- [20] Florens, J., J. Heckman, C. Meghir and E. Vytlačil (2002): “Instrumental Variables, Local Instrumental Variables and Control Functions,” IFS Working paper.

- [21] Hall, P. (1992): *The Bootstrap and Edgeworth Expansion*. Springer-Verlag New York.
- [22] Hall, P. and J. Horowitz (2003): “Nonparametric Methods for Inference in the Presence of Instrumental Variables,” Northwestern working paper.
- [23] Härdle W. and M. Jerison (1988): “Cross Section Engel Curves over time,” Discussion paper no. A-160. SFB 303, University of Bonn.
- [24] Härdle W. and O. Linton (1994): “Applied Nonparametric Models,” in Engle and McFadden, *Handbook of Econometrics*, Vol IV, North-Holland.
- [25] Härdle, W. and J. Marron (1990): “Semiparametric Comparison of Regression Curves,” *Annals of Statistics*, 18, 63-89.
- [26] Hausman, J., W.K. Newey, H. Ichimura and J. Powell (1991): “Identification and Estimation of Polynomial Errors in Variables Models,” *Journal of Econometrics*, 50, 273-296.
- [27] Heinrich, S. and T. Kuhn (1985): “Embedding maps between Holder spaces over metric compacta and eigenvalues of integral operators,” *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, Series A*, vol. 88, 47-62.
- [28] Hoeffding, W. (1977): “Some Incomplete and Boundedly Complete Families of Distributions,” *Annals of Statistics*, 5, 278-291.
- [29] Huang, J. (1998): “Projection Estimation in Multiple Regression with Application to Functional ANOVA Models,” *Annals of Statistics*, 26, 242-272.
- [30] Jorgenson, D.W., L.J. Lau, and T.M. Stoker (1982): “The Transcendental Logarithmic Model of Aggregate Consumer Behavior,” in *Advances in Econometrics*, Vol. 1, ed. by R. Basmann and G. Rhodes, Greenwich, Connecticut: JAI Press.
- [31] Korostelev, A. and A. Tsybakov (1993): *Minimax Theory of Image Reconstruction*, Lecture Notes in Statistics, v.82, Springer: New York.
- [32] Kress, R. (1999): *Linear Integral Equations*, 2nd ed. Springer: New York.
- [33] Lehmann, E.L. (1986): *Testing Statistical Hypothesis*, 2nd ed. Wiley: New York.
- [34] Leser, C. E. V. (1963): “Forms of Engel Functions,” *Econometrica*, 31, 694-703.
- [35] Lewbel, A. (1991): “The rank of demand systems: theory and nonparametric estimation,” *Econometrica*, 59, 711-730.
- [36] Mattner, L. (1993): “Some Incomplete but Bounded Complete Location Families,” *Annals of Statistics*, 21, 2158-2162.
- [37] Matzkin, R.L. (1994): “Restrictions of Economic Theory in Nonparametric Methods,” in *Handbook of Econometrics: Vol IV*, edited by R.F. Engle and D.L. McFadden, Elsevier Science, North Holland.

- [38] Meyer, Y. (1992): *Wavelets and Operators*. Cambridge University Press.
- [39] Newey, W.K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147-168.
- [40] Newey, W.K. and J. Powell (2003): "Instrumental Variables Estimation for Nonparametric Models," *Econometrica*, 71, 1565-1578.
- [41] Newey, W. Powell, J. and F. Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67, 565-604.
- [42] Pendakur, K. (1998): "Semiparametric Estimates and Tests of Base-independent Equivalence Scales," *Journal of Econometrics*, 88, 1-40.
- [43] Pinkse, C. and P. Robinson (1995): "Pooling Nonparametric Estimates of Regression Functions with a Similar Shape," *Advances in Econometrics and Quantitative Economics*, eds. G. Maddala, P. Phillips and T.N. Srinivisan, 172-195.
- [44] Politis, D.N., J.P. Romano and M. Wolf (1999): *Subsampling*. Springer-Verlag New York.
- [45] Robin, J.-M. and R. Smith (2000): "Tests of Rank," *Econometric Theory*, 16, 151-175.
- [46] Robinson, P. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931-954.
- [47] Stock, J.H., J.H. Wright and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20, 518-529.
- [48] Stock, J.H. and M. Yogo (2002): "Testing for Weak Instruments in Linear IV Regression," NBER Working Paper No. t0284
- [49] Working, H. (1943): "Statistical Laws of Family Expenditure," *Journal of the American Statistical Association*, 38, 43-56.

