SEMI-NONPARAMETRIC IV ESTIMATION OF SHAPE-INVARIANT ENGEL CURVES[1]

By Richard Blundell, Xiaohong Chen and Dennis Kristensen

This paper studies a shape-invariant Engel curve system with endogenous total expenditure, in which the shape-invariant specification involves a common shift parameter for each demographic group in a pooled system of nonparametric Engel curves. We focus on the identification and estimation of both the nonparametric shapes of the Engel curves and the parametric specification of the demographic scaling parameters. The identification condition relates to the bounded completeness, and 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 could have unbounded support. Root-n asymptotic normality and semiparametric efficiency of the parametric components are also given under a set of 'low-level' sufficient conditions. Our empirical application using the UK Family Expenditure Survey shows the importance of adjusting for endogeneity in terms of both the nonparametric curvatures and demographic parameters of systems of Engel curves.

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

1. INTRODUCTION

The empirical analysis of consumer behavior represents an important area for the application of semiparametric and nonparametric methods in economics. The Engel curve relationship, which describes the expansion path for commodity demands as the household's budget increases, is a key example of this and one that lies at the heart of the study of consumer behavior. Historically parametric specifications have been based on the Working-Leser (Working, 1943; Leser, 1963) form

for the Engel curve in which the budget share is linear in the logarithm of the total budget. This shape underpins the popular Almost Ideal and Translog models of consumer behavior developed by Deaton and Muellbauer (1980) and Jorgenson, Lau and Stoker (1982), respectively. More recent empirical studies have suggested further nonlinearities in the total budget variable are required to capture observed behavior at the microeconomic level (see, for example, Hausman, Newey, Ichimura and Powell, 1991; Lewbel, 1991; Banks, Blundell and Lewbel, 1997) and nonparametric methods are now commonly used in application (see Hausman and Newey, 1995; Deaton, 1997). Blundell, Browning and Crawford (2003, 2004) show how to use nonparametric Engel curves together with the Afriat-Varian analysis of revealed preference to identify consumer responses to relative price changes across the income distribution.

The motivation for the Instrumental Variable (IV) estimator developed in this paper comes from the endogeneity of the total budget variable in the analysis of consumer Engel curves. This variable is the total expenditure allocated by the consumer to the sub-group of commodities under study, e.g. non-durables and services. As such it is also a choice variable in the consumer's allocation of income across consumption goods and savings. Consequently, it is very likely to be jointly determined with individual demands and an endogenous regressor in the estimation of consumer expansion paths. If total expenditure is endogenous for individual commodity demands, then the conditional mean estimated by nonparametric Least Squares (LS) regression does not identify the economically meaningful 'structural' Engel curve relationship. The 'statistical' Engel curve does not recover the correct shape necessary for the analysis of expansion paths or revealed preference. However, the allocation model of income to individual consumption goods and to savings does suggest that exogenous sources of income will provide suitable instrumental variables for total expenditure in the Engel curve regression.

Our focus on semi-nonparametric IV estimation is due to the need to pool nonparametric Engel curves across households of different demographic types while allowing for the endogeneity of total expenditure. The Engel curve relationship is well known to differ by demographic type, see, for example, Blundell (1988). This variation is often used to estimate equivalence scales (see, for example, Pendakur, 1998). To capture a wide range of income variation in consumer expansion paths and also to estimate equivalence scales it is typical to pool Engel curves across different demographic types of households. Recently Blundell, Browning and Crawford (2003) have shown that, in a nonparametric budget share specification, demographics cannot enter additively into each Engel curve equation while retaining consistency with consumer optimization theory; they must also enter so as to scale

the total expenditure variable inside the nonparametric Engel curve for each commodity. A partially linear semiparametric formulation (Robinson, 1988) is therefore ruled out. An attractive alternative semiparametric formulation, which is also consistent with consumer theory, is one that is based on the shape-invariant specification (see Härdle and Marron, 1990; Pinkse and Robinson, 1995; Blundell, Duncan and Pendakur, 1998) in which demographics simply shift and scale each demand function without altering its overall shape. The aim of the present paper is to extend the existing work on shape-invariant Engel curves to allow for endogenous total expenditure regressor. Our 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. In a semi-nonparametric regression framework of the type adopted here, there are two alternative approaches to estimation under endogeneity - the IV and the Control Function (CF) approaches. Here we develop the IV approach for this semi-nonparametric Engel curve case.[2]

The IV approach is investigated in Newey and Powell (2003), Darolles, Florens and Renault (2006) and Hall and Horowitz (2005) 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 method of Newey and Powell (2003) and Ai and Chen (2003). Existing papers on Engel curve models typically consider kernel based methods assuming exogenous total expenditure. In this paper, given the endogeneity of total expenditure and the shape-invariant semi-nonparametric specification, we argue that the sieve method offers an attractive alternative to the kernel based methods. This is because the sieve method is not only easier to implement numerically but also capable of achieving the optimal convergence rates simultaneously for both nonparametric and parametric components of the model specification. Moreover, this semi-nonparametric form is common in economic applications.

A nonparametric IV regression is a difficult ill-posed inverse problem and has not, to our knowledge, 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 nontrivial. While Newey and Powell (2003) provide consistency of the sieve nonparametric IV estimators, and Ai and Chen (2003) obtain root-n asymptotic 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. Moreover, neither of

these two papers provide convergence rates of the sieve nonparametric IV estimator under root mean squared metric; nor do they discuss details of implementation and empirical application.

In our analysis of the semi-nonparametric shape-invariant Engel curve model, 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 (2006). 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 new 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 (2006) and Hall and Horowitz (2005).[3] Their estimation procedures and the sufficient conditions for convergence rates they derive are different from ours. In particular, they assume that the endogenous regressor has bounded support, while we allow the endogenous regressor to have unbounded support, which is natural in the semiparametric shape-invariant Engel curve case with endogenous log-total expenditure.

In our convergence rate analysis, 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 sieve nonparametric LS regression, but, for the sieve nonparametric IV regression, it is always greater than one, and increases with the complexity of the sieve space as well as the smoothness of the conditional expectation operator. The greater the "sieve measure of ill-posedness", the bigger is the variance and the slower is the mean squared convergence rate of the sieve IV estimator.

Two Monte Carlo studies are included to assess the performance of the sieve nonparametric IV estimator. The first simulation is designed to mimic the subsample of household without children from the British Family Expenditure Survey (FES) data, which is the data set used in our empirical application. 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. The second Monte Carlo design is similar to the first except that we draw the endogenous regressor and the instrumental variable jointly from a bivariate Gaussian density. This leads to a severely ill-posed inverse problem and the sieve measure of ill-posedness goes to infinity exponentially

fast, which translates into a logarithmic mean squared convergence rate of the sieve IV estimator. Nevertheless, the Monte Carlo results indicate that the slow convergence rates are mainly due to the large variances and not due to the biases. 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 variances 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 both simulations. These findings are consistent with our theoretical result on the convergence rate of sieve IV estimator. The Monte Carlo simulations also shed light on the choice of sieve bases and smoothing parameters, and demonstrate that the sieve IV estimator performs well even for the severely ill-posed inverse problem.

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 across eight groups of non-durables and services for a sample of families with and without children in the Family Expenditure Survey. This data records detailed information on expenditures, incomes and family composition and has been a central data source for many investigations 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. 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 interpretable estimate of the income equivalence scale.

The structure of the remaining paper is as follows. Section 2 specifies the semi-/nonparametric Engel curve model with endogenous total expenditure, and provides sufficient conditions for identification. Section 3 presents the Sieve Minimum Distance (SMD) procedure. Section 4 establishes root mean squared convergence rates of the sieve IV estimators of the nonparametric Engel curves. Section 5 obtains root-n asymptotic normality and efficiency of the estimators of the parametric parts in the system of shape-invariant Engel curves. Section 6 discusses the implementation of the SMD estimation of the system of shape-invariant Engel curves, and presents two Monte Carlo studies to assess the performance of the sieve nonparametric IV estimator. Section 7 presents the empirical application using the FES data, and Section 8 briefly concludes. All proofs are collected into the Appendix.

2. MODEL SPECIFICATION AND IDENTIFICATION

2.1. *The Model*

Let $\{(Y_{1il}, Y_{2i}, X_{1i})\}_{i=1}^{n}$ represent an i.i.d. sequence of n household observations on the budget share Y_{1il} of good $l = 1, ..., L \geq 1$ for each household i facing the same relative prices,[4] on the log of total expenditure Y_{2i}, and on a vector of household composition variables X_{1i}. Household expenditures typically display a large variation with demographic composition. For each commodity l, budget shares and total outlay are related by the general stochastic Engel curve

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

where G_l, $l = 1, ..., L$, are unknown functions that can be estimated using standard nonparametric LS regression method under the exogeneity of (Y_{2i}, X_{1i}) assumption: $E[\varepsilon_{il}|Y_{2i}, X_{1i}] = 0$. 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. Blundell, Browning and Crawford (2003, proposition 6) show that the following extended partially linear specification is consistent with consumer optimization theory:

$$Y_{1il} = h_l(Y_{2i} - \phi(X_{1i}'\theta_1)) + X_{1i}'\theta_{2,l} + \varepsilon_{il}, \tag{2}$$

where h_l, $l = 1, ..., L$, are unknown functions, $\phi(X_{1i}'\theta_1)$ is a known function up to a finite set of unknown parameters θ_1 and can be interpreted as the log of a general equivalence scale for household i; see, e.g., Pendakur (1998).[5] 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 θ_1 is the vector of corresponding equivalence scales. Notice that (2) reduces to an additive form for functions of the demographic variables X_{1i} only when h_l is linear. This corresponds to the Almost Ideal model or Translog models of Deaton and Muellbauer (1980) and Jorgenson, Lau and Stoker (1982). For nonlinear specifications of h_l, including the QUAIDS specification of Banks, Blundell and Lewbel (1997), the theoretical consistency result implies that the demographic terms must also enter in the function h_l as is the case for (2).

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_{1i} 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, e.g., Blundell (1988). There are other explanations for endogenous total expenditure. See Hausman, Newey, Ichimura and Powell (1991) and Newey (2001) for a measurement error story. In our application we consider households in which the head of household is working and 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). Blundell, Duncan and Pendakur (1998) have analyzed the parametric control function approach. In this paper, we consider the alternative nonparametric IV approach to solve the endogeneity problem. In particular, we consider semi-nonparametric IV estimation where $h_l(\cdot)$ is a unknown function and $\theta_1, \theta_{2,l}$ are unknown finite-dimensional parameters. Functions of X_{2i} are then used as instrumental variables. More precisely we shall assume:

$$E[\varepsilon_{il}|X_{1i}, X_{2i}] = 0, \quad l = 1, ..., L. \tag{3}$$

2.2. *Identification*

We first lay out the notation that will be adopted throughout the remaining discussion. Denote $Y_{1i} = (Y_{1i1}, ..., Y_{1iL})' \in \mathcal{R}^L$, $X_i = (X_{1i}', X_{2i}')' \in \mathcal{X}$ with $\dim(X_1), \dim(X_2) \geq 1$ and $Z_i = (Y_{1i}', Y_{2i}, X_i')'$. Let $\alpha \equiv (\theta, h_1, ..., h_L)$ denote all the unknown parameters of interest, and $\mathcal{A} \equiv \Theta \times \mathcal{H}_1 \times \cdots \times \mathcal{H}_L$ denote the parameter space, where $\theta \equiv (\theta_1', \theta_{2,1}', ..., \theta_{2,L}')'$ denote a vector of unknown finite-dimensional parameters and $\theta \in \Theta$, a compact subset of $\mathcal{R}^{d_\theta}$ with $d_\theta \equiv (1 + L)\dim(X_1)$, and $h_l \in \mathcal{H}_l$ denote unknown Engel curve associated with good l, $l = 1, ..., L$, where $\mathcal{H}_l$ is a subset of space of functions that are square integrable against the probability measure of Y_{2i} (to be specified later). Finally we denote $\rho \equiv (\rho_1, ..., \rho_L)' \in \mathcal{R}^L$, where for $l = 1, ..., L$,

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

with a known functional form $\phi(\cdot)$. For each household i facing the same relative prices and for goods $l = 1, .., L$, the Engel curve model satisfies (2) and (3) which we rewrite as:

$$E\left[\rho(Z_i, \alpha_o)|X_i\right] = 0, \tag{4}$$

where $\alpha_o \equiv (\theta_o, h_{o1}, ..., h_{oL}) \in \mathcal{A}$ is the true but unknown parameters of interest. For policy analysis, we would like to estimate θ_o, the Engel curve functions h_{ol}, and other linear functionals such as the average derivatives $E\left[\nabla h_{ol}(Y_{2i} - \phi(X_{1i}'\theta_{o1}))\right]$, $l = 1, ..., L$.

The first assumption is about identification of α_o.

ASSUMPTION I (IDENTIFICATION): $E[Y_{1il} - h_l(Y_{2i} - \phi(X_{1i}'\theta_1)) - X_{1i}'\theta_{2,l}|X_{1i}, X_{2i}] = 0$ for $l = 1, ..., L$ implies $\theta_1 = \theta_{o1}, \theta_{2,l} = \theta_{o2,l}$ and $h_l = h_{ol}$ a.s.

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 (4) and the following 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, $l = 1, .., L$, *and* ϕ *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* ϕ *is not periodic. Then Assumption I is satisfied.*

REMARK 1: (**i**) Condition (1) is normally referred to as *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 defined as: *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, the requirement that the conditional distribution of Y_2 given $X = (X_1, X_2)$ is complete is somewhat restrictive, since there are not many known families of distributions beyond the exponential family that are complete. Luckily, there are much larger families of distributions that are bounded complete. For instance, within the location family of absolutely continuous distributions (with respect to Lebesgue measure), they are bounded complete if and only if 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. In particular, a location family of absolutely continuous distributions with compact supports are bounded complete but of which the only complete distribution is a Dirac measure. Also, a family of nontrivial finite scale mixtures of the standard Gaussian distribution is bounded complete but not complete. See, e.g., Lehmann (1986, page 173), Hoeffding (1977) and Mattner (1993) for more examples of distributions which are bounded complete but not complete. For the identification of g in a purely nonparametric IV regression model $E[Y_1 - g(Y_2)|X] = 0$, Newey and Powell (2003)

and Darolles, Florens and Renault (2006) impose the "completeness" condition. However, if one restricts the unknown functions g to be sup-norm bounded (i.e. $\sup_y |g(y)| \leq$ const. $< \infty$), then $E[Y_1 - g(Y_2)|X] = 0$ uniquely identifies the unknown g if and only if the conditional distribution of Y_2 given X is bounded complete. Therefore, if economic theory or any prior information suggests the unknown $g()$ should be sup-norm bounded, then it will be much easier to identify the unknown g from $E[Y_1 - g(Y_2)|X] = 0$. For example, if some economic theory implies that g should be continuous and that Y_2 has bounded support, then g is identified from $E[Y_1 - g(Y_2)|X] = 0$ as long as the conditional distribution of Y_2 given X is bounded complete. Here, for the Engel curve application, even though Y_2 (log-total expenditure) could have unbounded support, but since Engel curves should be all bounded below by zero and above by one, it suffices to impose the weaker "bounded completeness" identification condition.

(ii) Conditions (2) - (5) are satisfied in our Engel curve study. In the empirical application in Section 7, we take $X_{1i} = 1$ or 0 to indicate if the i-th family has children or not. 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. Since our instrumental variable X_2 is a monotonic transformation of log gross earnings into [0,1] support, the dependence between Y_2 and X_2 is still strong. Finally, as noted in the introduction, nonlinear behavior in the Engel curve relationship is commonplace for many goods; see, e.g., Hausman, Newey, Ichimura and Powell (1991), Lewbel (1991) and Banks, Blundell and Lewbel (1997).

3. SIEVE MINIMUM DISTANCE ESTIMATION

Our estimation method corresponds to the one in Ai and Chen (2003) for semiparametric conditional moment restrictions and is similar to Newey and Powell's (2003) for nonparametric IV regression. First we approximate the unknown functions $h_l \in \mathcal{H}_l$ by $h_{l,n} \in \mathcal{H}_{l,n}$, $l = 1, ..., L$, where $\mathcal{H}_{l,n}$ is a sieve space for $\mathcal{H}_l$, in particular, we let $\mathcal{H}_{l,n}$ be some finite-dimensional approximation space (e.g. Fourier series, splines, wavelets, etc.) that becomes dense in $\mathcal{H}_l$ as sample size $n \to \infty$. Then for arbitrarily fixed candidate value $\alpha = (\theta, h_{1,n}, ..., h_{L,n})$ in the sieve parameter space $\mathcal{A}_n \equiv \Theta \times \mathcal{H}_{1,n} \times \cdots \times \mathcal{H}_{L,n}$, we estimate the population conditional moment function $m(x, \alpha) \equiv (m_1(x, \alpha), ..., m_L(x, \alpha))' \equiv E[\rho(Z_i, \alpha)|X_i = x]$ nonparametrically by $\widehat{m}(x, \alpha) \equiv (\widehat{m}_1(x, \alpha), ..., \widehat{m}_L(x, \alpha))'$. Finally we estimate the θ and the unknown sieve coefficients of $h_{l,n}$,

$l = 1, ..., L$, jointly by applying the SMD procedure:

$$\min_{\alpha \in \mathcal{A}_n} \frac{1}{n} \sum_{i=1}^{n} \widehat{m}(X_i, \alpha)' [\widehat{\Sigma}(X_i)]^{-1} \widehat{m}(X_i, \alpha), \tag{5}$$

where $\widehat{\Sigma}$ is some consistent estimator of some positive definite weighting matrix Σ. To obtain a semiparametric efficient estimator of θ_o, we may follow the three-step procedure proposed in the first version of Ai and Chen (2003):

STEP 1: Compute the identity weighted SMD estimator $\widehat{\alpha}_n = \arg\min_{\alpha \in \mathcal{A}_n} \sum_{i=1}^{n} \widehat{m}(X_i, \alpha)' \widehat{m}(X_i, \alpha)$.

STEP 2: Compute a nonparametric estimator $\widehat{\Sigma}_o(X)$ of the optimal weighting matrix $\Sigma_o(X) \equiv Var[\rho(Z, \alpha_o)|X]$ using $\widehat{\alpha}_n$ and any nonparametric regression procedures (such as kernel, nearest neighbor or series).

STEP 3: Obtain the optimally weighted SMD estimator $\widetilde{\alpha}_n = (\widetilde{\theta}_n, \widetilde{h}_{1,n}, ..., \widetilde{h}_{L,n}) \in \mathcal{A}_n$ by

$$\min_{\alpha \in \mathcal{A}_n} \frac{1}{n} \sum_{i=1}^{n} \widehat{m}(X_i, \alpha)' [\widehat{\Sigma}_o(X_i)]^{-1} \widehat{m}(X_i, \alpha), \tag{6}$$

using $\widehat{\alpha}_n = (\widehat{\theta}_n, \widehat{h}_{1,n}, ..., \widehat{h}_{L,n}) \in \mathcal{A}_n$ as the starting point.