## B. Tables

Table B.1: MC-study: MSE of sieve IV-estimator of nonlinear  $h$

P-spline, $k_n = 4$										
	$m_n = \text{cos-sin}, J_n = 13$					$m_n = \text{cos-sin}, J_n = 27$				
$\lambda$	0.80	0.10	0.01	0.001	0.00	0.80	0.10	0.01	0.001	0.00
Bias <sup>2</sup> ( $10^{-2} \times$ )	2.28	2.48	2.27	2.06	2.02	2.24	2.28	1.94	1.75	1.71
Var. ( $10^{-2} \times$ )	0.35	0.71	1.12	1.57	1.72	0.33	0.66	1.06	1.45	1.55
MSE ( $10^{-2} \times$ )	2.63	3.19	3.39	3.63	3.74	2.57	2.94	3.00	3.20	3.26
P-spline, $k_n = 5$										
	$m_n = \text{cos-sin}, J_n = 17$					$m_n = \text{cos-sin}, J_n = 27$				
$\lambda$	0.80	0.10	0.01	0.001	0.00	0.80	0.10	0.01	0.001	0.00
Bias <sup>2</sup> ( $10^{-2} \times$ )	2.15	2.29	1.90	1.62	1.56	2.12	2.16	1.71	1.48	1.46
Var. ( $10^{-2} \times$ )	0.41	0.78	1.21	2.38	3.63	0.40	0.74	1.21	2.34	3.14
MSE ( $10^{-2} \times$ )	2.56	3.07	3.11	4.00	5.19	2.52	2.90	2.92	3.82	4.60
P-spline, $k_n = 6$										
	$m_n = \text{cos-sin}, J_n = 19$					$m_n = \text{cos-sin}, J_n = 27$				
$\lambda$	0.80	0.10	0.01	0.001	0.00	0.80	0.10	0.01	0.001	0.00
Bias <sup>2</sup> ( $10^{-2} \times$ )	1.87	1.84	1.58	1.82	3.55	1.83	1.68	1.38	1.63	2.77
Var. ( $10^{-2} \times$ )	0.41	0.89	1.66	3.31	10.35	0.41	0.87	1.66	3.23	7.82
MSE ( $10^{-2} \times$ )	2.28	2.74	3.24	5.13	13.90	2.24	2.55	3.04	4.86	10.59
P-spline, $k_n = 7$										
	$m_n = \text{cos-sin}, J_n = 23$					$m_n = \text{cos-sin}, J_n = 27$				
$\lambda$	0.80	0.10	0.01	0.001	0.00	0.80	0.10	0.01	0.001	0.00
Bias <sup>2</sup> ( $10^{-2} \times$ )	1.78	1.64	1.40	1.99	4.06	1.76	1.57	1.32	1.91	3.90
Var. ( $10^{-2} \times$ )	0.41	0.90	1.94	4.18	20.21	0.41	0.89	1.93	4.13	15.79
MSE ( $10^{-2} \times$ )	2.19	2.54	3.34	6.17	24.27	2.17	2.46	3.25	6.04	19.69
P-spline, $k_n = 8$										
	$m_n = \text{cos-sin}, J_n = 25$					$m_n = \text{B-spline}, J_n = 25$				
$\lambda$	0.80	0.10	0.01	0.001	0.00	0.80	0.10	0.01	0.001	0.00
Bias <sup>2</sup> ( $10^{-2} \times$ )	1.75	1.58	1.37	2.01	3.71	1.66	1.17	0.83	1.00	1.19
1Var. ( $10^{-2} \times$ )	0.41	0.91	2.07	4.85	56.69	0.37	0.77	1.60	3.13	7.19
MSE ( $10^{-2} \times$ )	2.16	2.49	3.44	6.86	60.40	2.03	1.94	2.43	4.13	8.38
P-spline, $k_n = 9$										
	$m_n = \text{cos-sin}, J_n = 29$					$m_n = \text{B-spline}, J_n = 25$				
$\lambda$	0.80	0.10	0.01	0.001	0.00	0.80	0.10	0.01	0.001	0.00
Bias <sup>2</sup> ( $10^{-2} \times$ )	1.72	1.50	1.30	1.91	3.51	1.65	1.16	0.84	1.00	1.22
Var. ( $10^{-2} \times$ )	0.40	0.90	2.18	5.77	56.36	0.37	0.78	1.70	3.47	10.36
MSE ( $10^{-2} \times$ )	2.12	2.40	3.48	7.68	59.87	2.02	1.94	2.54	4.48	11.58