REMARK 2: **(i)** An equivalent but sometimes computationally simpler alternative to the procedure (5) is the profile SMD procedure: First, for each fixed $\theta \in \Theta$, we compute $\widehat{h}_{l,n}(\theta; \cdot)$, $l = 1, ..., L$ as the solution to

$$\min_{h_l \in \mathcal{H}_{l,n}, l=1,...,L} \sum_{i=1}^{n} \widehat{m}(X_i, \theta, h_1, ..., h_L)' [\widehat{\Sigma}(X_i)]^{-1} \widehat{m}(X_i, \theta, h_1, ..., h_L).$$

Second, we compute $\widehat{\theta}_n$ as the solution to

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

and estimate $h_{ol}(\cdot)$ by $\widehat{h}_{l,n} = \widehat{h}_{l,n}(\widehat{\theta}_n; \cdot)$ for $l = 1, ..., L$.

(ii) If total expenditure is assumed to be exogenous, then Y_2 is a "perfect IV" and we have $E[\rho(Z_i, \alpha_o)|X_{1i}, Y_{2i}] = 0$. In this case, we do not need to estimate α_o via (5). Instead we can

apply the following sieve generalized least squares (GLS) procedure:

$$\min_{\alpha \in \mathcal{A}_n} \frac{1}{n} \sum_{i=1}^{n} \rho(Z_i, \alpha)' \widehat{\Sigma}(X_i)^{-1} \rho(Z_i, \alpha), \tag{7}$$

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

3.1. *Possible Sieve Bases for h*

Since $h_{ol}(\cdot)$, $l = 1, ..., L$ have the same argument $Y_2 - \phi(X_1'\theta_1)$ and similar smoothness, to simplify notation, we assume that they belong to the same function space, $h_{ol} \in \mathcal{H} = \mathcal{H}_l$, hence they can be approximated by the same sieve space $\mathcal{H}_n = \mathcal{H}_{l,n}$ for $l = 1, ..., L$. In our empirical application Y_2 is log total expenditure, and a simple kernel nonparametric estimation of the density of Y_2 using our data set reveals an approximate 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 $h_{ol} \in \mathcal{H}$.

We now introduce several typical spaces for real-valued functions on $\mathcal{Y}$, where $\mathcal{Y}$ is either $\mathcal{R}$ or a bounded interval of $\mathcal{R}$. Let k be a nonnegative integer and $C^k(\mathcal{Y})$ be the space of $k-$times continuously differentiable functions. For any real-valued $r > 0$, let $k = [r]$ be the largest nonnegative integer such that $[r] < r$ and set $r' \equiv r - [r] \in (0, 1]$. A function h on $\mathcal{Y}$ is said to be in Hölder space $\Lambda^r(\mathcal{Y})$ if it is in $C^{[r]}(\mathcal{Y})$ and its $[r]-$th derivative, $\nabla^{[r]}h$, satisfies a Hölder condition with exponent r' (i.e., $|\nabla^{[r]}h(x) - \nabla^{[r]}h(y)| \leq c|x - y|^{r'}$ for all $x, y \in \mathcal{Y}$ for some finite $c > 0$). The space $\Lambda^r(\mathcal{Y})$ becomes a Banach space under the Hölder norm:

$$||h||_{\Lambda^r} \equiv \max_{j \leq [r]} \sup_x |\nabla^j h(x)| + \sup_{x \neq y} \frac{|\nabla^{[r]}h(x) - \nabla^{[r]}h(y)|}{|x - y|^{r-[r]}} < \infty.$$

Let $L_2(\mathcal{Y})$ be the space of functions with finite $||h||_{L_2} \equiv \sqrt{\int_{\mathcal{Y}} |h(y)|^2 dy}$. Let $W_2^k(\mathcal{Y})$ be the Sobolev space of functions in $L_2(\mathcal{Y})$ with their derivatives up to order k also in $L_2(\mathcal{Y})$; it becomes a Banach space under the norm $||h||_{W_2^k} \equiv \sum_{j=0}^{k} ||\nabla^j h||_{L_2} < \infty$. For any real-valued $r > 0$ with $k = [r] < r < k + 1$, let $W_2^r(\mathcal{Y})$ be the (fractional) Sobolev space of functions in $L_2(\mathcal{Y})$ with finite norm $||h||_{W_2^r}$:

$$||h||_{W_2^r} \equiv \sum_{j=0}^{[r]} ||\nabla^j h||_{L_2} + \left(\int_{\mathcal{Y}} \int_{\mathcal{Y}} \frac{|\nabla^{[r]}h(x) - \nabla^{[r]}h(y)|^2}{|x - y|^{2(r-[r])+1}} dx dy \right)^{1/2} < \infty. \tag{8}$$

Let $w()$ be a positive continuous weight function on $\mathcal{Y}$, and $L_2(\mathcal{Y}, w)$ be the weighted space of functions with finite norm $||h||_{L_2(\mathcal{Y},w)} \equiv ||h \times w^{1/2}||_{L_2}$. Denote $W_2^r(\mathcal{Y}, w)$ as the weighted Sobolev space of functions with finite norm $||h||_{W_2^r(\mathcal{Y},w)} \equiv ||h \times w^{1/2}||_{W_2^r}$. Also define a mixed weighted Sobolev space $W_2^r(\mathcal{Y}; w, leb) \equiv \{h \in L_2(\mathcal{Y}, w) : ||\nabla h||_{W_2^{r-1}} < \infty\}$.

Let $\Lambda_c^r(\mathcal{Y}) \equiv \{h \in \Lambda^r(\mathcal{Y}) : ||h||_{\Lambda^r} \leq c\}$ be a Hölder ball (of radius c) with smoothness r, and $W_{2,c}^r(\mathcal{Y}; w, leb) \equiv \left\{h \in W_2^r(\mathcal{Y}; w, leb) : ||h||_{L_2(\mathcal{Y},w)} + ||\nabla h||_{W_2^{r-1}} \leq c\right\}$ as a weighted Sobolev ball (of radius c) with smoothness r. Since Engel curves h_{ol}, $l = 1, ..., L$ are bounded between zero and one, and since consumer demand theory and many empirical studies suggest that h_{ol}, $l = 1, ..., L$ are reasonably smooth, we could assume either $h_{ol} \in \mathcal{H}^a = \{h \in \Lambda_c^r(\mathcal{R}) : 0 \leq h \leq 1\}$ for some $r > 1/2$, or $h_{ol} \in \mathcal{H}^b = \{h \in W_{2,c}^r(\mathcal{R}; f_{Y_2}, leb) : 0 \leq h \leq 1\}$ for some $r > 1$, here f_{Y_2} denotes the marginal density of Y_2. Then a sieve space $\mathcal{H}_n$ could take either the form

$$\mathcal{H}_n^a = \left\{ \begin{array}{c} h_n : \mathcal{R} \to [0,1], \ \sup_y |\nabla^{[r]} h_n(y)| \leq c, \\ h_n(Y_2 - \phi(X_1'\theta_1)) = \psi^{k_n}(Y_2 - \phi(X_1'\theta_1))'\Pi \end{array} \right\}, \tag{9}$$

or the form

$$\mathcal{H}_n^b = \left\{ \begin{array}{c} h_n : \mathcal{R} \to [0,1], \ ||\nabla^{[r]} h_n||_{L_2} \leq c, \\ h_n(Y_2 - \phi(X_1'\theta_1)) = \psi^{k_n}(Y_2 - \phi(X_1'\theta_1))'\Pi \end{array} \right\}, \tag{10}$$

where $\psi^{k_n}(\cdot)$ is a $k_n \times 1$-vector of known basis functions that are at least $\gamma = ([r] + 1)-$times differentiable, such as Fourier series, wavelets or B-splines, and Π is a $k_n \times 1$-vector of unknown sieve coefficients.

In theoretical sections 4 and 5 we have used wavelet sieve basis for $h_n \in \mathcal{H}_n$. Let $\gamma > 0$ be an integer. A real-valued function ψ is called a "mother wavelet" of degree γ if it satisfies: (a) $\int_{\mathcal{R}} y^k \psi(y) dy = 0$ for $0 \leq k \leq \gamma$; (b) ψ and all its derivatives up to order γ decrease rapidly as $|y| \to \infty$; (c) $\{2^{k/2}\psi(2^k y - j) : k, j \in \mathbb{Z}\}$ forms a Riesz basis of $L_2(\mathcal{R})$, that is, the linear span of $\{2^{k/2}\psi(2^k y - j) : k, j \in \mathbb{Z}\}$ is dense in $L_2(\mathcal{R})$ and

$$\left\| \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_{kj} 2^{k/2} \psi(2^k y - j) \right\|_{L_2(\mathcal{R})}^2 \asymp \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |a_{kj}|^2$$

for all doubly bi-infinite square-summable sequence $\{a_{kj} : k, j \in \mathbb{Z}\}$. (The notation $a_k \asymp b_k$ means $c_1 a_k \leq b_k \leq c_2 a_k$ for some finite positive constants c_1, c_2 that do not depend on the sequences $\{a_k\}$,

$\{b_k\}$.) A scaling function φ is called a "father wavelet" of degree γ if it satisfies: (a') $\int_{\mathcal{R}} \varphi(y)dy = 1$; (b') φ and all its derivatives up to order γ decrease rapidly as $|y| \to \infty$; (c') $\{\varphi(y-j) : j \in \mathbb{Z}\}$ forms a Riesz basis for a closed subspace of $L_2(\mathcal{R})$.

Orthogonal wavelets. Given an integer $\gamma > 0$, there exist a father wavelet φ of degree γ and a mother wavelet ψ of degree γ, both compactly supported, such that for any integer $k_0 \geq 0$, any function h in $L_2(\mathcal{R})$ has the following wavelet $\gamma-$ regular multiresolution expansion:

$$h(y) = \sum_{j=-\infty}^{\infty} a_{k_0 j}\varphi_{k_0 j}(y) + \sum_{k=k_0}^{\infty}\sum_{j=-\infty}^{\infty} b_{kj}\psi_{kj}(y), \quad y \in \mathcal{R},$$

where

$$\begin{aligned} a_{kj} &= \int_{\mathcal{R}} h(y)\varphi_{kj}(y)dy, \quad \varphi_{kj}(y) = 2^{k/2}\varphi(2^k y - j), \quad y \in \mathcal{R}, \\ b_{kj} &= \int_{\mathcal{R}} g(y)\psi_{kj}(y)dy, \quad \psi_{jk}(y) = 2^{k/2}\psi(2^k y - j), \quad y \in \mathcal{R}, \end{aligned}$$

and $\{\varphi_{k_0 j}, j \in \mathbb{Z}; \psi_{kj}, k \geq k_0, j \in \mathbb{Z}\}$ is an orthonormal basis of $L_2(\mathcal{R})$; see Meyer (1992, theorem 3.3). For an integer $K_n > k_0$, we consider the finite-dimensional linear space spanned by this wavelet basis of order $\gamma > r$

$$h_n(y) = \psi^{k_n}(y)'\Pi = \sum_{j=0}^{2^{K_n}-1} \alpha_{K_n,j}\varphi_{K_n,j}(y), \quad k_n = 2^{K_n}.$$

Cardinal B-spline wavelets. The cardinal B-spline wavelets of order $\gamma > r$

$$h_n(y) = \psi^{k_n}(y)'\Pi = \sum_{k=0}^{K_n}\sum_{j\in\mathcal{K}_n} \pi_{kj}2^{k/2}B_\gamma(2^k y - j), \quad k_n = 2^{K_n} + 1, \tag{11}$$

where $B_\gamma(\cdot)$ is the cardinal B-spline of order $\gamma > r$,

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

which is $\gamma - 1$ times differentiable and has support on $[0, \gamma]$. For any fixed integer $k = 0, 1, ..., K_n$, $\mathcal{K}_n$ is the set consisting of those j's such that the support of $z \to B_\gamma(2^k z - j)$ overlaps with the empirical support of $Y_2 - \phi(X_1'\theta_1)$, $j = \pm 1, \pm 2, ...$ The compact support of $B_\gamma(\cdot)$ ensures that $\#\mathcal{K}_n$

is finite for any fixed k.

In simulations and empirical studies we have also used the following two sieve bases for $h_n \in \mathcal{H}_n$:

Polynomial splines of order $q_n > r$

$$h_n(y) = \psi^{k_n}(y)'\Pi = \sum_{j=0}^{q_n} \pi_j (y)^j + \sum_{k=1}^{r_n} \pi_{q_n+k} (y - \nu_k)_+^{q_n}, \quad k_n = q_n + r_n + 1, \tag{12}$$

where $(y-\nu)_+^q = \max\{(y-\nu)^q, 0\}$ and $\{\nu_k\}_{k=1,...,r_n}$ are the knots. In the empirical application, for any given number of knots value r_n, the knots $\{\nu_k\}_{k=1,...,r_n}$ are simply chosen as the empirical quantiles of Y_2, i.e., $\nu_k = k/(r_n+1)$-th quantile of Y_2.

Hermite polynomials of order $k_n - 1$

$$h_n(y) = \psi^{k_n}(y)'\Pi = \sum_{j=0}^{k_n-1} \pi_j (y - \nu_1)^j \exp\left\{ -\frac{(y-\nu_1)^2}{2\nu_2^2} \right\}, \tag{13}$$

where in the Monte Carlo study ν_1 and ν_2^2 are chosen as the sample mean and sample variance of the data $\{Y_{2i}\}_{i=1}^n$.

3.2. *Sieve Least Squares Estimation for m and Σ_o*

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 present the sieve LS estimation as illustration.

For each fixed (X,α), we approximate $m(X,\alpha) = E[\rho(Z,\alpha)|X]$ by the function $m_n(X,\alpha) = \sum_{j\in\mathcal{J}_n} a_j(\alpha) p_{0j}(X)$, where p_{0j} are some known basis functions, and $J_n \equiv \#(\mathcal{J}_n) \to \infty$ slowly as $n \to \infty$. We write $m_n(X,\alpha) = p^{J_n}(X)'A$ where $p^{J_n}(X) = (p_{01}(X), ..., p_{0J_n}(X))'$, and $P = (p^{J_n}(X_1), ..., p^{J_n}(X_n))'$. Then the sieve LS estimator of $m(X,\alpha)$ is:

$$\widehat{m}(X,\alpha) = p^{J_n}(X)'(P'P)^- \sum_{i=1}^n p^{J_n}(X_i)\rho(Z_i,\alpha), \tag{14}$$

where $(P'P)^-$ denotes the generalized inverse of the matrix $P'P$. Similarly we can compute a sieve LS estimator $\hat{\Sigma}_o(X)$ of $\Sigma_o(X)$ in STEP 2 by regressing $\rho(Z,\hat{\alpha}_n)\rho(Z,\hat{\alpha}_n)'$ on $p^{J_n}(X)$:

$$\hat{\Sigma}_o(X) = p^{J_n}(X)'(P'P)^- \sum_{i=1}^n p^{J_n}(X_i)\rho(Z_i,\hat{\alpha})\rho(Z_i,\hat{\alpha})', \tag{15}$$

where $\hat{\alpha}_n$ is the SMD estimator obtained in STEP 1 or any other consistent estimator of α_o.

Many known sieve bases could be used as $p^{J_n}(X)$. In our empirical application $X = (X_1, X_2)'$ is a bivariate vector, where $X_1 \in \{0, 1\}$ and X_2 is the normal transformation of the log of gross earnings: $X_2 = \Phi(\text{log-gross earnings}) \in [0, 1]$. We take

$$p^{J_n}(X) \equiv \left(B^{J_{2n}}(X_2)', X_1 \times B^{J_{2n}}(X_2)'\right)', \quad J_n = 2J_{2n}, \tag{16}$$

where $B^{J_{2n}}(X_2)$ is a $J_{2n} \times 1$-vector of univariate B-splines or polynomial splines or wavelets or cosine series $\{\cos(\pi j X_2) : j = 0, 1, ..., J_{2n} - 1\}$.

4. CONVERGENCE RATE of SIEVE NONPARAMETRIC IV ESTIMATOR

In this section we study the convergence rate of the sieve IV estimator $\widehat{h}_{nl}$ of the unknown Engel curve h_{ol} using the subsample of $X_1 = 0$, where $\widehat{h}_{nl}$ is computed using $\widehat{\Sigma}(X) = I_L$ without loss of generality. We establish the convergence rate of $\widehat{h}_{n,l}$ under the mean squared error metric:

$$||h_l - h_{ol}||^2_{Y_2} = E\left[\{h_l(Y_2 - \phi(0)) - h_{ol}(Y_2 - \phi(0))\}^2\right].$$

In the following we denote $k_n \equiv \dim(\mathcal{H}_n)$, and assume:

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

ASSUMPTION 2: (i) $h_{ol} \in \mathcal{H} \equiv \{h \in \Lambda^r_c(\mathcal{R}) : 0 \leq h \leq 1\}$ for $l = 1, ..., L$ for some $r > 1/2$; (ii) $E[|Y_2|^{2a}] < \infty$ for some $a > r$.

ASSUMPTION 3: For any x_1 in the support of X_1, $E[Y_{1l}|X_1 = x_1, X_2 = \cdot], l = 1, ..., L$ belong to $\Lambda^{r_m}_c(\mathcal{X}_2)$, $r_m > 1/2$, and $E[h_n(Y_2 - \phi(X_1'\theta_1))|X_1 = x_1, X_2 = \cdot]$ belongs to $\Lambda^{r_m}_c(\mathcal{X}_2)$ for any $h_n \in \mathcal{H}_n$.

ASSUMPTION 4: (i) the smallest and the largest eigenvalues of $E\{B^{J_{2n}}(X_2)B^{J_{2n}}(X_2)'\}$ are bounded and bounded away from zero for each J_{2n}; (ii) $B^{J_{2n}}(X_2)$ is either a cosine series or a B-spline basis of order γ_b, with $\gamma_b > r_m > 1/2$; (iii) the density of X_2 is continuous, bounded and bounded away from zero over its support $\mathcal{X}_2$, which is a compact interval with non-empty interior.

ASSUMPTION 5: (i) $k_n \to \infty$, $J_{2n}/n \to 0$; (ii) $\lim_{n\to\infty}(J_{2n}/k_n) = c_0 > 1$; $\lim_{n\to\infty}(k_n^2/n) = 0$.

We define τ_n as a *sieve measure of ill-posedness*:

$$\tau_n \equiv \sup_{h \in \mathcal{H}_n : h \neq 0} \frac{\sqrt{E\{h(Y_2)\}^2}}{\sqrt{E\{E[h(Y_2)|X_1 = 0, X_2]\}^2}}, \tag{17}$$

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 $E[h(Y_2)|X_1 = 0, X_2] = h(Y_2)$ for all $h \in \mathcal{H}_n$). For example $\tau_n = 1$ when Y_2 is exogenous (and we take $Y_2 = X_2$). We note that the τ_n measure of ill-posedness, as given in (17), depends on the choice of sieve space $\mathcal{H}_n$. This is why we call it a "sieve measure of ill-posedness".

ASSUMPTION 6: For $l = 1, ..., L$, there is a $\{\psi^{k_n}\}'\Pi_o^l \in \mathcal{H}_n$ such that $\tau_n^2 \times E\{E[h_{ol}(Y_2) - \{\psi^{k_n}(Y_2)\}'\Pi_o^l|X_1 = 0, X_2]\}^2 \leq const.||h_{ol} - \{\psi^{k_n}\}'\Pi_o^l||_{Y_2}^2$.