Table B.2: MC-study: MSE of sieve IV-estimator of nonlinear  $h$

P-spline, $k_n = 4$	$m_n = \text{B-spline}, J_n = 15$					$m_n = \text{B-spline}, J_n = 25$				
$\lambda$	0.80	0.10	0.01	0.001	0.00	0.80	0.10	0.01	0.001	0.00
Bias <sup>2</sup> ( $10^{-2} \times$ )	2.24	2.31	1.20	1.78	1.75	2.20	2.01	1.52	1.32	1.29
Var. ( $10^{-2} \times$ )	0.33	0.64	1.02	1.38	1.46	0.30	0.56	0.86	1.06	1.10
MSE ( $10^{-2} \times$ )	2.56	2.95	3.00	3.16	3.21	2.50	2.57	2.38	2.38	2.39
P-spline, $k_n = 5$	$m_n = \text{B-spline}, J_n = 15$					$m_n = \text{B-spline}, J_n = 25$				
$\lambda$	0.80	0.10	0.01	0.001	0.00	0.80	0.10	0.01	0.001	0.00
Bias <sup>2</sup> ( $10^{-2} \times$ )	2.13	2.20	1.75	1.48	1.45	2.08	1.88	1.32	1.13	1.11
Var. ( $10^{-2} \times$ )	0.40	0.73	1.18	2.23	2.93	0.36	0.63	1.01	1.65	1.90
MSE ( $10^{-2} \times$ )	2.52	2.93	2.93	3.71	4.39	2.44	2.51	2.33	2.78	3.01
P-spline, $k_n = 6$	$m_n = \text{B-spline}, J_n = 15$					$m_n = \text{B-spline}, J_n = 25$				
$\lambda$	0.80	0.10	0.01	0.001	0.00	0.80	0.10	0.01	0.001	0.00
Bias <sup>2</sup> ( $10^{-2} \times$ )	1.84	1.72	1.43	1.66	2.86	1.76	1.31	0.95	0.98	1.04
Var. ( $10^{-2} \times$ )	0.40	0.85	1.59	3.11	7.52	0.36	0.73	1.27	2.18	3.04
MSE ( $10^{-2} \times$ )	2.23	2.57	3.01	4.77	10.38	2.12	2.04	2.22	3.16	4.08
P-spline, $k_n = 7$	$m_n = \text{B-spline}, J_n = 15$					$m_n = \text{B-spline}, J_n = 25$				
$\lambda$	0.80	0.10	0.01	0.001	0.00	0.80	0.10	0.01	0.001	0.00
Bias <sup>2</sup> ( $10^{-2} \times$ )	1.77	1.62	1.37	1.93	3.50	1.68	1.19	0.84	1.00	1.19
Var. ( $10^{-2} \times$ )	0.39	0.86	1.88	4.12	15.68	0.37	0.75	1.49	2.76	5.07
MSE ( $10^{-2} \times$ )	2.16	2.48	3.25	6.05	19.18	2.05	1.94	2.33	3.76	6.26
B-spline, $k_n = 9$	$m_n = \text{B-spline}, J_n = 15$					$m_n = \text{B-spline}, J_n = 25$				
$\lambda$	0.80	0.10	0.01	0.001	0.00	0.80	0.10	0.01	0.001	0.00
Bias <sup>2</sup> ( $10^{-2} \times$ )	0.47	1.10	1.51	1.47	0.70	0.50	0.78	0.89	0.93	0.86
Var. ( $10^{-2} \times$ )	0.61	1.88	5.58	11.83	39.19	0.58	0.19	3.16	4.18	5.39
MSE ( $10^{-2} \times$ )	1.09	2.98	7.09	13.30	39.89	1.08	2.37	4.05	5.11	6.25
B-spline, $k_n = 14$	$m_n = \text{B-spline}, J_n = 15$					$m_n = \text{B-spline}, J_n = 25$				
$\lambda$	0.80	0.10	0.01	0.001	0.00	0.80	0.10	0.01	0.001	0.00
Bias <sup>2</sup> ( $10^{-2} \times$ )	1.75	1.62	1.50	2.13	6.45	0.54	0.85	1.02	1.16	2.14
Var. ( $10^{-2} \times$ )	0.40	0.90	2.29	6.79	265.54	0.63	1.73	3.85	5.99	27.16
MSE ( $10^{-2} \times$ )	2.15	2.52	3.79	8.92	271.99	1.17	2.58	4.87	7.15	29.30

Table B.3: MC-study: MSE of sieve estimators of linear  $h$ .

B-spline, $k_n = 9$	IV-estimator, $J_n = 25$					LS-estimator				
$\lambda$	0.80	0.10	0.01	0.001	0.00	0.80	0.10	0.01	0.001	0.00
Bias <sup>2</sup> ( $10^{-2} \times$ )	0.07	0.03	0.03	0.03	0.04	0.48	0.47	0.47	0.47	0.47
Var. ( $10^{-2} \times$ )	0.07	0.20	0.50	0.81	1.18	0.02	0.03	0.03	0.03	0.03
MSE ( $10^{-2} \times$ )	0.14	0.23	0.53	0.84	1.22	0.50	0.50	0.51	0.50	0.50
P-spline, $k_n = 6$	IV-estimator, $J_n = 25$					LS-estimator				
$\lambda$	0.80	0.10	0.01	0.001	0.00	0.80	0.10	0.01	0.001	0.00
Bias <sup>2</sup> ( $10^{-2} \times$ )	0.03	0.03	0.03	0.03	0.04	0.50	0.49	0.48	0.48	0.48
Var. ( $10^{-2} \times$ )	0.04	0.08	0.14	0.29	0.47	0.01	0.02	0.02	0.02	0.02
MSE ( $10^{-2} \times$ )	0.07	0.10	0.17	0.33	0.51	0.51	0.51	0.50	0.50	0.50

Table B.4: MC-study: MSE of sieve estimators of nonlinear  $h$ .

B-spline, $k_n = 9$	IV-estimator, $J_n = 25$					LS-estimator				
$\lambda$	0.80	0.10	0.01	0.001	0.00	0.80	0.10	0.01	0.001	0.00
Bias <sup>2</sup> ( $10^{-2} \times$ )	0.50	0.78	0.90	0.94	0.86	18.77	18.42	18.48	18.39	18.37
Var. ( $10^{-2} \times$ )	0.58	1.59	3.16	4.18	5.39	0.09	0.11	0.11	0.12	0.12
MSE ( $10^{-2} \times$ )	1.08	2.37	4.06	5.12	6.25	18.86	18.54	18.50	18.51	18.49
P-spline, $k_n = 6$	IV-estimator, $J_n = 25$					LS-estimator				
$\lambda$	0.80	0.10	0.01	0.001	0.00	0.80	0.10	0.01	0.001	0.00
Bias <sup>2</sup> ( $10^{-2} \times$ )	1.76	1.31	0.95	0.98	1.04	19.50	18.82	18.53	18.49	18.48
Var. ( $10^{-2} \times$ )	0.36	0.73	1.27	2.19	3.04	0.06	0.07	0.09	0.09	0.09
MSE ( $10^{-2} \times$ )	2.12	2.04	2.22	3.17	4.08	19.56	18.89	18.62	18.58	18.57

## C. Figures

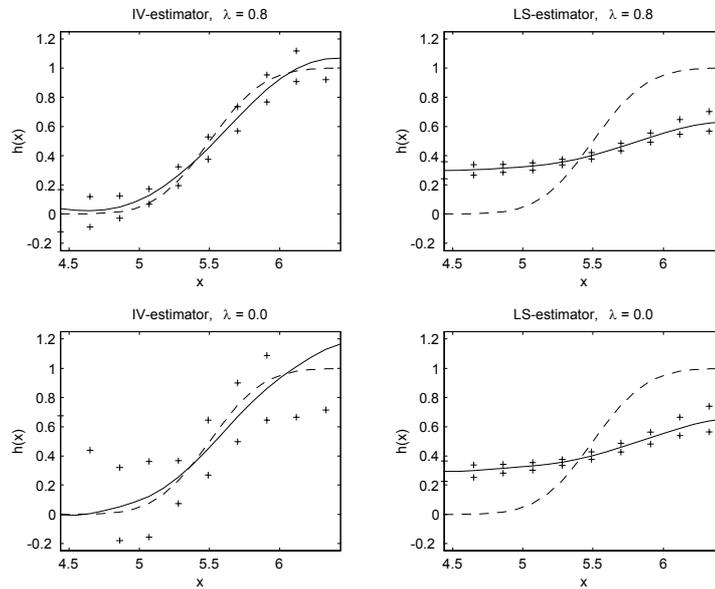


Figure C.1: MC-study: LS- and IV-estimator of non-linear  $h$ ,  $\lambda = 0.8, 0.0$

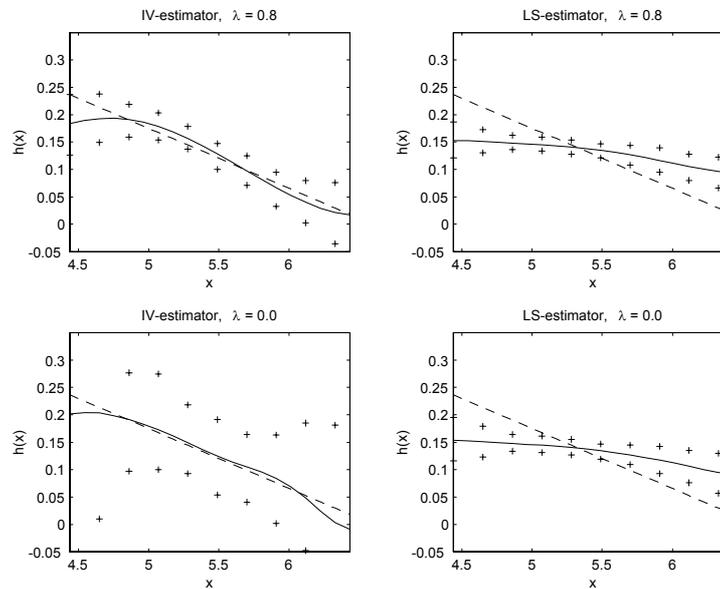


Figure C.2: MC-study: LS- and IV-estimator of linear  $h$ ,  $\lambda = 0.8, 0.0$