THEOREM 2: *Let $\widehat{\alpha}$ be the identity weighted SMD estimator with the sieve space $\mathcal{H}_n$ given in (9)-(11). Suppose Assumptions 1, 2(i)(ii), 3, 4, 5(i)(ii) and 6 are satisfied. Then*

$$||\widehat{h}_{nl} - h_{ol}||_{Y_2} = O_p\left(k_n^{-r} + \tau_n \times \sqrt{k_n/n}\right) \quad \textit{for all } l = 1, ..., L.$$

(1) Mildly ill-posed case: if $\tau_n = O\left((k_n)^s L(k_n)\right)$ for some finite $s > 0$ and $L(k_n) = 0$ or a slowly varying function that goes to ∞ slower than any polynomial order, then

$$||\widehat{h}_{nl} - h_{ol}||_{Y_2} = O_p\left(n^{-\frac{r}{2(r+s)+1}} L(n^{\frac{1}{2(r+s)+1}})\right) \quad \textit{provided} \quad k_n = O\left(n^{\frac{1}{2(r+s)+1}}\right);$$

(2) Severely ill-posed case: if $\tau_n = O\left(\exp(k_n)\right)$ and $r_m = \infty$, then

$$||\widehat{h}_{nl} - h_{ol}||_{Y_2} = O_p([\log n]^{-r}) \quad \textit{provided} \quad k_n = O(\log n).$$

REMARK 3: (**i**) For exogenous total expenditure Y_2, we have $\tau_n = 1$. Theorem 2 implies $||\widehat{h}_{nl} - h_{ol}||_{Y_2} = O_p(k_n^{-r} + \sqrt{k_n/n})$. If $k_n = O(n^{1/(2r+1)})$ then $||\widehat{h}_{nl} - h_{ol}||_{Y_2} = O_p(n^{-r/(2r+1)})$, which coincides with the well-known optimal rate of Stone (1982) for nonparametric LS regression; see also Newey (1997, theorem 1) or Chen and Shen (1998, theorem 1). Comparing this rate for the exogenous Y_2 case to that for the endogenous Y_2 case in Theorem 2, we note that the bias part k_n^{-r} is of the same order, however, the standard deviation part blows up from $\sqrt{k_n/n}$ in the exogenous case to $\tau_n\sqrt{k_n/n}$ in the endogenous case.

(**ii**) Without further assumption on the conditional expectation operator $E[h(Y_2)|X_1 = 0, X_2]$, one generally does not know the speed of divergence of the sieve measure of ill-posedness τ_n. Nevertheless,

τ_n can be easily estimated from the data by

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

where, for any fixed $h_n \in \mathcal{H}_n$, $\widehat{E}[h_n(Y_2)|X_1 = 0, X_2]$ 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 $B^{J_{2n}}(X_2)$. In most applications, $\log(\tau_{k_n})$ behaves either as $s \times \log(k_n) + o(\log(k_n))$ for the mildly ill-posed case (with some finite $s > 0$), or as $c \times k_n + o(k_n)$ for the severely ill-posed case (with some finite $c > 0$). For $k_n = 2, 3, 4, ...$, we can compute $\log(\widehat{\tau}_{k_n})$, and regress it on either $\log(k_n)$ or k_n to estimate the speed of divergence of τ_{k_n}.

In the following we provide some sufficient conditions to bound the sieve measure of ill-posedness τ_n. Let f_{0,X_2,Y_2}, f_{0,X_2}, f_{0,Y_2} respectively denote the joint probability density of (X_2, Y_2) (with respect to Lebesgue measure on $\mathcal{X}_2 \times \mathcal{R}$), marginal densities of X_2 and Y_2, all conditioning on $X_1 = 0$. Denote the conditional expectation operator as $\{Th\}(X_2) \equiv E[h(Y_2)|X_2, X_1 = 0]$, which maps $L_2(\mathcal{R}, f_{0,Y_2})$ into $L_2(\mathcal{X}_2, f_{0,X_2})$. Denote the adjoint operator of T as T^*, $\{T^*g\}(Y_2) \equiv E[g(X_2)|Y_2, X_1 = 0]$, which maps $L_2(\mathcal{X}_2, f_{0,X_2})$ into $L_2(\mathcal{R}, f_{0,Y_2})$. Assumptions 1(iii) and 4(iii) imply that T and T^* are compact operators. Therefore T has the singular value decomposition $\{\mu_k; \phi_{1k}, \phi_{0k}\}_{k=1}^{\infty}$, where $\{\mu_k\}_{k=1}^{\infty}$ are the singular numbers arranged in non-increasing order ($\mu_k \geq \mu_{k+1} \searrow 0$), $\{\phi_{1k}(y_2)\}_{k=1}^{\infty}$ and $\{\phi_{0k}(x_2)\}_{k=1}^{\infty}$ are eigenfunction (orthonormal) bases for $L_2(\mathcal{R}, f_{0,Y_2})$ and $L_2(\mathcal{X}_2, f_{0,X_2})$ respectively. See Appendix A for details.

LEMMA 1: *Let Assumptions 1(iii) and 4(iii) hold. Then: (1) $\tau_n \geq 1/\mu_{k_n}$; (2) If the sieve space $\mathcal{H}_n$ spans the linear subspace* (in $L_2(\mathcal{R}, f_{0,Y_2})$) *generated by $\{\phi_{1k} : k = 1, ..., k_n\}$, then: (2.i) $\tau_n \leq 1/\mu_{k_n}$, and (2.ii) Assumption 6 is satisfied.*

REMARK 4: **(i)** If the joint density f_{0,X_2,Y_2} is bivariate Gaussian with non-zero constant correlation coefficient, then $\mu_k \asymp \exp(-ck)$ for some constant $c > 0$. This corresponds to the so-called severely ill-posed inverse problem. Noting that $E[Y_1 - h_{ol}(Y_2)|X_2] = 0$ if and only if $E[Y_1 - h_{ol}(Y_2)|\Phi(X_2)] = 0$ and the conditional expectations operator $E[\bullet|\Phi(X_2)]$ shares the same singular values $\{\mu_k\}$ as those of $E[\bullet|X_2]$, we could use Hermite polynomial sieve for $h_{ol}(Y_2)$ and Lemma 1 implies that $\tau_n \asymp 1/\mu_{k_n} \asymp \exp(ck_n)$. Notice that $r_m = \infty$ in this case, hence we could use cosine sieve for $E[Y_1 - h_l(Y_2)|\Phi(X_2)]$ to satisfy Assumption 4(ii). Now Theorem 2 (2) becomes applicable and $||\widehat{h}_{nl} - h_{ol}||_{Y_2} = O_p([\log n]^{-r})$ provided $k_n = O(\log n)$. This logarithmic rate is shown to be

optimal in Efromovich and Koltchinskii (2001, corollary 6.3 and example 3.2 with their ε, α and β being our $n^{-1/2}$, r and 1), in the context of an ill-posed inverse white-noise model with an unknown operator that can be estimated by a training sample.

(ii) According to Theorem 2, in order for the sieve nonparametric IV estimator to perform well for a given sample size n, we should choose the sieve space $\mathcal{H}_n$ to best approximate the structural function h_o, and at the same time to have the best order of sieve measure of ill-posedness τ_n. Lemma 1 suggests that the sieve nonparametric IV estimator could reach optimal convergence rate when $\mathcal{H}_n$ coincides with the linear subspace generated by the eigenfunctions $\{\phi_{1k} : k = 1, ..., k_n\}$. Nevertheless, it does not rule out the existence of other sieve spaces that possess the best approximation error rate k_n^{-r} and the best order of τ_n. In fact, for an ill-posed inverse white-noise model with a known operator T, Donoho (1995) establishes the optimal convergence rate for a sieve estimator using a wavelet sieve $\mathcal{H}_n$ that best approximates the function h_o, but differs from the linear subspace generated by the eigenfunctions $\{\phi_{1k} : k = 1, ..., k_n\}$. See Cohen, Hoffmann and Reiss (2004) for similar results. In subsection 6.2 we present a Monte Carlo study to evaluate the finite sample performance of the sieve IV estimator when (X_2, Y_2) is drawn from a bivariate Gaussian density. We find that the sieve IV estimator using cardinal B-splines for both $h_o(Y_2)$ and $E[Y_1 - h_o(Y_2)|X_2]$ performs well and is comparable to the one using the ideal Hermite polynomial (eigenfunction) sieves.

Even if one knows the conditional expectation operator T, one might still not know its corresponding singular value system $\{\mu_k; \phi_{1k}, \phi_{0k}\}_{k=1}^{\infty}$ explicitly. Nevertheless, we can still bound τ_n using the smoothness of the operator T^*T and the property of the sieve space $\mathcal{H}_n$. We provide such examples in the next two theorems. Let $\widehat{h}$ denote the Fourier transform of h (i.e., $\widehat{h}(\xi) = \frac{1}{\sqrt{2\pi}} \int \exp\{-iy\xi\} h(y) dy$), and $(g)^\vee$ the inverse Fourier transform of g (i.e., $(g)^\vee(y) = \frac{1}{\sqrt{2\pi}} \int \exp\{iy\xi\} g(\xi) d\xi$). For any $s \in \mathcal{R}$, a fractional Sobolev space $W_2^s(\mathcal{R})$ is the space of functions h in $L_2(\mathcal{R})$ such that $(1+|\cdot|^2)^{s/2}\widehat{h}(\cdot) \in L_2(\mathcal{R})$; i.e., $W_2^s(\mathcal{R}) = \{h \in L_2(\mathcal{R}) : ||h||_{W_2^s(\mathcal{R})} \equiv \left\| \left((1 + |\cdot|^2)^{s/2} \widehat{h}(\cdot) \right)^\vee \right\|_{L_2} < \infty\}$. This definition is equivalent to that of (8) when $s > 0$. Note that for $s > 0$, the norm $||h||_{W_2^{-s}(\mathcal{R})}$ is a shrinkage in the Fourier domain. It is known that $W_2^{-s}(\mathcal{R})$ is the dual space of $W_2^s(\mathcal{R})$. See Triebel (1983) for various equivalent definitions of $W_2^s(\mathcal{R})$ and the corresponding space $W_2^s(\mathcal{Y})$ for $\mathcal{Y}$ being an interval of $\mathcal{R}$.

THEOREM 3: *Let Assumptions 1(iii) and 4(iii) hold. Suppose that f_{0,Y_2} is bounded above and bounded away from zero over its support $\mathcal{Y}$, which is a bounded interval of $\mathcal{R}$ with non-empty interior. Suppose that there is a finite constant $s > 0$ such that $||(T^*T)^{1/2}h||_{Y_2} \asymp ||h||_{W_2^{-s}(\mathcal{Y})}$ for all $h \in L_2(\mathcal{Y}, f_{0,Y_2})$. Then: (1) $\mu_k \asymp (k)^{-s}$;*

(2) Let the sieve space $\mathcal{H}_n$ be either cosine sieve, or spline sieve of order $\gamma > s$, or wavelet sieve

of order $\gamma > s$. Then: (i) $\tau_n \leq const.(k_n)^s$; (ii) Assumption 6 is satisfied; (iii) If Assumptions 1, 2(i)(ii), 3, 4 and 5(i)(ii) hold with $k_n = O(n^{\frac{1}{2(r+s)+1}})$, then $||\widehat{h}_{nl} - h_{ol}||_{Y_2} = O_p\left(n^{-\frac{r}{2(r+s)+1}}\right)$.

We note that the condition "$||(T^*T)^{1/2}h||_{Y_2} \asymp ||h||_{W_2^{-s}(\mathcal{Y})}$ for all $h \in L_2(\mathcal{Y}, f_{0,Y_2})$" means that the operator $(T^*T)^{1/2}$ maps $L_2(\mathcal{Y}, f_{0,Y_2})$ isomorphically onto $W_2^s(\mathcal{Y})$. This condition implies an eigenvalue decay order $\mu_k \asymp (k)^{-s}$, but it does not provide expressions of the eigenfunctions of $(T^*T)^{1/2}$. Nevertheless, according to part (2) of Theorem 3, several commonly used sieve spaces $\mathcal{H}_n$ that have the best approximation error rate k_n^{-r} also have the best order of sieve measure of ill-posedness τ_n. The convergence rate of Theorem 3 achieves the minimax optimal rate obtained in Chen and Reiss (2007) for the ideal case of a nonparametric instrumental regression with a known conditional expectation operator T.

Recently, Hall and Horowitz (2005) and Darolles, Florens and Renault (2006) propose alternative nonparametric IV estimators and obtain convergence rates of their estimators. Due to the different estimation procedures and different assumptions, it is beyond the scope of this paper to clarify the exact relations between their convergence rates and ours as stated in Theorem 3. Nevertheless, we point out that Hall and Horowitz (2005) establish rate optimality of their estimator without assuming the existence of any derivatives of h_o, instead they impose a condition on h_o in terms of the decay speed of its Fourier coefficients relative to the eigenfunction basis $\{\phi_{1k} : k \geq 1\}$ of the operator T^*T. We assume $h_o \in \Lambda_c^r(\mathcal{Y})$ for some $r > 1/2$ (see Assumption 2(i)), which requires the existence of $[r]-$th derivative, $\nabla^{[r]}h_o$, where $[r]$ is the largest nonnegative integer such that $[r] < r$. Moreover, as a sufficient condition to obtain root-n consistent estimation of the equivalence scale parameter, θ_{o1}, in condition (3) of Theorem 1 we assume the existence of the first derivative of h_o (also see Section 5). Therefore, our estimation of shape-invariant Engel curves imposes derivative smoothness condition on h_o that is not assumed in Hall and Horowitz (2005).

Theorem 3 assumes that the support of f_{0,Y_2} is a bounded interval of $\mathcal{R}$. The following result relaxes this assumption.

THEOREM 4: *Suppose that Assumptions 1, 2(i)(ii), 3, 4 and 5(i)(ii) hold.* Let $w_a(y) \equiv (1+y^2)^{-a}$ for *some $a > 1/2$. Suppose that $f_{0,Y_2}(y) \asymp w_a(y)$ for large $|y|$. Suppose that there are finite constants c, $s > 0$ such that $||(T^*T)^{1/2}h||_{Y_2} \geq c||h||_{W_2^{-s}(\mathcal{R},w_a)}$ for all $h \in L_2(\mathcal{R}, f_{0,Y_2})$. Let the sieve space $\mathcal{H}_n$ be given in (9)-(11) with $\gamma > s$. Then: (1) $\tau_n \leq const.(k_n)^s$; (2) If Assumption 6 is satisfied, then $||\widehat{h}_{nl} - h_{ol}||_{Y_2} = O_p\left(n^{-\frac{r}{2(r+s)+1}}\right)$ provided $k_n = O\left(n^{\frac{1}{2(r+s)+1}}\right)$.*

The relation between the smoothness of T^*T and singular values μ_k becomes more complicated when f_{0,Y_2} has unbounded support $\mathcal{R}$. See Chen, Hansen and Scheinkman (2005) for some results.

5. ASYMPTOTIC NORMALITY AND EFFICIENCY OF $\hat{\theta}$

For the system of shape-invariant Engel curve with exogenous total expenditure model, Blundell, Browning and Crawford (2003) have established $\sqrt{n}$-asymptotic normality of their kernel LS estimator of θ_o. However, there is no published results on whether θ_o could be estimated at $\sqrt{n}$-rate for the endogenous total expenditure case, nor on how to obtain efficient estimation of θ_o under heteroskedasticity of unknown form. In this section, for the system of shape-invariant Engel curve with possibly endogenous total expenditure model, we provide relatively low-level sufficient conditions for $\hat{\theta}$ to be $\sqrt{n}$-asymptotically normally distributed and for the three-step estimator to be semiparametrically efficient. The following Propositions 1 - 3 can be obtained by applying the general theory of Ai and Chen (2003). We refer readers to our working paper version, Blundell, Chen and Kristensen (2003), for the proofs.

5.1. *Asymptotic Normality and Efficiency*

We first impose the following conditions:

ASSUMPTION 2: (iii) conditions (3) - (5) of Theorem 1 hold; (iv) $\theta_o = (\theta'_{o1}, \theta'_{o2,1}, ..., \theta'_{o2,L})' \in int(\Theta)$.

ASSUMPTION 5: (iii) $J_n \geq (1 + L^{-1}) \dim(X_1) + k_n$; (iv) $k_n^2 \ln(n)/\sqrt{n} \to 0$, $k_n^{-r_m} = o\left(n^{-1/4}\right)$.

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; (iii) $\Sigma_o(X) = Var\left[\rho(Z, \alpha_o)|X\right]$ is finite positive definite.

PROPOSITION 1: *Suppose Assumptions 1-7 and B1-B2 stated in the Appendix B are satisfied. Let $\widehat{\alpha}$ be the SMD estimator (5) with the sieve space given by (9). Then $\sqrt{n}(\widehat{\theta}_n - \theta_o) \Longrightarrow \mathcal{N}(0, V^{-1})$, where V is given in (28) in the Appendix B.*

Before we state the semiparametric efficiency of the three-step estimator $\widetilde{\theta}$ obtained in (6), we need some additional notation. We define $D_{w^1}(X, \alpha_o)$ and $D_{w^{2,l}}(X, \alpha_o)$ as the $L \times \dim(X_1)$-matrix valued functions given by

$$D_{w^1}(X, \alpha_o) = E\left[\nabla h_o(Y_2 - \phi(X_1'\theta_{o1}))\nabla\phi(X_1'\theta_{o1})X_1' + w^1(Y_2 - \phi(X_1'\theta_{o1})) \mid X\right] \tag{18}$$

with

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

and for $l = 1, ..., L$,

$$D_{w^{2,l}}(X, \alpha_o) = E\left[-e_l X_1' + w^{2,l}(Y_2 - \phi(X_1'\theta_{o1})) \mid X\right], \tag{19}$$

where e_l denotes the $L \times 1$-vector with 0's everywhere except 1 in the l-th element; and $w^1(Y_2 - \phi(X_1'\theta_{o1}))$, $w^{2,l}(Y_2 - \phi(X_1'\theta_{o1}))$ are $L \times \dim(X_1)$-matrix valued squared integrable functions of $Y_2 - \phi(X_1'\theta_{o1})$. Let $D_{w_k^j}(X, \alpha_o)$ be the k-th column of $D_{w^j}(X, \alpha_o)$. Let $w^{oj}(Y_2 - \phi(X_1'\theta_{o1})) = (w_1^{oj}(Y_2 - \phi(X_1'\theta_{o1})), ..., w_{\dim(X_1)}^{oj}(Y_2 - \phi(X_1'\theta_{o1})))$ be given by

$$w_k^{oj} = \arg \inf_{\substack{w_k^j \in \mathcal{H} - \{h_o\} \\ w_k^j \neq 0}} E\left[D_{w_k^j}(X, \alpha_o)' \Sigma_o(X)^{-1} D_{w_k^j}(X, \alpha_o)\right] \text{ for } k = 1, ..., \dim(X_1).$$

Finally we denote

$$D_{w^o}(X, \alpha_o) = (D_{w^{o1}}(X, \alpha_o), D_{w^{o2,1}}(X, \alpha_o), ..., D_{w^{o2,L}}(X, \alpha_o)) \tag{20}$$

as the $L \times (1 + L)\dim(X_1)$-matrix valued function, and $w^o = (w^{o1}, w^{o2,1}, ..., w^{o2,L})$.

PROPOSITION 2: *Let $\tilde{\alpha}_n = (\tilde{\theta}_n, \tilde{h}_n)$ be the three-step estimator (6). Suppose all conditions of Proposition 1 are satisfied with $\Sigma = \Sigma_o$. Then $\tilde{\theta}_n$ is semiparametrically efficient, and $\sqrt{n}(\tilde{\theta}_n - \theta_o) \Longrightarrow \mathcal{N}(0, V_o^{-1})$, where $V_o = E\left[D_{w^o}(X, \alpha_o)'[\Sigma_o(X)]^{-1} D_{w^o}(X, \alpha_o)\right]$.*

5.2. *Consistent Covariance Estimator*

To conduct any statistical inference using the semiparametric efficient estimator $\widetilde{\theta}$, we need a consistent estimator $\widetilde{V}_o$ of V_o. Let $\widehat{D}_{w^j}(X, \widetilde{\alpha})$ be a consistent estimator of $D_{w^j}(X, \alpha_o)$. For example we can use the sieve LS estimator

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

where for $j = 1$,

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

with

$$\nabla\widetilde{h}(Y_{2i}-\phi(X_{1i}'\widetilde{\theta}_1))=\left[\begin{array}{l}\nabla\widetilde{h}_1(Y_{2i}-\phi(X_{1i}'\widetilde{\theta}_1))\\ ...\\ \nabla\widetilde{h}_L(Y_{2i}-\phi(X_{1i}'\widetilde{\theta}_1))\end{array}\right]\ L\times 1 \text{ vector},$$

and for $l=1,...,L$,

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

Let $\widehat{D}_{w_k^j}(X,\widetilde{\alpha})$ denote the k-th column of $\widehat{D}_{w^j}(X,\widetilde{\alpha})$, and $\widetilde{w}^j=(\widetilde{w}_1^j,...,\widetilde{w}_{\dim(X_1)}^j)$ the solution to

$$\inf_{w_k^j\in\mathcal{H}_n:w_k^j\neq 0}\frac{1}{n}\sum_{i=1}^n\widehat{D}_{w_k^j}(X_i,\widetilde{\alpha})'\widehat{\Sigma}_o(X_i)^{-1}\widehat{D}_{w_k^j}(X_i,\widetilde{\alpha})\quad\text{for } k=1,...,\dim(X_1).$$

Then $\widehat{D}_{\widetilde{w}}(X,\widetilde{\alpha})=(\widehat{D}_{\widetilde{w}^1}(X,\widetilde{\alpha}),\widehat{D}_{\widetilde{w}^{21}}(X,\widetilde{\alpha}),...,\widehat{D}_{\widetilde{w}^{2L}}(X,\widetilde{\alpha}))$ is a consistent estimator of $D_{w^o}(X,\alpha_o)$ given in (20). Finally Let $\widetilde{V}_o=\frac{1}{n}\sum_{i=1}^n\widehat{D}_{\widetilde{w}}(X_i,\widetilde{\alpha})'\widehat{\Sigma}_o(X_i)^{-1}\widehat{D}_{\widetilde{w}}(X_i,\widetilde{\alpha})$.

PROPOSITION 3: *Under the conditions of Proposition 2,* $\tilde{V}_o=V_o+o_p(1)$.

6. IMPLEMENTATION AND SIMULATIONS

6.1. *Implementation*

The SMD estimation of the system of shape-invariant Engel curves can be easily implemented using a matrix-oriented software package such as Gauss or Matlab. Here we only discuss the implementation for the endogenous case; the exogenous case follows along the same lines.

Since the $l-$th unknown Engel curve $h_l(\cdot)$ only enters the $l-$th conditional moment function $m_l(x,\theta,h_l)\equiv E\{Y_{1l}-h_l(Y_2-\phi(X_1'\theta_1))-X_1'\theta_{2,l}|X=x\}$ linearly, STEP 1 of the SMD procedure can be easily implemented using the profile approach:

STEP 1A: for each fixed $\theta\in\Theta$, compute $\widehat{h}_{l,n}(\theta;\cdot)=\arg\min_{h_l\in\mathcal{H}_n}\sum_{i=1}^n[\widehat{m}_l(X_i,\theta,h_l)]^2$ for $l=1,...,L$;

STEP 1B: compute $\widehat{\theta}_n=\arg\min_{\theta\in\Theta}\sum_{l=1}^L\sum_{i=1}^n[\widehat{m}_l(X_i,\theta,\widehat{h}_{l,n}(\theta;\cdot))]^2$, and estimate $h_{ol}(\cdot)$ by $\widehat{h}_{l,n}=\widehat{h}_{l,n}(\widehat{\theta}_n;\cdot)$ for $l=1,...,L$.

Recall that any candidate Engel curve function $h_{l,n}(\cdot)=\psi^{k_n}(\cdot)'\Pi^l\in\mathcal{H}_n$, $l=1,...,L$, is subject to two constraints: (i) $0\leq h_{l,n}<1$ and (ii) $\sup_y|\nabla^{[r]}h_{l,n}(y)|\leq c$ for $\mathcal{H}_n$ in (9) or $||\nabla^{[r]}h_{l,n}||_{L_2}\leq c$ for $\mathcal{H}_n$

in (10). In simulations and empirical applications, we have imposed one weaker but easier-to-compute constraint: $\frac{1}{n}\sum_{i=1}^n |h_{l,n}(Y_{2i})|^2 + \int \left|\nabla^{[r]} h_{l,n}(y)\right|^2 dy \leq D_n$, where D_n could grow slowly with sample size n (say $D_n = \log n$ or $\log\log n$). Let $C_0 = \frac{1}{n}\sum_{i=1}^n \psi^{k_n}(Y_{2i} - \phi(X_{1i}'\theta_1))\psi^{k_n}(Y_{2i} - \phi(X_{1i}'\theta_1))'$ and $C_{[r]} = \int\{\nabla^{[r]}\psi^{k_n}(y)\nabla^{[r]}\psi^{k_n}(y)'\}dy$ (for spline basis, $C_{[r]}$ can also be the self-adjoint difference Gram matrix; see, e.g., Schumaker, 1993). Let $C = C_0 + C_{[r]}$ be $k_n \times k_n-$matrices. Then the constraint becomes $(\Pi^l)'C\Pi^l \leq D_n$ for a known bound D_n. If we use the sieve LS estimator $\widehat{m}_l$ for m_l, then the STEP 1A procedure becomes: for $l = 1, ..., L$, compute $\widehat{h}_{l,n}(\theta;.) = \psi^{k_n}(.)'\widehat{\Pi}_D^l(\theta)$, where $\widehat{\Pi}_D^l(\theta)$ solves

$$\min_{\Pi^l,\lambda} \left(\mathcal{Y}_{1l}(\theta_{2,l}) - \Psi(\theta_1)\Pi^l\right)' P(P'P)^- P' \left(\mathcal{Y}_{1l}(\theta_{2,l}) - \Psi(\theta_1)\Pi^l\right) + \lambda\{(\Pi^l)'C\Pi^l - D_n\} \tag{21}$$

with $\mathcal{Y}_{1l}(\theta_{2,l}) = (Y_{11l} - X_{11}'\theta_{2,l}, ..., Y_{1nl} - X_{1n}'\theta_{2,l})'$, $\Psi(\theta_1) = (\psi^{k_n}(Y_{21} - \phi(X_{11}'\theta_1)), ..., \psi^{k_n}(Y_{2n} - \phi(X_{1n}'\theta_1)))'$, and $\lambda \geq 0$ being the Lagrange multiplier. The problem (21) has a simple closed-form solution:

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

where $\widehat{\lambda}$ satisfies $\{\widehat{\Pi}_D^l(\theta)\}'C\{\widehat{\Pi}_D^l(\theta)\} = D_n$.

In practice the bound D_n might be unknown. In our simulation studies and empirical applications we actually solved the following equivalent problem to (21) for a few possible values of $\lambda \in [0, 1]$:

$$\min_{\Pi^l} \left(\mathcal{Y}_{1l}(\theta_{2,l}) - \Psi(\theta_1)\Pi^l\right)' P(P'P)^- P' \left(\mathcal{Y}_{1l}(\theta_{2,l}) - \Psi(\theta_1)\Pi^l\right) + \lambda\Pi^{l\prime}C\Pi^l, \tag{22}$$

and the solution is: $\widehat{\Pi}_\lambda^l(\theta) = (\Psi(\theta_1)'P(P'P)^-P'\Psi(\theta_1) + \lambda C)^- \Psi(\theta_1)'P(P'P)^-P'\mathcal{Y}_{1l}(\theta_{2,l})$. We note that when $\lambda = 0$ (i.e. without smoothness constraints), this solution is simply the well known IV/2SLS estimator.

Let $\widehat{\Pi}^l(\theta)$ denote either the solution $\widehat{\Pi}_D^l(\theta)$ to the problem (21) or the solution $\widehat{\Pi}_\lambda^l(\theta)$ to the problem (22). Then $\widehat{h}_{l,n}(\theta;.) = \psi^{k_n}(.)'\widehat{\Pi}^l(\theta)$ is the profile SMD estimator in STEP 1A. Next, we plug $(\widehat{h}_{1,n}(\theta;.), ..., \widehat{h}_{L,n}(\theta;.))'$ into the STEP 1B problem:

$$\min_{\theta\in\Theta} \sum_{l=1}^L \left(\mathcal{Y}_{1l}(\theta_{2,l}) - \Psi(\theta_1)\widehat{\Pi}^l(\theta)\right)' P(P'P)^- P' \left(\mathcal{Y}_{1l}(\theta_{2,l}) - \Psi(\theta_1)\widehat{\Pi}^l(\theta)\right). \tag{23}$$

The solution $\widehat{\theta}$ to the problem (23) will be a $\sqrt{n}$-consistent estimator for θ_o, and the corresponding SMD estimator for h_{ol} is $\widehat{h}_{l,n}(\widehat{\theta};.) = \psi^{k_n}(.)'\widehat{\Pi}^l(\widehat{\theta})$, $l = 1, ..., L$. To solve the problem (23), one needs to run a numerical routine since θ enters nonlinearly, but it is relatively easily performed compared

to optimizing over both Π and θ simultaneously.

In STEP 2, we use the above sieve profile estimator $(\widehat{\Pi}, \widehat{\theta})$ to compute a consistent estimator $\widehat{\Sigma}_o(X)$ for $\Sigma_o(X)$. In STEP 3, we could use the above sieve profile estimator $(\widehat{\Pi}, \widehat{\theta})$ as a starting point to solve the optimally weighted SMD problem (6) simultaneously over (Π, θ). Actually this optimization can again be solved easily using the profile approach.

In the actual implementation of the above procedures, one has to specify sieve bases $(p^{J_n}(X), \psi^{k_n}(Y_2))$, and smoothing parameters $(J_n,\ k_n)$. The theoretical results obtained in Sections 4 and 5 provide some guidelines about such choices for the case of known bound D_n. In particular, one should choose $J_n \geq k_n$ to ensure identification and to satisfy Assumption 5, 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" $\psi^{k_n}(Y_2)$. In the simulations and empirical applications we find that $J_n = c_0 k_n$, $c_0 \approx 2$ or 3 works fine. We also consider the case of unknown bound D_n; hence, we have to select λ in addition to $(J_n,\ k_n)$. There is a certain interdependence between k_n and λ; a high number k_n could potentially lead to overfitting (i.e. the estimated Engel curve h becomes wiggly and the variance gets big), but this can be controlled for with a slightly bigger penalization weight λ. For a given choice of k_n, one may try out different values of $\lambda \in [0, 1]$ 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 LS regression setting. However, there is no theoretical justification for such a procedure in the endogenous case. In fact, even for the sieve nonparametric LS regressions, how to optimally choose k_n and λ simultaneously is still an open question. Finally, for the semiparametric efficient estimation of θ, one can choose slightly bigger J_n, k_n (or smaller λ) than those used for the purely nonparametric IV estimation of h.

6.2. *Two Monte Carlo Studies*

Before applying the SMD estimators of the shape-invariant Engel curves to the British FES data set, we assess the performance of the purely nonparametric IV estimator in two small simulation studies. We are particularly 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 choices 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 $J_n = 2J_{2n}$, k_n and λ will lead to similar estimated functions that are all consistent estimates of the true unknown function.

6.2.1. *Simulation 1:* (Y_2, X_2) *is drawn from the data*

The first Monte Carlo design will mimic the specific FES data set; see the empirical section for a detailed description of the data. All we need to know here is that the data set consists of two subsamples of households: one consisting of families with no children and one of families with 1-2 children such that $X_1 \in \{0, 1\}$ and $J_n = 2J_{2n}$. In this simulation study, we 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 drawn in our simulation study. The model we simulate is given by

$$Y_1 = h_o(Y_2) + \varepsilon, \quad \varepsilon = E[h_o(Y_2)|X_2] - h_o(Y_2) + v, \tag{24}$$

where $v \sim \mathcal{N}(0, 0.01)$ 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 (24) for two choices of h_o, one is linear and the other is nonlinear:

$$\text{(i) } h_o(y_2) = -0.1095y_2 + 0.7229, \quad \text{(ii) } h_o(y_2) = \Phi((y_2 - 5.5)/0.3),$$

where Φ 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 $y_2 = \max_i \{Y_{2i}\}$. Model (i) closely mimics the estimate obtained for food-in expenditure in the empirical application.

The sieve nonparametric IV estimation of $h_o(y_2)$ is very simple. First, we approximate $h_o(Y_2)$ by $h_n(Y_2) = \psi^{k_n}(Y_2)'\Pi$ and $m(X_2) = E[Y_1 - h(Y_2)|X_2]$ by $m_n(X_2) = B^{J_{2n}}(X_2)'A$. We then obtain an estimator of Π by solving

$$\min_{\Pi} [\mathcal{Y}_1 - \Psi\Pi]'\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'[\mathcal{Y}_1 - \Psi\Pi] + \lambda(\Pi)'C(\Pi),$$

where $\mathcal{Y}_1 = (Y_{11}, ..., Y_{1n})'$, $\Psi = (\psi^{k_n}(Y_{21}), ..., \psi^{k_n}(Y_{2n}))'$, $\mathbf{B} = (B^{J_{2n}}(X_{21}), ..., B^{J_{2n}}(X_{2n}))'$, and the smoothness penalization matrix $C = C_0 + C_2$ with C_0, C_2 defined in Subsection 6.1. This problem has the solution $\widehat{\Pi}_\lambda = (\Psi'\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\Psi + \lambda C)^- \Psi'\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathcal{Y}_1$, such that $\widehat{h}(y_2) = \psi^{k_n}(y_2)'\widehat{\Pi}_\lambda$.

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 $B^{J_{2n}}(X_2)$ and $\psi^{k_n}(Y_2)$ for the conditional mean m and h respectively, all yielding similar results as long as the sieve approximating terms J_{2n} and k_n and the penalization weight λ are similar. We here report the results for a few combinations: h is approximated by either a 3nd 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 cosine basis with $J_{2n} \approx 3k_n$,27, or a 4th order B-spline with $J_{2n} = 15, 25$. To check for the robustness of the sieve IV estimators towards the choice of smoothness penalization, we also present the results for different penalization weights $\lambda = 0.8$, 0.4, 0.1, 0.01, 0.001, 0.0. Table 4 reports the integrated squared bias, integrated variance and integrated MSE based on the 100 simulations for the sieve IV estimators of nonlinear h,[6] where h was estimated using either polynomial spline or B-splines with different smoothing parameters k_n, J_{2n} and λ. 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 not very sensitive towards the choice of the penalization weight λ; however the variance increases as λ decreases; and $\lambda = 0.8$ yields the best performance in terms of the integrated MSE for large $k_n \geq 8$. We obtained similar results for sieve IV estimators of the linear h. The sieve IV estimates for the nonlinear and the linear h are plotted in Figures 1 and 2 with a $k_n = 9$-dimensional B-spline for h, $J_{2n} = 25$-dimensional B-spline for m, $\lambda = 0.8$ and 0.0. Corresponding tables and plots for other choices of k_n, J_{2n} and other bases functions for h and m were very similar, and the full set of results can be obtained from the authors upon request.

From these results it is also 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 λ, which again indicates that for a given value k_n of sieve terms in approximating h, the penalization weight λ 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 (2005) test for weak instruments in the parametric linear 2SLS regression problem, suggests the presence of weak instruments under the specification $\lambda = 0.0$, $J_{2n} = 15$ 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 1 and 2 show no indication of any bias towards LS. This is also confirmed by Table 4 where with $\lambda = 0.0$, the bias of sieve IV generally decreases as J_{2n} increases from 15 to 25. Finally Table 5 reports the integrated squared bias, integrated variance and integrated MSE of the sieve IV and the inconsistent sieve LS estimators with $k_n = 6$, 9. This shows that the sieve LS estimator is not sensitive toward the choice of λ, consistently having a small variance but a large bias compared to the sieve IV estimator leading to its MSE being greater than the one of the IV estimator.

The results of the first Monte Carlo study can be summarized as follows: (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 $J_{2n} = c_0 k_n$ 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 λ 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_{2n}, k_n, λ 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 k_n, λ 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 $J_{2n} = c_0 k_n$ 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 $k_n = 4, 5, 6$, $J_{2n} \approx 3k_n$ and $\lambda \in [0.0, 0.001]$ will satisfy Assumption 5 (with $r = 2$ say) for $\sqrt{n}-$normality and efficiency of θ 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 given the Monte Carlo design since the sieve LS estimator of the linear and the nonlinear h is 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.

We want to make sure the simulated data sets do mimic the real FES data set with no children subsample such that the insights we learnt about the smoothing parameters k_n, J_{2n}, λ 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, we know one important ingredient for nonparametric IV regression is the decay behavior of the singular values $\{\mu_k\}$ associated with the conditional expectations operator $T(h)(\cdot) = E[h(Y_2)|X_2 = \cdot]$. Therefore, we estimated the singular values $\{\mu_k\}$ associated with T. We restricted h to $h(Y_2) = \psi^{k_n}(Y_2)'\Pi$ and imposed smoothness constraints on it via the aforementioned penalization matrix C and Lagrange multiplier λ, while the operator T was approximated using $B^{J_{2n}}(X_2)$. Then for each simulated data set, we estimated $1 = \mu_1^2 \geq \mu_2^2 \geq ... \geq \mu_{k_n}^2$ by the solutions to the eigenvalue problem

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

where Ψ and $\mathbf{B}$ are defined earlier, $\psi^{k_n}(Y_2)$ is a 3rd order B-spline basis with $k_n = 14$, and $B^{J_{2n}}(X_2)$ is a 4th order B-spline basis with $J_{2n} = 25$. We repeated this 100 times. We found that the eigenvalue estimates using simulated data match well with the ones estimated using the real FES data of no-kids subsample. See Blundell, Chen and Kristensen (2003) for these estimates.

Another important point in relation to the weak IV issue is that the singular values for the pooled FES data decay less rapidly than the ones for the no-kids subsample. Consequently, we expect our estimator to be better behaved in the pooled sample.