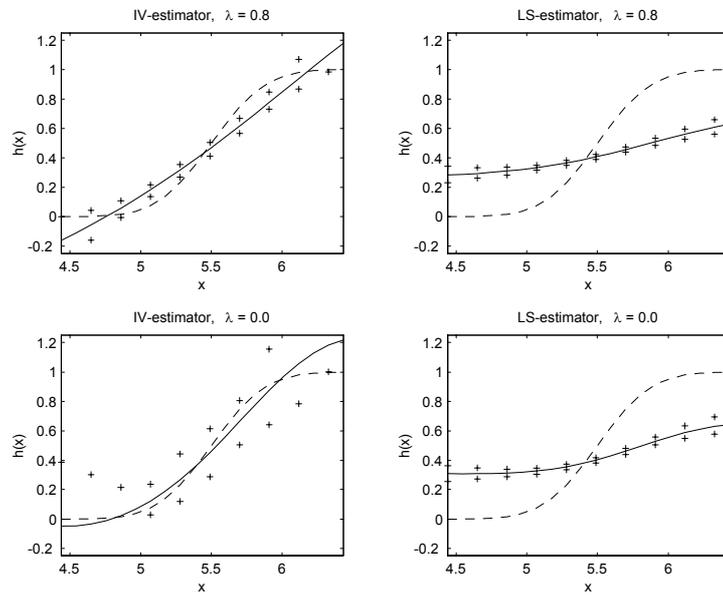


Figure C.3: MC-study: LS- and IV-estimator of non-linear  $h$ ,  $\lambda = 0.8, 0.0$ ;  $k_n = 5$

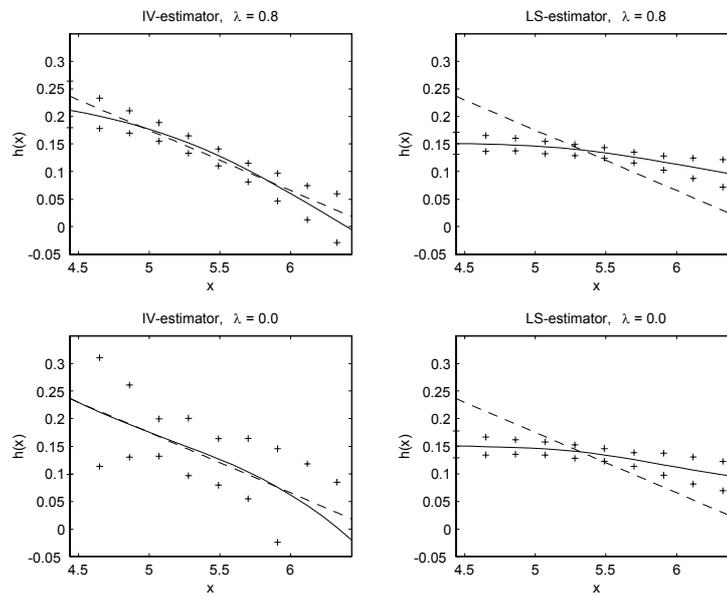


Figure C.4: MC-study: LS- and IV-estimator of linear  $h$ ,  $\lambda = 0.8, 0.0$ ;  $k = 5$

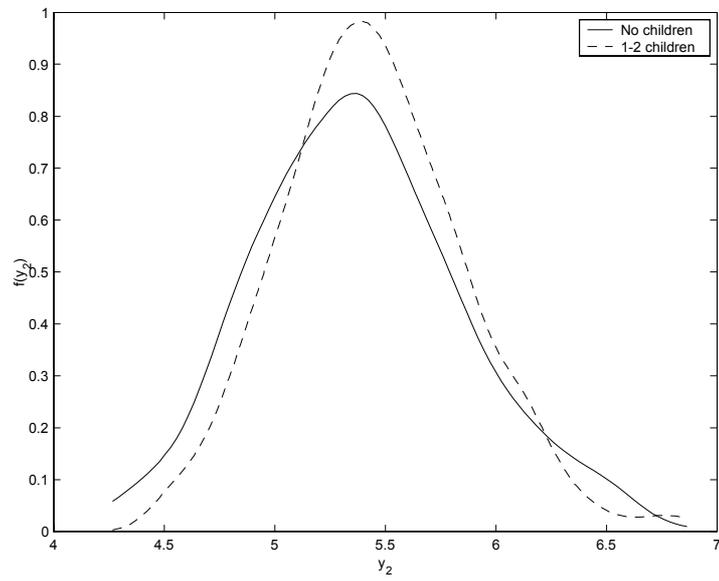


Figure C.5: Kernel estimate of density for  $Y_2$ .