6.2.2. *Simulation 2:* (Y_2, X_2) *is drawn from a bivariate Gaussian density*

From Remark 4 in Section 4, we know that the theoretical convergence rate of the sieve IV estimator will be very slow (logarithmic) if the joint density of log-total expenditure and log-earnings were in fact bivariate Gaussian. We now investigate the finite sample performance of the sieve IV estimator in this severely ill-posed inverse setting. We maintain the Monte Carlo design from above, only now we draw (Y_2, X_2) from a bivariate Gaussian density instead of from the kernel density estimate. The mean and covariance matrix of the Gaussian density is chosen as the empirical counterparts of the no-kids subsample. We consider four different cases for both linear and nonlinear h: *(1a)* use X_2 as an instrument and B-splines for both $h(Y_2)$ and $m(X_2) = E[Y_1 - h(Y_2)|X_2]$; *(1b)*

use X_2 as an instrument and Hermite polynomials for both $h(Y_2)$ and $m(X_2)$; *(2a)* use $\tilde{X}_2 = \Phi(X_2)$ as an instrument and B-splines for both $h(Y_2)$ and $m(\widetilde{X}_2) = E[Y_1 - h(Y_2)|\widetilde{X}_2]$; *(2b)* use $\tilde{X}_2 = \Phi(X_2)$ as an instrument, Hermite polynomials for $h(Y_2)$ and cosine basis for $m(\widetilde{X}_2)$. For each case, we have tried different combinations of smoothing parameters: $k_n = 6, 9$, $J_{2n} = 23$ and $\lambda = 0.8$, 0.4, 0.1, 0.01, 0.001, 0.0. See Tables C.3 and C.4 and Figure D.3 for some of the results with nonlinear h.

Our findings for both linear and nonlinear h are as follows: (1) Regardless the choices of instrument (X_2 or $\tilde{X}_2$) and the smoothing parameters (λ, k_n), the sieve IV estimates of h using cardinal B-spline bases for both h and m perform very well. (2) For $\lambda = 0.0$, the sieve IV estimates of h under cases (1a), (2a) and (2b) all perform well. However, the sieve IV estimate using the ideal Hermite polynomial sieves for both h and m performs poorly due to big variance. (3) It has slight advantage to use $\tilde{X}_2 = \Phi(X_2) \in [0, 1]$ as an instrument, especially so if one uses Hermite polynomial sieve for h. (4) For $\lambda \in [0.001, 0.4]$, the sieve IV estimates of h under all four cases perform well, that is, the choices of sieve bases are not very important as long as $\lambda \in [0.001, 0.4]$. (5) For all four cases with fixed $k_n = 6, 9$, $J_{2n} = 23$, as λ increases from 0.0 to 0.8, the variances of sieve IV estimates reduce fast, but the squared biases of sieve IV estimates increase a little bit. (6) There are many combinations of sieve bases and smoothing parameters J_{2n}, k_n, λ which can reduce the variance part and lead to a small integrated MSE of the consistent sieve IV estimator. However, there is no combination of sieve bases and smoothing parameters k_n, λ which can reduce the large bias of the inconsistent sieve LS estimator.

We already demonstrated that the sieve IV estimation is very easy to implement, the Monte Carlo results now indicate that the sieve IV estimators perform well even in the severely ill-posed inverse setting.

7. AN EMPIRICAL INVESTIGATION

In this section we apply the results developed in the previous sections to the problem of estimating a system of shape-invariant Engel curves. In this application total expenditure is allowed to be endogenous and gross earnings of the head of household is used as an instrument. We begin our analysis with a data description and examination of the quality of the gross earnings as an instrument. We then present semi-nonparametric estimates of Engel curves under the two alternative assumptions: (i) Y_2 is exogenous and (ii) Y_2 is endogenous. We assess the importance of allowing for endogeneity both on the nonparametric shape of the Engel curve and on the parametric components that represent the demographic parameters.

7.1. *The Engel Curve Data*

In our application we consider $L = 7$ broad categories of non-durables and services: food-out, food-in, alcohol, fuel, leisure goods and services, fares (including expenses on public transports), and travel (excluding expenses on public transports). The data set is drawn from the British FES. We have applied the SMD estimator on data from the survey for three different years and the qualitative conclusions are the same. But for the purposes of this discussion we select a single year (1995). 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 in work. We exclude all those households with three or more children. So our demographic variable, X_1, is simply 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 $\phi(X_{1i}'\theta_1)$ as $X_{1i}\theta_1$. The log of total expenditure on non-durables and services is our measure of the continuous endogenous explanatory variable 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 1 presents descriptive statistics for the main variables used in the empirical analysis. It shows the smaller share of food-out, alcohol, fares, 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 θ in our semiparametric model to reflect these.

Figures 4-5 present plots of the kernel estimates for the joint density of log-total expenditure and log-earnings; see Härdle and Linton (1994) for a review of the kernel method. On each graph there is also a series estimate of $E[Y_2|$ log-earnings$]$. 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. We see that the joint density is also smooth and, together with the conditional mean, confirms our belief that the gross earnings variable should be a good choice for our instrumental variable. Since the kernel estimate of the density of log-earnings is close to normal, we have taken the instrumental variable $X_2 = \Phi(\text{log-earnings})$ in the empirical applications.

7.2. *Quality of the Instrument*

Our sieve IV estimator is similar to the parametric linear IV regression once the basis functions $p^{J_n}(X)$ and $\psi^{k_n}(Y_2)$, and the smoothing parameters J_n, k_n and λ are chosen. As a consequence, we

Table 1: Data Descriptives

	couples w/o children		couples w/ children	
	mean	std.	mean	std.
budget shares:				
food-in	0.1776	0.0950	0.2256	0.0938
food-out	0.0829	0.0591	0.0790	0.0555
alcohol	0.0712	0.0791	0.0496	0.0543
fares	0.0216	0.0499	0.0137	0.0399
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
expenditure and income:				
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	

consider the following regression,

$$\psi^{k_n-1}(Y_2) = Ap^{J_n}(X) + e, \quad E[e|X] = 0, \tag{25}$$

where $\psi^{k_n-1}(Y_2)$ is a vector of endogenous regressors of dimension $k_n - 1$ (excluding the constant one from the original sieve basis $\psi^{k_n}(Y_2)$) and $p^{J_n}(X)$ is a vector of instrumental variables with dimension J_n. Here we choose $\psi^{k_n}(Y_2) = B_3^{k_n}(Y_2)$ as a 3rd order B-spline basis with $k_n = 9$ (and $\lambda = 0.4$), and let $p^{J_n}(X)' = [B_4^{J_{2n}}(X_2)', X_1 B_4^{J_{2n}}(X_2)']$ where $B_4^{J_{2n}}(X_2)$ is a 4th order B-spline basis for functions of X_2 with the number of sieve terms $J_{2n} = 15$, hence the length of the vector $p^{J_n}(X)$ is actually $J_n = 2J_{2n} = 30$ for the full sample including both types of households. We then perform two tests for the quality of instruments.

In the first we test the hypothesis $H_0 : \text{rank}(A) = r^*$ by applying the result in Robin and Smith (2000). As reported in Blundell, Chen and Kristensen (2003) we easily reject H_0 for $r^* = 0, ..., 8$. Note that the Robin-Smith (2000) test is based on the assumptions that the model (25) is correctly specified with both k_n and J_n being fixed and finite known numbers, and that A is estimated root-n consistently with asymptotically normal distribution. However, our basic setup of unknown $h()$ implies that the model (25) 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 $E[h_o(Y_2)|X] = 0$. Second, we perform the Stock and Yogo's (2005)

test of the null hypothesis $H_0 : B^{J_{2n}}(X_2)$ are weak instruments for $\psi^{k_n-1}(Y_2)$. This is done with $\psi^{k_n-1}(Y_2) = (Y_2, Y_2^2, Y_2^3)$. The test statistic is 4.5647 for households without children, and 10.9535 for those with children. With number of instruments $J_{2n} = 15$, the 5% critical values are given by 10.33 and 4.37 (for a 10% and 30 % maximal bias relative to OLS respectively). When pooling the two household groups with number of instruments becoming $2J_{2n} = 30$, the test statistic equals 14.0615 while the 5% critical value is 10.77 (for a 10% maximal bias relative to OLS). Note that the Stock and Yogo's (2005) test is conservative, and is based on the parametric 2SLS estimator under the assumption that the model (25) is correctly specified with k_n being fixed and finite known numbers. See Blundell, Chen and Kristensen (2003) for further discussion.

Together 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 that is used for the SMD estimator of the shape-invariant Engel curve model developed in this paper. This is given further consideration when we report the estimated curves below.

7.3. *Estimation Results*

The estimation of the system was performed as follows: First we obtained a profile estimator of $\alpha = (\theta, \Pi)$ using the closed form sieve IV solution of $\widehat{\Pi}(\theta)$. We then used this estimator as a starting point for the numerical optimization procedure employed to obtain simultaneous estimates of θ and Π. In most cases however, the simultaneous estimates are practically identical to the initial profile ones. Note also that we do 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 θ, we run the 3-step procedure described in Section 3. In the 2nd step, the conditional covariance matrix, $\Sigma_o(X)$, is estimated using either the sieve method or the kernel method. The results reported here are based on kernel estimates, but $\hat{\theta}$ proves to be fairly robust to the choice of estimator for $\Sigma_o(X)$. To improve the estimates we run an iterative procedure, repeating Step 2 and 3 until $\hat{\theta}$ converges towards a stable level. In this application, we require that $||\hat{\theta}^{(i+1)} - \hat{\theta}^{(i)}|| < 0.005$ where $\hat{\theta}^{(i)}$ denotes the estimate obtained in the i-th iteration; this convergence criterion is satisfied after 7-10 iterations.

Figures 6 to 9 present the range of estimated curves for four of the goods in our system. 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 use a 3rd order B-spline sieve for h with number of sieve terms $k_n = 9$, and a 4th order B-spline of dimension $J_{2n} = 15$ for m. We penalize both the level and the 2nd order derivative of h (i.e., $C = C_0 + C_2$), with penalization weight $\lambda = 0.4$. All plots are with identity weighting, $\widehat{\Sigma}(X_i) = I_L$. 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 right hand side plot in each panel that represents the sieve IV estimates under the shape-invariant restriction.

Together with the estimated Engel curves, we also report their 95% pointwise confidence bands. The bands are obtained using the nonparametric bootstrap based on 1000 resamples. In each resample, $n = 1655$ observations are drawn from the original data set with replacement, and then h is reestimated. We perform this with $\theta = \widehat{\theta}$ fixed at its efficient estimated value, since $\widehat{\theta}$ is $\sqrt{n}$-consistent while $\widehat{h}$ has a slower than $\sqrt{n}$-convergence rate; hence, this has no effect asymptotically. For simplicity, we use the same J_{2n} and k_n in the estimation of h using the bootstrap sample, however, to control for the asymptotic bias in $\widehat{h} - h_o$, we slightly decreased λ in the bootstrap sample so in effect we were overfitting (or undersmoothing in kernel literature);[7] see, e.g., Hall (1992, Section 4.5) for theoretical justification of this undersmoothing procedure for kernel least squares regression. In the exogenous case, by Theorem 2 in Newey (1997), $\widehat{h}(y_2)$ has a pointwise asymptotically normal distribution; hence, in this case the bootstrap yields consistent estimates of the true confidence bands; see, e.g., Theorem 1.2.1 in Politis, et al. (1999). In the endogenous case, we have no theoretical justification for the bootstrap since we have not derived a pointwise asymptotic distribution of $\widehat{h}(y_2)$, but we conjecture such a result 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.[8]

Several interesting features are present in the plots. As may be expected the estimated shares of 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 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.

To check the robustness of our Engel curve estimates with respect to the choices of sieve basis functions and the smoothing parameters, we carry out extensive sensitivity analysis. We approximate h by several different sieve bases such as wavelet cardinal B-splines, polynomial splines and Hermite polynomials. For the endogenous case, we also approximate the conditional mean m by cosine series and wavelet cardinal B-splines. To examine the robustness of the sieve estimator of the conditional mean, we also employ kernel regression method to estimate m; this gives very similar results that are available from the authors on request. 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 $J_{2n} = 15$ to 27. Generally the smoothness parameter λ should increase slightly as k_n grows. We try out different values of λ for each value of k_n in the range 5 to 9; the shapes of the estimated Engel curves are fairly robust towards the choice of λ.[9] In Figure 10 we report four semiparametric IV estimates of Engel curves for the leisure goods that correspond to four different values of the smoothing parameter $\lambda = 0.1$, 0.3, 0.5 and 0.7. As noted above, the value of λ used in Figure 9 is 0.4. There appears to be some overall robustness in the Engel curve shape to the choice of λ although for low values of λ the curve appears too variable, which is not too surprising since $k_n = 9$ here. Analyses display a similar pattern for the other goods and are available from the authors on request.

The semiparametric efficient estimates of θ are given in Table 2. These estimates have been obtained using the same functional bases and the same $J_{2n} = 15$ and $k_n = 9$ as used to obtain the Engel curves h estimates, except with a smaller λ. The estimates of θ 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 θ 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 $n(\widehat{\theta}_{LS} - \widehat{\theta}'_{IV})(\widehat{V}_{LS} - \widehat{V}_{IV})^{-1}(\widehat{\theta}_{LS} - \widehat{\theta}_{IV}) \overset{asy.}{\sim} \chi^2_{L+1}$ under the null. This Hausman test for the exogeneity of Y_2 produces a statistic of 880.06 with critical value $\chi^2_8(95) = 15.5$, and we reject the null-hypothesis. That is, the data supports the hypothesis that Y_2 is endogenous. The results show a strong impact on θ_1 of allowing for endogeneity. This parameter measures the general log equivalence scale for the presence of children with a couple normalized to unity. The LS estimate is implausibly low whereas the IV estimate is very plausible and represents an equivalence scale of about .45, normalized 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 θ_2; e.g. the negative value of θ_2 for alcohol shows the decline in the overall alcohol budget share, given total equalized expenditure, that occurs for larger households.

Table 2: Efficient estimates of θ in the exog. and endog. case

	Semiparametric IV		Semiparametric LS	
	coefficient	std. ($10^{-3}\times$)	coefficient	std. ($10^{-3}\times$)
θ_1	0.3698	57.4712	0.1058	34.3810
θ_2 - food-in	0.0213	6.5406	0.0461	4.8861
θ_2 - food-out	0.0006	3.6744	-0.0046	2.4182
θ_2 - alcohol	-0.0216	4.5047	-0.0239	2.5322
θ_2 - fares	-0.0023	2.5089	-0.0092	1.4027
θ_2 - fuel	-0.0035	2.7611	0.0054	1.9069
θ_2 - leisure	0.0388	10.9148	-0.0016	6.2392
θ_2 - travel	-0.0384	5.9912	-0.0226	3.9748

To check the robustness of our θ estimates with respect to the choices of sieve basis functions ψ^{k_n} and p^{J_n}, we also 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 $J_{2n} = 15$ to 27. The estimates $\widehat{\theta}$ are very similar to the ones reported in Table 2, and are also stable as k_n increases in both the exogenous and the endogenous cases.[10] To conserve space, here we only report a small sensitivity check in terms of θ estimates under the endogeneity. Although the θ estimates reported in Table 3 are obtained using smaller penalization λ than those reported in Table 2, the estimated values are virtually the same.

Table 3: Estimates of θ: Sensitivity Analysis

	Semiparametric IV							
	θ estimates							
k_n, J_{2n}	Food-in	Food-out	Alcohol	Fares	Fuel	Leisure	Travel	θ_1
B-spl 9, B-spl 15	0.0207	0.0003	-0.0210	-0.0019	-0.0038	0.0422	-0.0393	0.3834
B-spl 9, B-spl 25	0.0171	-0.0005	-0.0233	-0.0005	-0.0027	0.0489	-0.0419	0.4113
B-spl 8, Cos 20	0.0204	-0.0005	-0.0248	-0.0009	-0.0029	0.0458	-0.0387	0.3989
P-spl 8, B-spl 15	0.0209	-0.0019	-0.0222	-0.0004	-0.0029	0.0429	-0.0359	0.3981
P-spl 5, B-spl 25	0.0191	-0.0002	-0.0285	-0.0011	-0.0038	0.0496	-0.0399	0.4088

Finally, we briefly mention the results of a number of further comparisons that we have carried out. The first implements the control function approach of Newey, Powell and Vella (1999). As one might expect, this gives estimates that are between our sieve IV-estimates (for the endogenous case) and

the sieve LS estimates (for the exogenous case). Second, we compare our semi-nonparametric model with a parametric quadratic model of the QUAIDS-class proposed in Banks, Blundell and Lewbel (1997). A test on θ rejects the QUAIDS-model in favour of the semi-nonparametric model. Third, we implement the SMD procedure using three different years of FES data sets, and try both gross earnings and disposable income as the instrument. The empirical findings are again surprisingly robust to the choice of sieve bases and smoothing parameters in the sense of being qualitatively similar, and are available from the authors on request.

8. CONCLUSIONS

In this paper we study the sieve semi- /nonparametric IV estimation of a shape-invariant Engel curve system with endogenous total expenditure. We provide identification and establish the nonparametric convergence rate and semiparametric efficiency properties of our estimators under relatively "low-level" sufficient conditions. We also present Monte Carlo simulation results that shed some lights on the choice of smoothing parameters and the performance of the sieve nonparametric IV estimator.

In our application to consumer behavior in the UK FES we show the importance of allowing for endogeneity and document the relatively simple steps involved in implementing the sieve semi-nonparametric IV method. The shape-invariant system of Engel curve specification, that pools across demographic groups, enable us to estimate the parametric effects of equivalence scales and the demographic impacts accurately and efficiently. 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. We also contrast our sieve IV estimator with the sieve LS estimator that 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, e.g., Matzkin (1994).

Dept. of Economics, University College, London WC1E 6BT, U.K.; r.blundell@ucl.ac.uk,

Dept. of Economics, New York University, 19 W. 4th Street, 6th Floor, New York, NY 10012, U.S.A.; xiaohong.chen@nyu.edu

and

Dept. of Economics, Columbia University, 420 W. 118th Street, New York, NY 10027, U.S.A.; dk2313@columbia.edu

Manuscript received October, 2003; final revision received April, 2007.

APPENDIX A: PROOFS

PROOF OF THEOREM 1: Without loss of generality and given condition (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:

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

since ϕ is known, conditions (1) and (2) imply:

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

Now $h_{ol}(\cdot)$ is identified. This and conditions (3) and (4), together with conditional moment restriction (4), identify θ_{o1} and $\theta_{o2,l}$. Since for all $l = 1, ..., L$,

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

this and (4) imply:

$$E[h_{ol}(Y_{2i} - \phi(\theta_{o1})) - h_{ol}(Y_{2i} - \phi(\theta_1)) + (\theta_{o2,l} - \theta_{2,l}) \mid X_{1i} = 1, X_{2i}] = 0 \tag{26}$$

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

$$E[\nabla h_{ol^*}(Y_{2i} - \phi(\overline{\theta}_{o1})) \mid X_{1i} = 1, X_{2i}] \times (\phi(\theta_1) - \phi(\theta_{o1})) + (\theta_{o2,l^*} - \theta_{2,l^*}) = 0$$

where $\overline{\theta}_{o1}$ is some value between θ_{o1} and θ_1. Again by conditions (1) and (2) and $\nabla h_{ol^*}(Y_{2i} - \phi(\overline{\theta}_{o1})) \neq \text{const.}$ (in particular $\neq 0$) imply that $E[\nabla h_{ol^*}(Y_{2i} - \phi(\overline{\theta}_{o1})) \mid X_{1i} = 1, X_{2i}] \neq \text{const.}$ (in particular $\neq 0$), hence

$$\phi(\theta_1) - \phi(\theta_{o1}) = 0 \text{ and } \theta_{o2,l^*} - \theta_{2,l^*} = 0$$

by condition (5) we have $\theta_1 - \theta_{o1} = 0$, this and (26) imply $\theta_{o2,l} - \theta_{2,l} = 0$ for $l = 1, ..., L$. *Q.E.D.*

PROOF OF 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 $l = 1, ..., L$. To simplify notation further, we assume $\phi(0) = 0$ and suppress the conditioning variable $X_1 = 0$ and drop the subscript l. We denote $\widehat{h} = \arg\min_{h \in \mathcal{H}_n} \frac{1}{n} \sum_{i=1}^n \widehat{m}(X_{2i}, h)^2$ where

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

and with $p^{J_n}(0, X_2) = B^{J_{2n}}(X_2)$, $\mathbf{B} = (B^{J_{2n}}(X_{21}), ..., B^{J_{2n}}(X_{2n}))'$, $J_n = 2J_{2n}$,

$$\begin{aligned} \widehat{m}(X_{2i}, h) &= \sum_{t=1}^n \{Y_{1t} - h(Y_{2t})\} B^{J_{2n}}(X_{2t})' (\mathbf{B}'\mathbf{B})^{-1} B^{J_{2n}}(X_{2i}) \\ &= \sum_{t=1}^n \{Y_{1t} - \{\psi^{k_n}(Y_{2t})\}'\Pi\} B^{J_{2n}}(X_{2t})' (\mathbf{B}'\mathbf{B})^{-1} B^{J_{2n}}(X_{2i}) \\ &= \widehat{E}[Y_1|X_{2i}] - \widehat{E}[\psi^{k_n}(Y_2)|X_{2i}]'\Pi = \widehat{E}[Y_1|X_{2i}] - \widehat{E}[h(Y_2)|X_{2i}]. \end{aligned}$$

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

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

We first state three claims.

Claim 1: *(i) Under Assumption 2(i)(ii), and the sieve space $\mathcal{H}_n$ given by (9)-(11), we have: There is a finite $c > 0$ such that for any $h \in \mathcal{H}$, there is a $\{\psi^{k_n}(Y_2)\}'\Pi \in \mathcal{H}_n$ satisfying $\left\|h(\cdot) - \{\psi^{k_n}(\cdot)\}'\Pi\right\|_{Y_2} \leq ck_n^{-r}$;*

(ii) Under Assumption 4, we have: there is a finite $c' > 0$ such that for any $g \in \Lambda_c^{r_m}(\mathcal{X}_2)$, there is a $B^{J_{2n}}(X_2)'A$ such that $\left\|g(\cdot) - B^{J_{2n}}(\cdot)'A\right\|_{X_2} \leq c' J_{2n}^{-r_m}$.

Claim 2: *Under Assumptions 1, 2(i)(ii), 3, 4 and 5(i)(ii), we have*:

$$\begin{aligned} (i) \quad ||g_o - \widehat{g}||_{X_2} &= O_p\left(J_{2n}^{-r_m} + \sqrt{J_{2n}/n}\right); \\ (ii) \quad \sup_{h \in \mathcal{H}_n} ||\{\widehat{T} - T\}h||_{X_2} &= O_p\left(J_{2n}^{-r_m} + \sqrt{J_{2n}/n}\right); \\ (iii) \quad \sup_{h \in \mathcal{H}_n} ||\widehat{m}(\cdot, h) - m(\cdot, h)||_{X_2} &= O_p\left(J_{2n}^{-r_m} + \sqrt{J_{2n}/n}\right). \end{aligned}$$

Claim 3: *Under Assumptions 1, 2(i)(ii), 3, 4 and 5(i)(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* $n \to \infty$*.*

We now apply Claims 1, 2 and 3 to continue the proof of Theorem 2. Denote $||h||_{Y_2} \equiv \sqrt{E[h(Y_2)]^2}$. By the triangular inequality and Assumption 2, there is a $\{\psi^{k_n}\}'\Pi_o \in \mathcal{H}_n$ such that

$$||\widehat{h} - h_o||_{Y_2} \leq ||h_o - \{\psi^{k_n}\}'\Pi_o||_{Y_2} + ||\widehat{h} - \{\psi^{k_n}\}'\Pi_o||_{Y_2}.$$

Next by the definition of τ_n and the triangular inequality,

$$\begin{aligned} & ||\widehat{h} - \{\psi^{k_n}\}'\Pi_o||_{Y_2} \leq \tau_n \times ||T\{\widehat{h} - \{\psi^{k_n}\}'\Pi_o\}||_{X_2} \\ \leq \;& \tau_n \times \{||T\widehat{h} - \widehat{g}||_{X_2} + ||\widehat{g} - g_o||_{X_2} + ||Th_o - T\{\psi^{k_n}\}'\Pi_o||_{X_2}\} \\ \leq \;& \tau_n \times \{||\{T - \widehat{T}\}\widehat{h}||_{X_2} + ||\widehat{T}\widehat{h} - \widehat{g}||_{X_2} + ||\widehat{g} - g_o||_{X_2} + ||T\{h_o - \{\psi^{k_n}\}'\Pi_o\}||_{X_2}\}. \end{aligned}$$

By Assumptions 1(iii) and 3, we have $Th_o = g_o \in \Lambda_c^{r_m}(\mathcal{X}_2)$. Also by the definition of $\mathcal{H}_n$ and Assumption 3, we have $Th \in \Lambda_c^{r_m}(\mathcal{X}_2)$ for all $h \in \mathcal{H}_n$. Under Claim 3 and by the definition of $\widehat{h}$, we have:

$$\begin{aligned} ||\widehat{T}\widehat{h} - \widehat{g}||_{X_2}\{1 + o_p(1)\} &= ||\widehat{T}\widehat{h} - \widehat{g}||_{n,X_2} \\ &\leq ||\widehat{T}\{\psi^{k_n}\}'\Pi_o - \widehat{g}||_{n,X_2} = ||\widehat{T}\{\psi^{k_n}\}'\Pi_o - \widehat{g}||_{X_2}\{1 + o_p(1)\}. \end{aligned}$$

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

$$\begin{aligned} & ||\widehat{T}\{\psi^{k_n}\}'\Pi_o - \widehat{g}||_{X_2} = ||\widehat{m}(\cdot, \{\psi^{k_n}\}'\Pi_o)||_{X_2} \\ \leq \;& ||\widehat{m}(\cdot, \{\psi^{k_n}\}'\Pi_o) - m(\cdot, \{\psi^{k_n}\}'\Pi_o)||_{X_2} + ||m(\cdot, \{\psi^{k_n}\}'\Pi_o) - m(\cdot, h_o)||_{X_2} \\ = \;& ||\widehat{m}(\cdot, \{\psi^{k_n}\}'\Pi_o) - m(\cdot, \{\psi^{k_n}\}'\Pi_o)||_{X_2} + ||T\{h_o - \{\psi^{k_n}\}'\Pi_o\}||_{X_2}. \end{aligned}$$

These and Claim 2 imply

$$||\widehat{h} - \{\psi^{k_n}\}'\Pi_o||_{Y_2} \leq \tau_n \times \left\{O_p\left(J_{2n}^{-r_m} + \sqrt{J_{2n}/n}\right) + ||T\{h_o - \{\psi^{k_n}\}'\Pi_o\}||_{X_2} \times \{2 + o_p(1)\}\right\}.$$

Hence

$$||\widehat{h} - h_o||_{Y_2} \leq ||h_o - \{\psi^{k_n}\}'\Pi_o||_{Y_2} + \tau_n \times O_p\left(J_{2n}^{-r_m} + \sqrt{J_{2n}/n} + ||T\{h_o - \{\psi^{k_n}\}'\Pi_o\}||_{X_2}\right).$$

This, Assumptions 2(i)(ii), 5(i)(ii) and 6, and Claim 1(i) imply:

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

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

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

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

see, e.g., Chen, Hansen and Scheinkman (1997). Then:

$$\begin{aligned}
& \int [h(y) - \{\psi^{k_n}(y)\}'\Pi]^2 f_{0,Y_2}(y)dy \\
= & \int [h(y) - \{\psi^{k_n}(y)\}'\Pi]^2 (1+y^2)^{-a} \left[(1+y^2)^a f_{0,Y_2}(y)\right] dy \\
\leq & \left[\sup_y \left|[h(y) - \{\psi^{k_n}(y)\}'\Pi](1+y^2)^{-a/2}\right|\right]^2 \int (1+y^2)^a f_{0,Y_2}(y)dy
\end{aligned}$$

hence

$$\begin{aligned}
\left\|h - \{\psi^{k_n}\}'\Pi\right\|_{Y_2} &\asymp \sqrt{\int [h(y) - \{\psi^{k_n}(y)\}'\Pi]^2 f_{0,Y_2}(y)dy} \\
&\leq C \sup_y \left|[h(y) - \{\psi^{k_n}(y)\}'\Pi](1+y^2)^{-a/2}\right| \leq C' k_n^{-r}.
\end{aligned}$$

(ii) See Timan (1963) for Fourier series and Schumaker (1993) for spline sieve.

Proof of Claim 2: (i) By Assumptions 1, 2(i)(ii) and 3, $g_o = Th_o \in \Lambda_c^{r_m}(\mathcal{X}_2)$. This together with Assumptions 4 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_{2n}, his $\zeta_0(K) =$ our $\sqrt{J_{2n}}$, and his $K^{-\alpha} =$ our $J_{2n}^{-r_m}$, hence we obtain result (i).

(ii) By the definition of $\mathcal{H}_n$ and Assumption 3, we have $Th \in \Lambda_c^{r_m}(\mathcal{X}_2)$ 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 $\widehat{T}h$ is simply the sieve LS regression of $h(Y_2)$ on $B^{J_{2n}}(X_2)$. 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(i)(ii) and 3, we have $m(\cdot, h) \in$

$\Lambda_c^{r_m}(\mathcal{X}_2)$ 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 $\widehat{m}(\cdot, h)$ is simply the sieve LS regression of $Y_1 - h(Y_2)$ on $B^{J_{2n}}(X_2)$, hence $\widehat{m}(\cdot, h)$ belongs to the closed linear span of $B^{J_{2n}}(X_2)$ with probability approaching to one. Now we go through the proof of Lemma 4 in Huang (1998) with his A_n = our $\sqrt{J_{2n}}$ and his N_n = our J_{2n}. Under our Assumptions 1 - 4 and 5(ii), we notice that Huang's result actually holds uniformly over $h \in \mathcal{H}_n$, hence we obtain Claim 3. *Q.E.D.*

For any h, $h' \in L_2(\mathcal{R}, f_{0,Y_2})$, define the inner product $\langle h, h' \rangle_{Y_2} \equiv \int_{\mathcal{R}} h(y)h'(y)f_{0,Y_2}dy$ and the norm $||h||_{Y_2} = \sqrt{\langle h, h \rangle_{Y_2}}$. Similarly we define the inner product $\langle , \rangle_{X_2}$ and the norm $|| \cdot ||_{X_2}$on $L_2(\mathcal{R}, f_{0,X_2})$. Let $f_{0,Y_2|X_2}$ denote the conditional density of Y_2 given $\{X_2, X_1 = 0\}$. Assumptions 1(iii) and 4(iii) imply that the densities f_{0,Y_2}, f_{0,X_2}, f_{0,X_2,Y_2} and $f_{0,Y_2|X_2}$ are all continuous. Let $L_\infty(\mathcal{R})$ and $L_\infty(\mathcal{X}_2)$ be the spaces of bounded measurable functions of Y_2 and X_2 respectively. Under the identification condition, the conditional expectation operator, $\{Th\}(X_2) = \int h(y)f_{0,Y_2|X_2}(y|X_2)dy$, is a bounded linear operator from $L_\infty(\mathcal{R})$ into $L_\infty(\mathcal{X}_2)$, and is invertible on its range (i.e., $N(T) = \{h : Th = 0\} = \{0\}$). It is obvious that T is also a bounded operator from $L_2(\mathcal{R}, f_{0,Y_2})$ into $L_2(\mathcal{X}_2, f_{0,X_2})$, with its kernel function defined as

$$t(x_2, y_2) \equiv \frac{f_{0,X_2,Y_2}(x_2, y_2)}{f_{0,X_2}(x_2)f_{0,Y_2}(y_2)} = \frac{f_{0,Y_2|X_2}(y_2|x_2)}{f_{0,Y_2}(y_2)} \quad \text{for any } (x_2, y_2) \in \mathcal{X}_2 \times \mathcal{R}.$$

Denote the kernel function of the self-adjoint operator $TT^* : L_2(\mathcal{X}_2, f_{0,X_2}) \to L_2(\mathcal{X}_2, f_{0,X_2})$ as

$$t_2(w, x_2) \equiv \int_{\mathcal{R}} t(w, y)t(x_2, y)f_{0,Y_2}(y)dy \quad \text{for any } (w, x_2) \in \mathcal{X}_2 \times \mathcal{X}_2.$$

Assumptions 1(iii) and 4(iii) imply that $t_2(w, x_2)$ is continuous on $\mathcal{X}_2 \times \mathcal{X}_2$, where $\mathcal{X}_2$ is a compact interval. Hence TT^* is a compact operator (see e.g., theorem 3.1.5 in Zimmer, 1990). This implies that T, T^* and T^*T are all compact. It is well-known that the compact operators T^*T and TT^* share the same eigenvalue sequence $\{\mu_k^2\}_{k=1}^\infty$ with $\mu_1^2 = 1 \geq \mu_2^2 \geq \mu_3^2 \geq ... \searrow 0$, and for all $k \geq 1$,

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

The corresponding eigenfunctions, $\{\phi_{1k}\}_{k=1}^\infty$ (of T^*T) and $\{\phi_{0k}\}_{k=1}^\infty$ (of TT^*), are orthonormal bases for $L_2(\mathcal{R}, f_{0,Y_2})$ and $L_2(\mathcal{X}_2, f_{0,X_2})$ respectively. The non-negative value of $\sqrt{\mu_k^2}$ is called the $k-$th singular number of T (and of T^*), which is also denoted as μ_k to simplify notation, and $\{\mu_k; \phi_{1k}, \phi_{0k}\}_{k=1}^\infty$ is called the singular value system of T. By the identification condition we have $\mu_k > 0$ for all $k \geq 1$. Therefore, the inverse operator T^{-1} is not bounded.

Proof of Lemma 1: As already mentioned, Assumptions 1(iii) and 4(iii) imply that the operators T, T^*, TT^* and T^*T are all compact with the singular value system $\{\mu_k; \phi_{1k}, \phi_{0k}\}_{k=1}^\infty$. For any $h \in L_2(\mathcal{R}, f_{0,Y_2})$,

$g \in L_2(\mathcal{X}_2, f_{0,X_2})$, we have

$$(Th)(x_2) = \sum_{k=1}^{\infty} \mu_k \langle h, \phi_{1k} \rangle_{Y_2} \phi_{0k}(x_2), \quad (T^* g)(y_2) = \sum_{k=1}^{\infty} \mu_k \langle g, \phi_{0k} \rangle_{X_2} \phi_{1k}(y_2),$$

$$\frac{f_{0,X_2,Y_2}(x_2, y_2)}{f_{0,X_2}(x_2) f_{0,Y_2}(y_2)} = \sum_{k=1}^{\infty} \mu_k \phi_{0k}(x_2) \phi_{1k}(y_2), \quad \mu_1^2 = 1 \geq \mu_2^2 \geq \mu_3^2 \geq ... \searrow 0.$$

(1) Let $clsp\{\mathcal{H}_n\} \equiv \left\{ \psi^{k_n}(y)'\Pi \in L_2(\mathcal{R}, f_{0,Y_2}) : \Pi \in \mathcal{R}^{k_n} \right\}$ be the closed linear subspace (in $L_2(\mathcal{R}, f_{0,Y_2})$) that is generated by the sieve basis functions $\psi^{k_n}(\cdot)$ for $\mathcal{H}_n$. It is easy to see that

$$\tau_n^2 = \sup_{h \in clsp\{\mathcal{H}_n\}: h \neq 0} \frac{E[h(Y_2)]^2}{E[\{Th\}(X_2)]^2} = \frac{1}{\inf_{h \in clsp\{\mathcal{H}_n\}: ||h||_{Y_2} = 1} ||Th||_{X_2}^2}. \tag{27}$$

Let $\mathcal{P}_{k_n - 1} = clsp\{\phi_{1k} : k = 1, ..., k_n - 1\}$ and $(\mathcal{P}_{k_n-1})^{\perp}$ be its orthogonal complement in $L_2(\mathcal{R}, f_{0,Y_2})$. Since $\dim(clsp\{\mathcal{H}_n\}) = k_n$, there is a $\widetilde{h} \in clsp\{\mathcal{H}_n\} \cap (\mathcal{P}_{k_n-1})^{\perp}$ with $||\widetilde{h}||_{Y_2} = 1$ such that

$$\begin{aligned} \frac{1}{(\tau_n)^2} &= \inf_{h \in clsp\{\mathcal{H}_n\}: ||h||_{Y_2} = 1} ||Th||_{X_2}^2 \leq ||T\widetilde{h}||_{X_2}^2 \leq \sup_{h \in (\mathcal{P}_{k_n-1})^{\perp}: ||h||_{Y_2} = 1} ||Th||_{X_2}^2 \\ &= \sup_{h \in (\mathcal{P}_{k_n-1})^{\perp}: ||h||_{Y_2} = 1} \langle T^* Th, h \rangle_{Y_2} = (\mu_{k_n})^2. \end{aligned}$$

Hence $\tau_n \geq \frac{1}{\mu_{k_n}}$.

(2) Under the stated assumptions, for any $h_n \in clsp\{\mathcal{H}_n\}$, we have the alternative representation: $h_n(y_2) = \{\psi^{k_n}(y_2)\}'\Pi = \sum_{k=1}^{k_n} \langle h_n, \phi_{1k} \rangle_{Y_2} \phi_{1k}(y_2)$.