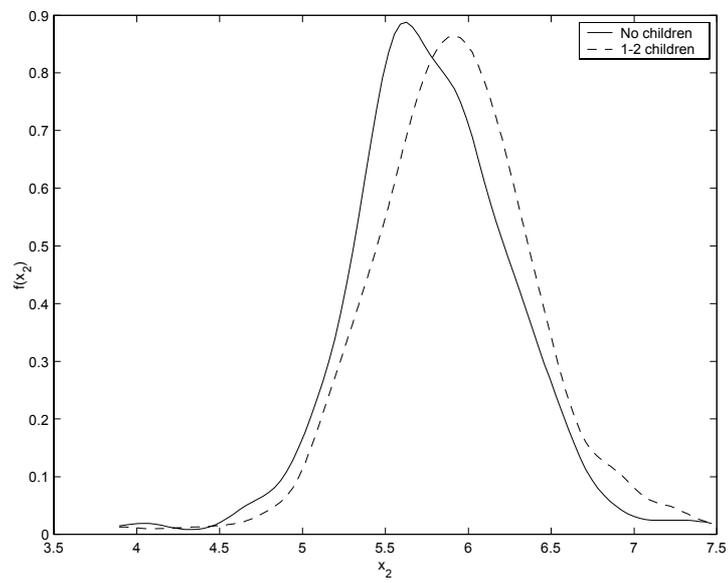


Figure C.6: Kernel estimate of density for  $X_2$ .

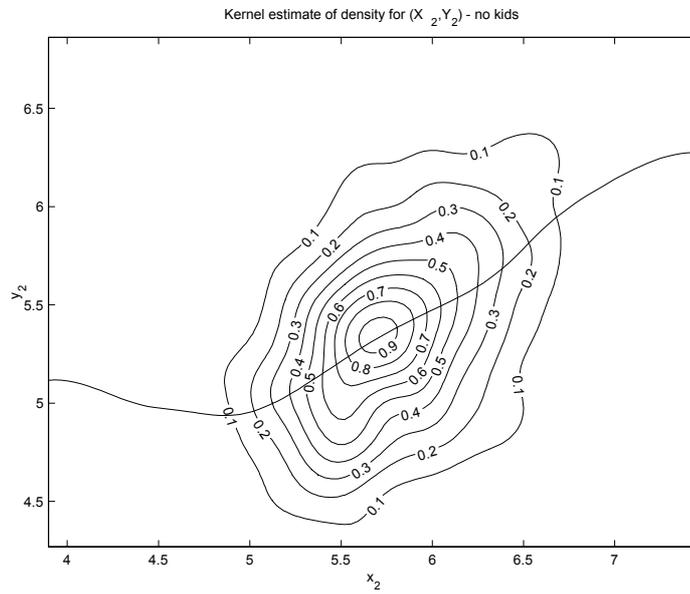


Figure C.7: Kernel estimate of density for  $(X_2, Y_2)$ ; series estimate of  $E[Y_2|X_2]$

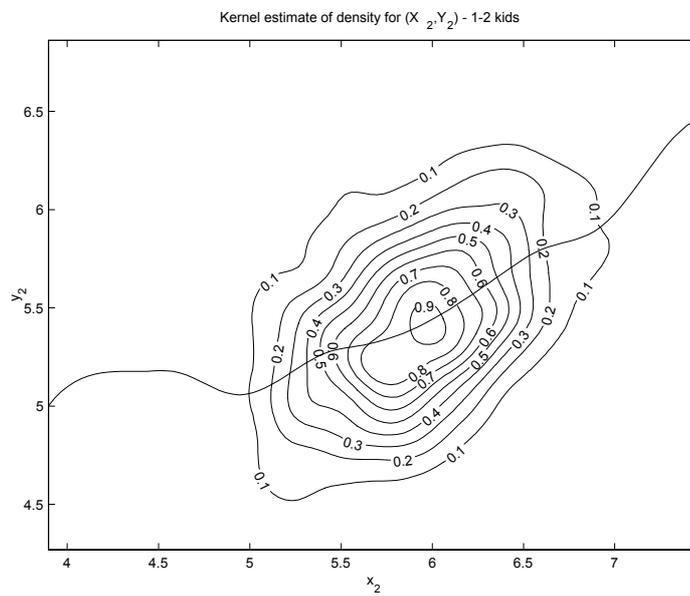


Figure C.8: Kernel estimate of density for  $(X_2, Y_2)$ ; series estimate of  $E[Y_2|X_2]$