(2.i) For any $h \in clsp\{\mathcal{H}_n\}$ with $h \neq 0$,

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

$$||Th||_{X_2}^2 = \sum_{j=1}^{k_n} \mu_j^2 \{ \left\langle h, \phi_{1j} \right\rangle_{Y_2} \}^2 \geq \mu_{k_n}^2 \sum_{j=1}^{k_n} \{ \left\langle h, \phi_{1j} \right\rangle_{Y_2} \}^2 = \mu_{k_n}^2 \left\| h \right\|_{Y_2}^2.$$

Hence

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

(2.ii) Notice that

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

Hence,

$$\begin{aligned} T\{h_o - \{\psi^{k_n}\}'\Pi_o\}(x_2) &\equiv E[h_o(Y_2) - \{\psi^{k_n}(Y_2)\}'\Pi_o | X_2 = x_2] \\ &= \int (h_o(y_2) - \{\psi^{k_n}(y_2)\}'\Pi_o) \frac{f_{0,X_2,Y_2}(x_2, y_2)}{f_{0,X_2}(x_2) f_{0,Y_2}(y_2)} f_{0,Y_2}(y_2) dy_2 \\ &= \sum_{k=k_n+1}^{\infty} \langle h_o, \phi_{1k} \rangle_{Y_2} \times \mu_k \phi_{0k}(x_2), \end{aligned}$$

and

$$\begin{aligned} ||T\{\{\psi^{k_n}\}'\Pi_o - h_o\}||_{X_2}^2 &= \sum_{k=k_n+1}^{\infty} \{\langle h_o, \phi_{1k} \rangle_{Y_2}\}^2 \times \mu_k^2 \\ &\leq \mu_{k_n}^2 \times \sum_{k=k_n+1}^{\infty} \{\langle h_o, \phi_{1k} \rangle_{Y_2}\}^2 = \mu_{k_n}^2 \times ||h_o - \{\psi^{k_n}\}'\Pi_o||_{Y_2}^2. \end{aligned}$$

Therefore,

$$||T\{\{\psi^{k_n}\}'\Pi_o - h_o\}||_{X_2} \leq \mu_{k_n} \times ||h_o - \{\psi^{k_n}\}'\Pi_o||_{Y_2}.$$

This and Result (2.i) imply Assumption 6. *Q.E.D.*

PROOF OF THEOREM 3: (1) By theorem II.5.10 of Edmunds and Evans (1987, page 91), μ_k, the $k-$th eigenvalue of $(T^*T)^{1/2}$ equals the $k-$th approximation number of $(T^*T)^{1/2}$. Recall that $(T^*T)^{1/2}$ maps $L_2(\mathcal{Y}, f_{0,Y_2})$ isomorphically onto $W_2^s(\mathcal{Y})$, and f_{0,Y_2} is bounded above and bounded away from zero over its support $\mathcal{Y}$, which is a bounded interval of $\mathcal{R}$. We have that the $k-$th approximation number of $(T^*T)^{1/2}$ equals the $k-$th approximation number of the compact embedding operator mapping from $W_2^s(\mathcal{Y})$ into $L_2(\mathcal{Y})$, which is k^{-s} by theorem 3.3.4 of Edmunds and Triebel (1996, page 119, equation (2)).

(2.i) Recall that the sieve measure of ill-posedness is

$$\tau_n = \sup_{h \in clsp\{\mathcal{H}_n\}: h \neq 0} \frac{||h||_{Y_2}}{||Th||_{X_2}} = \sup_{h \in clsp\{\mathcal{H}_n\}: h \neq 0} \frac{||h||_{Y_2}}{||(T^*T)^{1/2} h||_{Y_2}}.$$

Since f_{0,Y_2} is bounded above and bounded away from zero over its support $\mathcal{Y}$, which is a bounded interval of $\mathcal{R}$, by duality, we have for any $h \in clsp\{\mathcal{H}_n\}$,

$$||h||_{Y_2} = \sup_{f \in clsp\{\mathcal{H}_n\}: ||f||_{Y_2} \leq 1} \left| \langle h, f \rangle_{Y_2} \right| \leq \sup_{f \in clsp\{\mathcal{H}_n\}: ||f||_{Y_2} \leq 1} \left\{ ||h||_{W_2^{-s}(\mathcal{Y})} \times ||f||_{W_2^s(\mathcal{Y})} \right\}.$$

Since $||(T^*T)^{1/2}h||_{Y_2} \asymp ||h||_{W_2^{-s}(\mathcal{Y})}$, we have

$$\begin{aligned}\tau_n &\asymp \sup_{h\in clsp\{\mathcal{H}_n\}:h\neq 0} \frac{||h||_{Y_2}}{||h||_{W_2^{-s}(\mathcal{Y})}} \leq \sup_{f\in clsp\{\mathcal{H}_n\}:||f||_{Y_2}\leq 1} ||f||_{W_2^s(\mathcal{Y})} \\ &= \sup_{f\in clsp\{\mathcal{H}_n\}} \frac{||f||_{W_2^s(\mathcal{Y})}}{||f||_{Y_2}} \asymp \sup_{f\in clsp\{\mathcal{H}_n\}} \frac{||f||_{W_2^s(\mathcal{Y})}}{||f||_{L_2(\mathcal{Y})}}.\end{aligned}$$

When the support $\mathcal{Y}$ is a bounded interval of $\mathcal{R}$, the inverse inequalities (or Bernstein inequalities) hold when the sieve space $clsp\{\mathcal{H}_n\}$ is either wavelets with $\gamma > s$, or B-splines with $\gamma > s$, or cosine bases. That is, there is a constant $c > 0$ such that $||f||_{W_2^s(\mathcal{Y})} \leq c(k_n)^s||f||_{L_2(\mathcal{Y})}$ for all $f \in clsp\{\mathcal{H}_n\}$; see, e.g., Meyer (1992) for wavelet sieve and Schumaker (1993) for spline and cosine sieves. Therefore we obtain:

$$\tau_n \asymp \sup_{h\in clsp\{\mathcal{H}_n\}:h\neq 0} \frac{||h||_{Y_2}}{||h||_{W_2^{-s}(\mathcal{Y})}} \leq \sup_{f\in clsp\{\mathcal{H}_n\}} \frac{||f||_{W_2^s(\mathcal{Y})}}{||f||_{L_2(\mathcal{Y})}} \leq c(k_n)^s.$$

(2.ii) Notice that

$$(k_n)^{-r_m} \leq ||T\{h_o - \Pi_o'\psi^{k_n}\}||_{X_2} \equiv ||(T^*T)^{1/2}\{h_o - \Pi_o'\psi^{k_n}\}||_{Y_2} \asymp ||h_o - \Pi_o'\psi^{k_n}||_{W_2^{-s}(\mathcal{Y})}.$$

By duality, we have

$$\begin{aligned}||h_o - \Pi_o'\psi^{k_n}||_{W_2^{-s}(\mathcal{Y})} &= \sup_{f:||f||_{W_2^s(\mathcal{Y})}\leq 1} \left|\left\langle h_o - \Pi_o'\psi^{k_n}, f\right\rangle_{L_2(\mathcal{Y})}\right| \\ &= \sup_{f:||f||_{W_2^s(\mathcal{Y})}\leq 1} \left|\left\langle h_o - \Pi_o'\psi^{k_n}, f - P_{\mathcal{H}_n}f\right\rangle_{L_2(\mathcal{Y})}\right| \\ &\leq ||h_o - \Pi_o'\psi^{k_n}||_{L_2(\mathcal{Y})} \times \sup_{f:||f||_{W_2^s(\mathcal{Y})}\leq 1} ||f - P_{\mathcal{H}_n}f||_{L_2(\mathcal{Y})},\end{aligned}$$

where $P_{\mathcal{H}_n}$ denotes the $L_2(\mathcal{Y})-$orthogonal projection onto the sieve space $\mathcal{H}_n$. When the support $\mathcal{Y}$ is a bounded interval of $\mathcal{R}$, and when the sieve space $\mathcal{H}_n$ is either wavelets with $\gamma > s$, or B-splines with $\gamma > s$, or cosine bases, we have the approximation error: $||f - P_{\mathcal{H}_n}f||_{L_2(\mathcal{Y})} \leq c \times (k_n)^{-s}$ for all $f \in W_2^s(\mathcal{Y})$ with $||f||_{W_2^s(\mathcal{Y})} \leq 1$, where the constant c does not depend on f; see e.g. Meyer (1992) for wavelet sieve and Schumaker (1993) for spline and cosine sieves. Since $f_{0,Y_2}(y)$ is continuous, and is bounded above and below over its support $\mathcal{Y}$ that is a bounded interval of $\mathcal{R}$, we have $||h_o - \Pi_o'\psi^{k_n}||_{L_2(\mathcal{Y})} \asymp ||h_o - \Pi_o'\psi^{k_n}||_{Y_2}$, hence

$$||T\{h_o - \Pi_o'\psi^{k_n}\}||_{X_2} \asymp ||h_o - \Pi_o'\psi^{k_n}||_{W_2^{-s}(\mathcal{Y})} \leq c(k_n)^{-s} \times ||h_o - \Pi_o'\psi^{k_n}||_{Y_2}.$$

This and Result (2.i) imply

$$\tau_n \times ||T\{h_o - \Pi_o' \psi^{k_n}\}||_{X_2} \leq c \times ||h_o - \Pi_o' \psi^{k_n}||_{Y_2}.$$

Result (2.iii) immediately follows from Results (2.i), (2.ii) and Theorem 2 (1). *Q.E.D.*

PROOF OF THEOREM 4: (1) Since $||(T^*T)^{1/2}h||_{Y_2} \geq c||h||_{W_2^{-s}(\mathcal{R},w_a)}$, we have

$$\tau_n = \sup_{h \in clsp\{\mathcal{H}_n\}: h \neq 0} \frac{||h||_{Y_2}}{||Th||_{X_2}} = \sup_{h \in clsp\{\mathcal{H}_n\}: h \neq 0} \frac{||h||_{Y_2}}{||(T^*T)^{1/2}h||_{Y_2}} \leq \sup_{h \in clsp\{\mathcal{H}_n\}: h \neq 0} \frac{||h||_{Y_2}}{c||h||_{W_2^{-s}(\mathcal{R},w_a)}}.$$

Since $f_{0,Y_2}(y)$ is continuous on $\mathcal{R}$ and $f_{0,Y_2}(y) \asymp w_a(y)$ for large $|y|$, by duality theorem we have for any $h \in clsp\{\mathcal{H}_n\}$,

$$||h||_{Y_2} = \sup_{f \in clsp\{\mathcal{H}_n\}: ||f||_{Y_2} \leq 1} \left|\langle h, f \rangle_{Y_2}\right| \leq const. \sup_{f \in clsp\{\mathcal{H}_n\}: ||f||_{Y_2} \leq 1} \left\{ ||h||_{W_2^{-s}(\mathcal{R},w_a)} \times ||f||_{W_2^s(\mathcal{R},w_a)} \right\}.$$

Thus

$$\begin{aligned} \tau_n &\leq \sup_{h \in clsp\{\mathcal{H}_n\}: h \neq 0} \frac{||h||_{Y_2}}{c||h||_{W_2^{-s}(\mathcal{R},w_a)}} \leq const. \sup_{f \in clsp\{\mathcal{H}_n\}: ||f||_{Y_2} \leq 1} ||f||_{W_2^s(\mathcal{R},w_a)} \\ &= c' \sup_{f \in clsp\{\mathcal{H}_n\}} \frac{||f||_{W_2^s(\mathcal{R},w_a)}}{||f||_{Y_2}} = c' \sup_{f \in clsp\{\mathcal{H}_n\}} \frac{||f\sqrt{w_a}||_{W_2^s}}{||f\sqrt{f_{0,Y_2}}||_{L_2}} \leq c' \sup_{f \in clsp\{\mathcal{H}_n\}} \frac{||f\sqrt{w_a}||_{W_2^s}}{||f\sqrt{w_a}||_{L_2}}. \end{aligned}$$

Since the sieve space $\mathcal{H}_n$ is given in (9)-(11) with $\gamma > s > 0$, we can apply the inverse inequality for wavelets and B-spline wavelets spaces (see e.g. theorem 2.3 in Meyer, 1992) and obtain:

$$\tau_n \leq c' \sup_{f \in clsp\{\mathcal{H}_n\}} \frac{||f\sqrt{w_a}||_{W_2^s}}{||f\sqrt{w_a}||_{L_2}} \leq const.(k_n)^s.$$

(2) Immediately follows from Result (1) and Theorem 2 (1). *Q.E.D.*

APPENDIX B: ADDITIONAL ASSUMPTIONS FOR SECTION 5

Define $D_w(X, \alpha_o) = (D_{w^1}(X, \alpha_o), D_{w^{2,1}}(X, \alpha_o), ..., D_{w^{2,L}}(X, \alpha_o))$ where $D_{w^j}(X, \alpha_o)$ is given in (18) and (19); this is a $L \times (1 + L)\dim(X_1)$-matrix valued function. Recall that $w^1(\cdot)$ and $w^{2,l}(\cdot)$, $l = 1, ..., L$, are $L \times \dim(X_1)$-matrix valued squared integrable functions of $Y_2 - \phi(X_1'\theta_{o1})$. Let $w^* = \left(w^{*1}, w^{*2,1}, ..., w^{*2,L}\right)$ where $w^{*j}(Y_2 - \phi(X_1'\theta_{o1})) = (w_1^{*j}(Y_2 - \phi(X_1'\theta_{o1})),....,w_{\dim(X_1)}^{*j}(Y_2 - \phi(X_1'\theta_{o1})))$ is given by

$$w_k^{*j} = \arg \inf_{\substack{w_k^j \in \mathcal{H} - \{h_o\} \\ w_k^j \neq 0}} E\left[D_{w_k^j}(X, \alpha_o)'\Sigma(X)^{-1}D_{w_k^j}(X, \alpha_o)\right] \text{ for } k = 1, ..., \dim(X_1).$$

The asymptotic variance expression for V in Proposition 1 is:

$$V \equiv \left[\begin{array}{c} E\{D_{w^*}(X,\alpha_o)'\Sigma(X)^{-1}D_{w^*}(X,\alpha_o)\}\times \\ \left(E\{D_{w^*}(X,\alpha_o)'\Sigma(X)^{-1}\Sigma_o(X)\Sigma(X)^{-1}D_{w^*}(X,\alpha_o)\}\right)^{-1}\times \\ E\{D_{w^*}(X,\alpha_o)'\Sigma(X)^{-1}D_{w^*}(X,\alpha_o)\} \end{array} \right]. \tag{28}$$

Below are the additional conditions imposed for Propositions 1 - 3:

ASSUMPTION B1: $E[D_{w^*}(X,\alpha_o)'\Sigma(X)^{-1}D_{w^*}(X,\alpha_o)]$ is positive definite.

We note that Assumption B1 is actually a consequence of a more primitive assumption N1 stated in the old version of this paper; see Lemma 1 in Blundell, Chen and Kristensen (2003).

ASSUMPTION B2: For all θ_1 within a $o(n^{-1/4})$-shrinking neighborhood of θ_{o1} and for $x_1 = 0, 1$, (i) $E[w_l^*(Y_2 - \phi(X_1'\theta_1))|X_1 = x_1, X_2 = \cdot]$ belongs to $\Lambda_c^{r_m}(\mathcal{X}_2)$ for $l = 1, ..., L$; (ii) $\nabla\psi^{k_n}(Y_2 - \phi(X_1'\theta_1))$ is continuous in θ_1; (iii) $E[h_n(Y_2 - \phi(X_1'\theta_1))|X_1 = x_1, X_2 = \cdot]$ belongs to $\Lambda_c^{r_m}(\mathcal{X}_2)$ for any $h_n \in \mathcal{H}_n$.

The proofs of Propositions 1 - 3 are direct applications of the theory in Ai and Chen (2003), and can be found in Blundell, Chen and Kristensen (2003); hence we omit them.

REFERENCES

AI, C. AND X. CHEN (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795-1844.

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.

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.

BLUNDELL, R. (1988): "Consumer Behavior: Theory and Empirical Evidence," *Economic Journal*, 98, 16-65.

BLUNDELL, R., M. BROWNING AND I. CRAWFORD (2003): "Nonparametric Engel Curves and Revealed Preference," *Econometrica*, 71, 205-240.

——— (2004), "Best Nonparametric Bounds on Demand Responses", Walras-Bowley Lecture, IFS Working Paper, 05/20.

Blundell, R. X. Chen and D. Kristensen (2003): "Semiparametric Engel Curves with Endogenous Expenditure", CeMMAP Working Paper CWP15/03, UCL, October.

Blundell, R., A. Duncan and K. Pendakur (1998): "Semiparametric Estimation and Consumer Demand," *Journal of Applied Econometrics*, 13, 435-461.

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.

Carrasco, M., J.-P. Florens and E. Renault (2006): "Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization", in J.J. Heckman and E.E. Leamer (eds.), *The Handbook of Econometrics*, vol. 6. North-Holland, Amsterdam, forthcoming.

Cavalier, L., G. Golubev, D. Picard and A. Tsybakov (2002): "Oracle Inequalities in Inverse Problems," *Annals of Statistics*, 30, 843-874.

Chen, X. and M. Reiss (2007): "On Rate Optimality for Nonparametric Ill-posed Inverse Problems in Econometrics," manuscript, New York University, Dept. of Economics.

Chen, X. and X. Shen (1998): "Sieve Extremum Estimates for Weakly Dependent Data," *Econometrica*, 66, 289-314.

Chen, X., L.P. Hansen, and J. Scheinkman (1997): "Shape-preserving Estimation of Diffusions," manuscript, University of Chicago, Dept. of Economics.

——— (2005): "Principal Components and the Long Run," manuscript, University of Chicago, Dept. of Economics.

Chui, C.K. (1992): *An Introduction to Wavelets.* Academic Press.

Cohen, A., M. Hoffmann, and M. Reiss (2004): "Adaptive wavelet Galerkin methods for linear inverse problems," *SIAM J. Numer. Anal.* 42, 1479-1501.

Darolles, S., J.-P. Florens and E. Renault (2006): "Nonparametric Instrumental Regression," mimeo, GREMAQ, University of Toulouse.

Deaton A. S. (1997): The Analysis of Household Surveys: A Microeconometric Approach to Development Policy, World Bank Research Publications, Johns Hopkins Press.

Deaton A. S. and J. Muellbauer (1980): "An Almost Ideal Demand System," *American Economic Review*, 70, 312-36.

Donoho, D. (1995): "Nonlinear Solution of Linear Inverse Problems by Wavelet-Vaguelette Decomposition," *Appl. Comput. Harmonic Anal.* 2, 101-126.

Edmunds, D. and W. Evans (1987): *Spectral Theory and Differential Operators*, Oxford University Press: Oxford.

Edmunds, D. and H. Triebel (1996): *Function Spaces, Entropy Numbers, Differential Operators*, Cambridge University Press: Cambridge.

Efromovich, S. and V. Koltchinskii (2001): "On Inverse Problems with Unknown Operators," *IEEE Trans. on Information Theory*, 47, 2876-2893.

Engl, H., M. Hanke and A. Neubauer (1996): *Regularization of Inverse Problems*, Kluwer Academic Publishers: London.

Eubank, L. (1988): *Nonparametric Regression and Spline Smoothing.* Marcel Dekker.

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.

Florens, J., J. Heckman, C. Meghir and E. Vytlacil (2002): "Instrumental Variables, Local Instrumental Variables and Control Functions," IFS Working paper.

Hall, P. (1992): *The Bootstrap and Edgeworth Expansion.* Springer-Verlag New York.

Hall, P. and J. Horowitz (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *Annals of Statistics*, 33, 2904-2929.

Härdle W. and O. Linton (1994): "Applied Nonparametric Models," in Engle and McFadden, *Handbook of Econometrics*, Vol IV, North-Holland.

Härdle, W. and J. Marron (1990): "Semiparametric Comparison of Regression Curves," *Annals of Statistics*, 18, 63-89.

Hausman, J.A. and Newey, W.K. (1995), "Nonparametric Estimation of Exact Consumer Surplus and Deadweight Loss", *Econometrica* 63, 1445-1476.

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.

Hoeffding, W. (1977): "Some Incomplete and Boundedly Complete Families of Distributions," *Annals of Statistics*, 5, 278-291.

Huang, J. (1998): "Projection Estimation in Multiple Regression with Application to Functional ANOVA Models," *Annals of Statistics*, 26, 242-272.

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.

Korostelev, A. and A. Tsybakov (1993): *Minimax Theory of Image Reconstruction*, Lecture Notes in Statistics, v.82. Springer: New York.