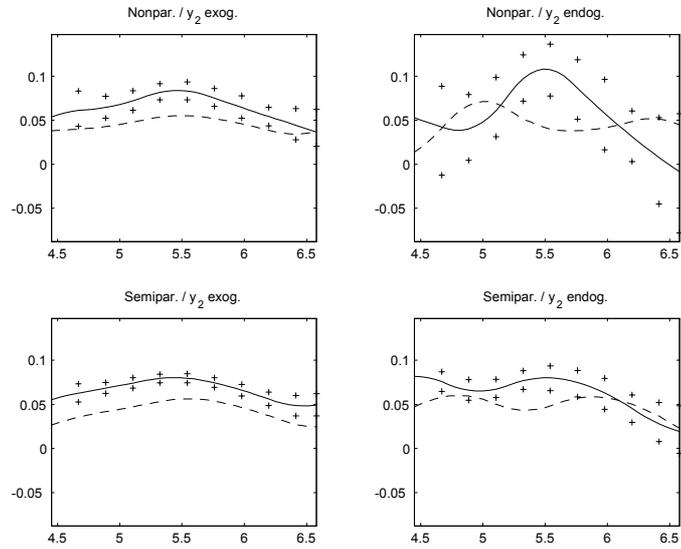


Figure C.9: Engel curve for alcohol. -- w/ children, — w/o children, ++ 95%-confidence bands

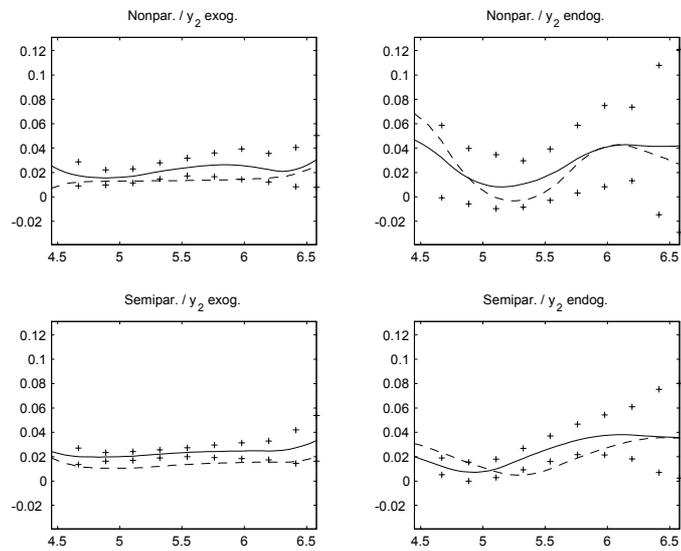


Figure C.10: Engel curve for fares. -- w/ children, — w/o children, ++ 95%-confidence bands

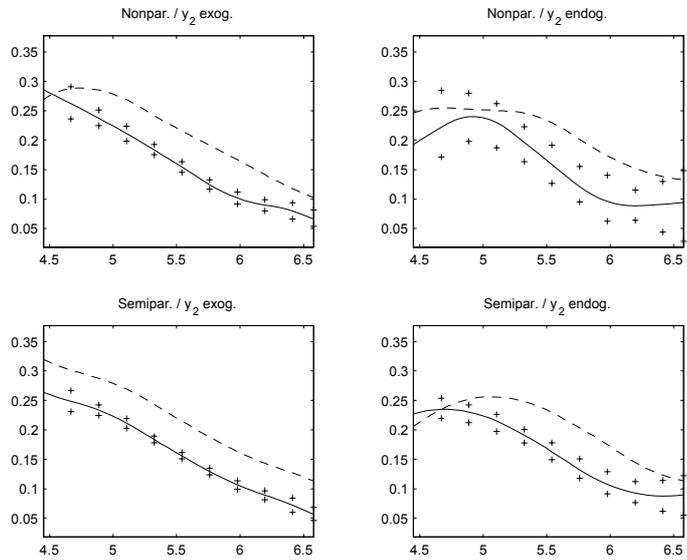


Figure C.11: Engel curve for food-in. - - w/ children, — w/o children, ++ 95%-confidence bands

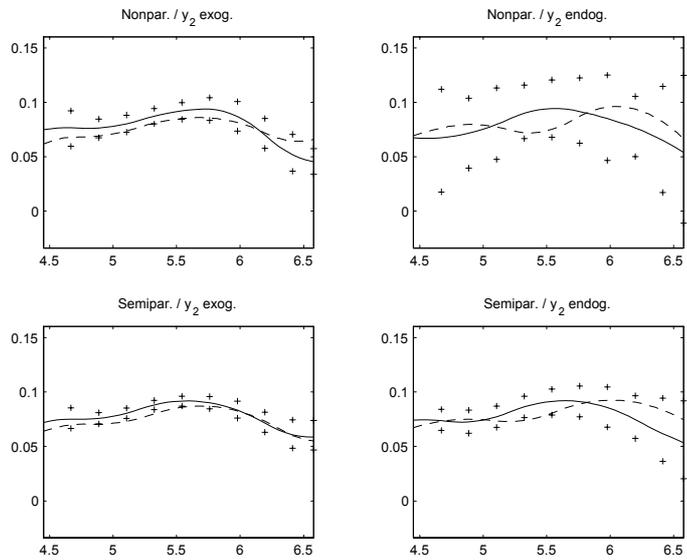


Figure C.12: Engel curve for food-out. - - w/ children, — w/o children, ++ 95%-confidence bands

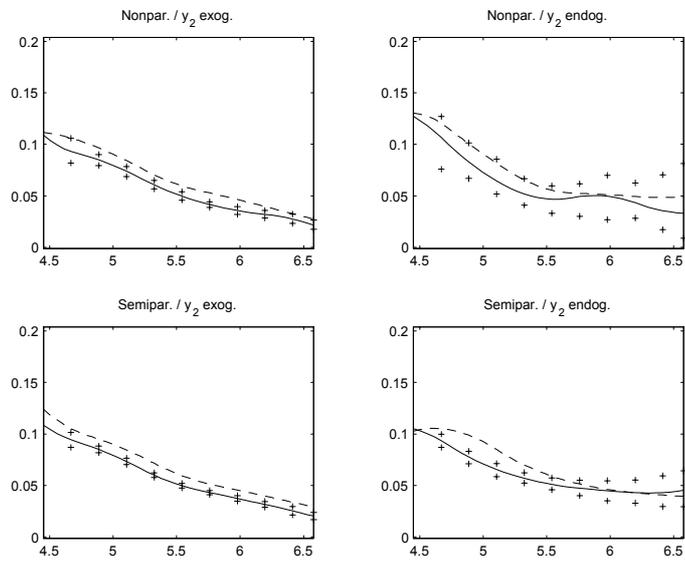


Figure C.13: Engel curve for fuel. -- w/ children, — w/o children, ++ 95%-confidence bands

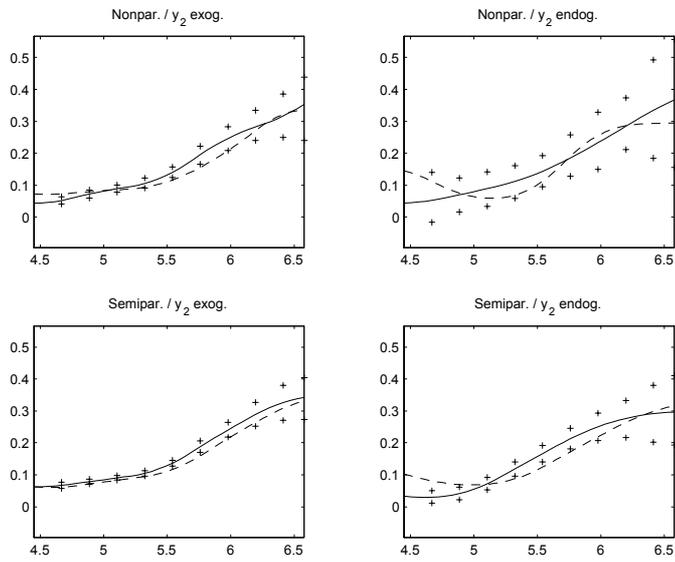


Figure C.14: Engel curve for leisure. -- w/ children, — w/o children, ++ 95%-confidence bands

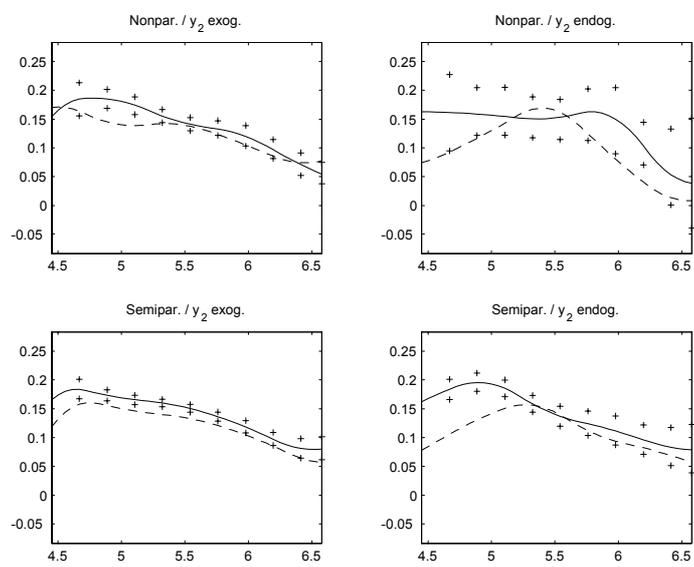


Figure C.15: Engel curve for motor. - - w/ children, — w/o children, ++ 95%-confidence bands