Kress, R. (1999): *Linear Integral Equations*, 2nd ed. Springer: New York.

Lehmann, E.L. (1986): *Testing Statistical Hypothesis*, 2nd ed. Wiley: New York.

Leser, C. E. V. (1963), "Forms of Engel Functions," *Econometrica*, 31, 694-703.

Lewbel, A. (1991): "The rank of Demand Systems: Theory and Nonparametric Estimation," *Econometrica*, 59, 711-730.

Mattner, L. (1993): "Some Incomplete but Bounded Complete Location Families," *Annals of Statistics*, 21, 2158-2162.

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.

Meyer, Y. (1992): *Wavelets and Operators*. Cambridge University Press.

Newey, W.K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147-168.

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

Newey, W.K. and J. Powell (2003): "Instrumental Variables Estimation for Nonparametric Models," *Econometrica*, 71, 1565-1578.

Newey, W. Powell, J. and F. Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67, 565-604.

Pendakur, K. (1998): "Semiparametric Estimates and Tests of Base-independent Equivalence Scales," *Journal of Econometrics*, 88, 1-40.

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.

Politis, D.N., J.P. Romano and M. Wolf (1999): *Subsampling*. Springer-Verlag New York.

Robin, J.-M. and R. Smith (2000): "Tests of Rank," *Econometric Theory*, 16, 151-175.

Robinson, P. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931-954.

Schumaker, L. (1993): *Spline Functions: Basic Theory*, revised ed. Krieger Publishing Co.

Stock, J.H. and M. Yogo (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, eds. D.W.K. Andrews and J.H. Stock. Cambridge: Cambridge University Press.

Stone, C.J. (1982): "Optimal Global Rates of Convergence for Nonparametric Regression", *The Annals of Statistics,* 10, 1040-1053.

Timan, A.F. (1963): *Theory of Approximation of Functions of a Real Variable*, MacMillan, New York.

Triebel, H. (1983): *Theory of Function Spaces*, Birkhauser, Basel.

Working, H. (1943), "Statistical Laws of Family Expenditure", *Journal of the American Statistical Association*, 38, 43-56.

Zimmer, R.J. (1990): *Essential Results of Functional Analysis,* Chicago Lectures in Mathematics, Chicago.

1 Tables

Table 4: MC-study1: MSE of sieve IV-estimator of nonlinear h

	m_n=B-spline, $J_{2n} = 15$					m_n=B-spline, $J_{2n} = 25$				
	λ					λ				
h_n - k_n	0.80	0.10	0.01	0.001	0.00	0.80	0.10	0.01	0.001	0.00
P-spline - 4	2.24	2.31	1.20	1.78	1.75	2.20	2.01	1.52	1.32	1.29
	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 - 6	1.84	1.72	1.43	1.66	2.86	1.76	1.31	0.95	0.98	1.04
	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
B-spline - 9	0.47	1.10	1.51	1.47	0.70	0.50	0.78	0.89	0.93	0.86
	0.61	1.88	5.58	11.83	39.19	0.58	1.59	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 - 14	1.75	1.62	1.50	2.13	6.45	0.54	0.85	1.02	1.16	2.14
	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 5: MC-study1: MSE of sieve estimators of nonlinear h.

	IV-estimator, $J_{2n} = 25$					LS-estimator				
	λ					λ				
h_n - k_n	0.80	0.10	0.01	0.001	0.00	0.80	0.10	0.01	0.001	0.00
B-spline - 9	0.50	0.78	0.90	0.94	0.86	18.77	18.42	18.48	18.39	18.37
	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 - 6	1.76	1.31	0.95	0.98	1.04	19.50	18.82	18.53	18.49	18.48
	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

Notes: The three elements in each cell are, from top to bottom, integrated squared bias ($10^{-2}\times$), integrated variance ($10^{-2}\times$) and integrated mean square error (MSE) ($10^{-2}\times$).

Table 6: MC-study2: MSE of sieve IV-estimators of nonlinear h, bivariate Normal design

$k_n = 6$, $J_{2n} = 23$	h_n=B-spline, m_n=B-spline					h_n=Herm., m_n=Cosine or Herm.				
	λ					λ				
	0.40	0.10	0.01	0.001	0.00	0.40	0.10	0.01	0.001	0.00
uniform X_2						m_n=Cosine				
Bias^2 ($10^{-2}\times$)	0.73	0.43	0.39	0.51	0.53	1.62	0.77	0.77	0.77	0.69
Var. ($10^{-2}\times$)	0.18	0.32	0.47	4.71	7.06	0.27	0.73	1.02	1.48	2.35
MSE ($10^{-2}\times$)	0.91	0.75	0.86	5.22	7.59	1.89	1.50	1.79	2.25	3.04
normal X_2						m_n=Hermite				
Bias^2 ($10^{-2}\times$)	0.83	0.47	0.37	0.15	0.15	2.30	0.83	0.84	0.95	0.67
Var. ($10^{-2}\times$)	0.20	0.33	0.46	2.69	3.73	0.51	1.81	2.26	5.04	21.23
MSE ($10^{-2}\times$)	1.03	0.80	0.83	2.84	3.88	2.81	2.64	3.10	5.99	21.90

Table 7: MC-study2: MSE of sieve estimators of nonlinear h, bivariate Normal design.

$k_n = 9$ normal X_2	IV-estimator, $J_{2n} = 23$					LS-estimator				
	λ					λ				
	0.40	0.10	0.01	0.001	0.00	0.40	0.10	0.01	0.001	0.00
h_n=Bspl, m_n=Bspl										
Bias^2 ($10^{-2}\times$)	0.69	0.43	0.23	0.17	0.17	19.40	19.24	19.11	19.12	19.12
Var. ($10^{-2}\times$)	0.34	0.52	1.24	6.23	6.57	0.07	0.08	0.09	0.11	0.11
MSE ($10^{-2}\times$)	1.03	0.95	1.47	6.40	6.74	19.47	19.32	19.20	19.23	19.23
h_n=Herm, m_n=Herm										
Bias^2 ($10^{-2}\times$)	1.48	0.52	0.56	1.63	1.76	16.69	18.34	18.67	18.90	19.17
Var. ($10^{-2}\times$)	0.41	1.76	2.66	5.50	23.62	0.06	0.05	0.06	0.06	0.08
MSE ($10^{-2}\times$)	1.89	2.28	3.22	7.13	25.38	16.75	18.39	18.73	18.96	19.25

2 Figures

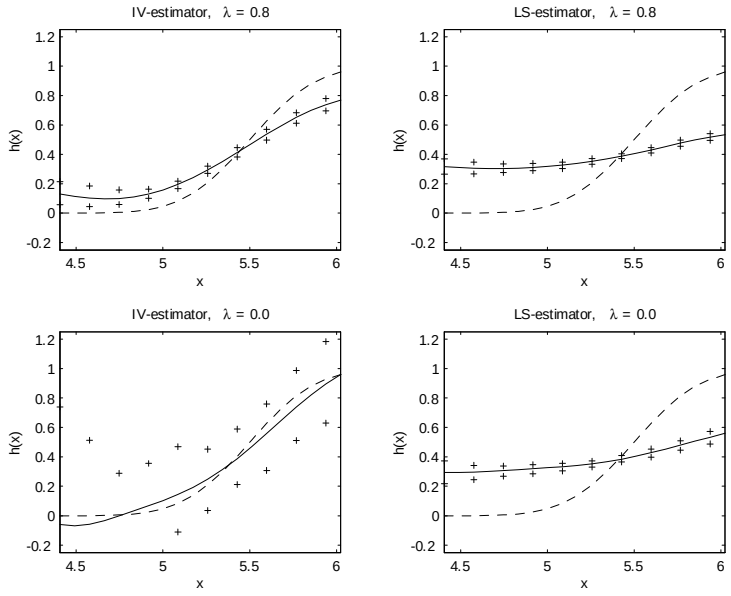

Figure 1: MC-study1: LS- and IV-estimator of nonlinear h, $k_n = 9$.

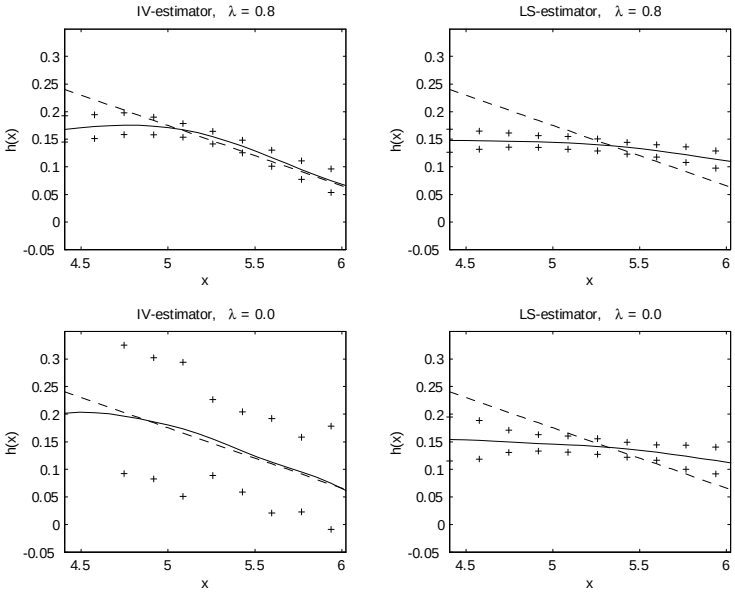

Figure 2: MC-study1: LS- and IV-estimator of linear h, $k_n = 9$.

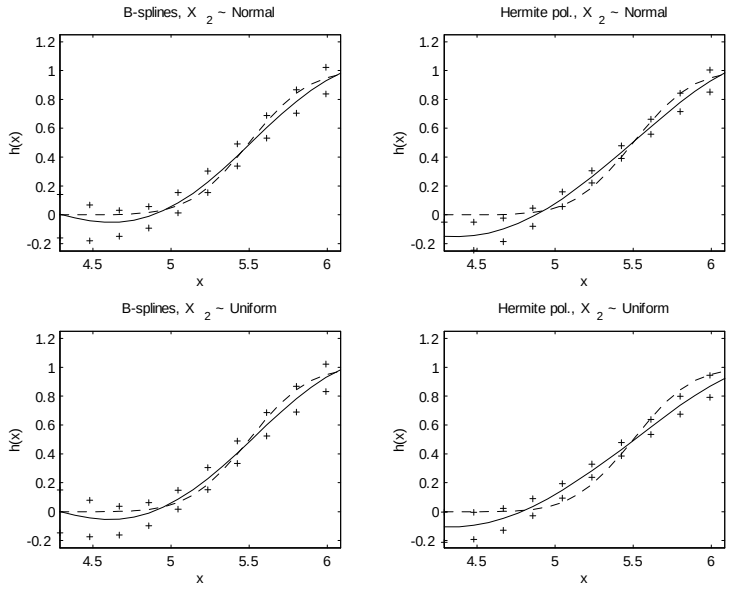

Figure 3: MC-study2: IV-estimator of nonlinear h, instrument $X_2/\Phi(X_2)$, $k_n = 6$, $\lambda = 0.0$.

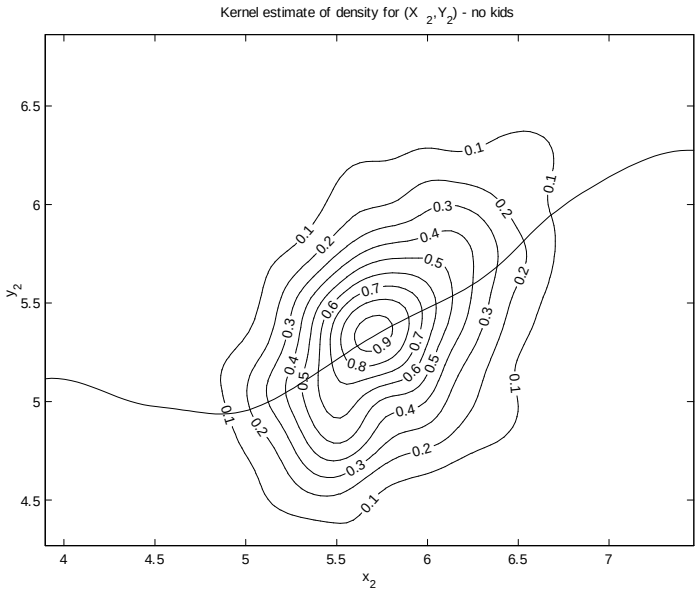

Figure 4: Kernel density estimate of $(X_2 = \log\text{-earning}, Y_2)$; series estimate of $E[Y_2|X_2]$

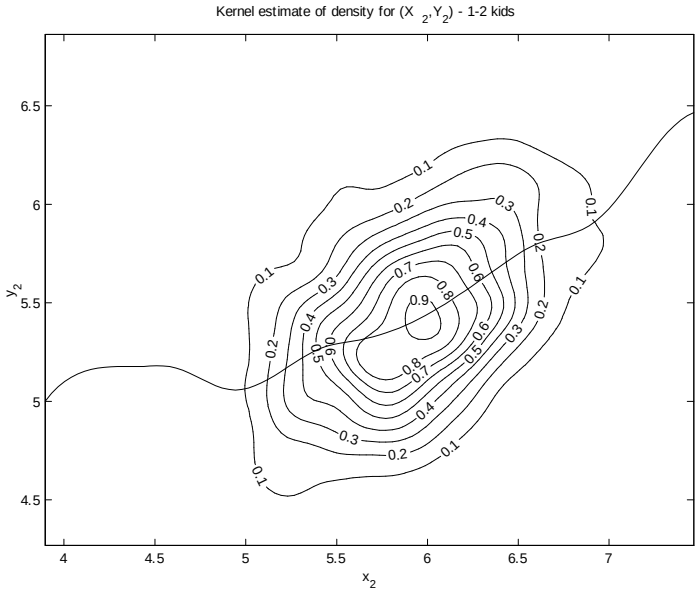

Figure 5: Kernel density estimate of ($X_2 = \log$-earning, Y_2); series estimate of $E[Y_2|X_2]$

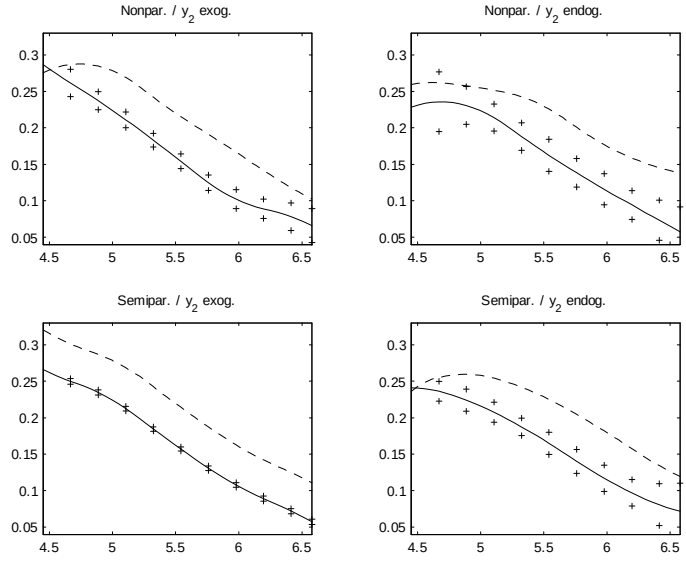

Figure 6: Engel curve for food-in. - - w/ children, — w/o children, ++ 95%-confidence bands

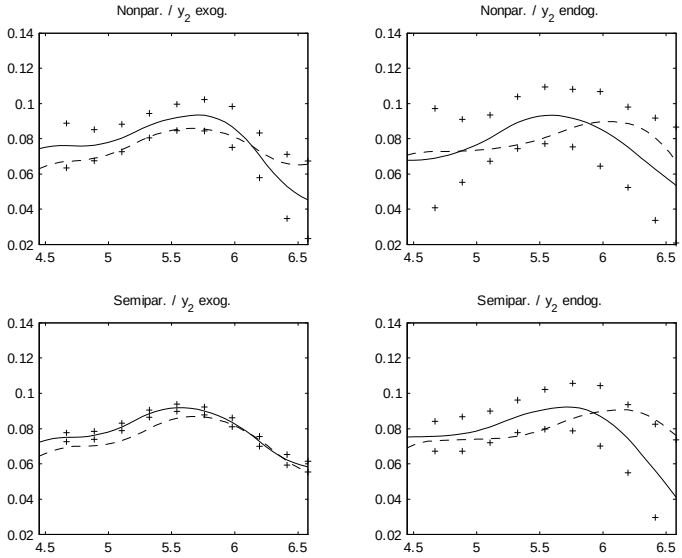

Figure 7: Engel curve for food-out. - - w/ children, — w/o children, ++ 95%-confidence bands

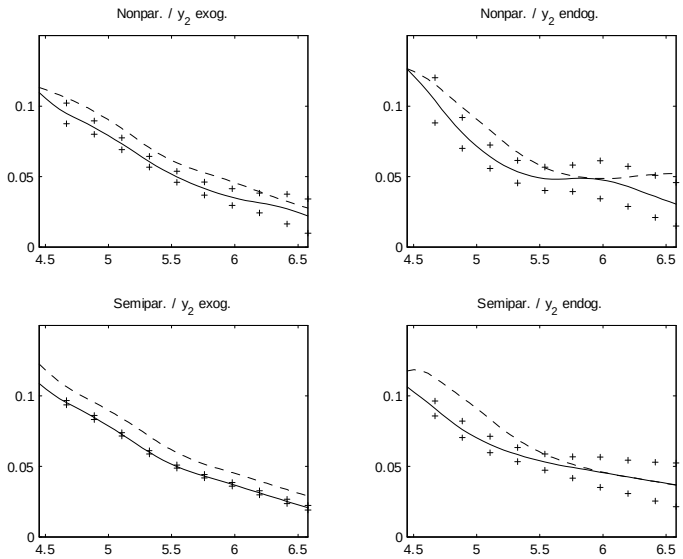

Figure 8: Engel curve for fuel. - - w/ children, — w/o children, ++ 95%-confidence bands

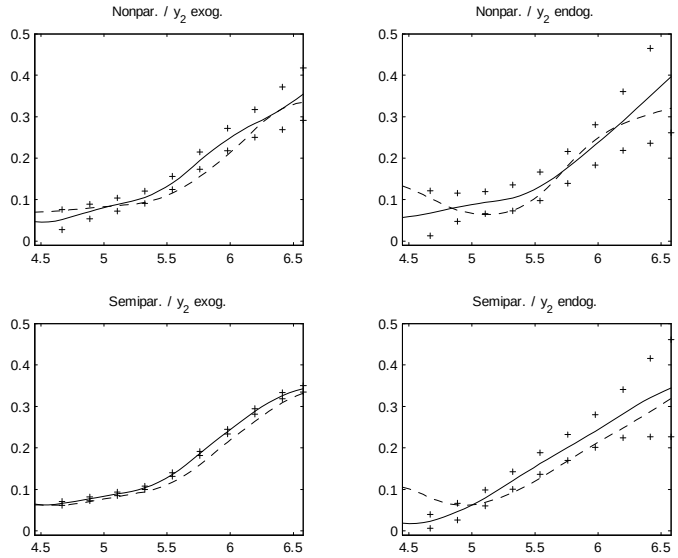

Figure 9: Engel curve for leisure. - - w/ children, — w/o children, ++ 95%-confidence bands

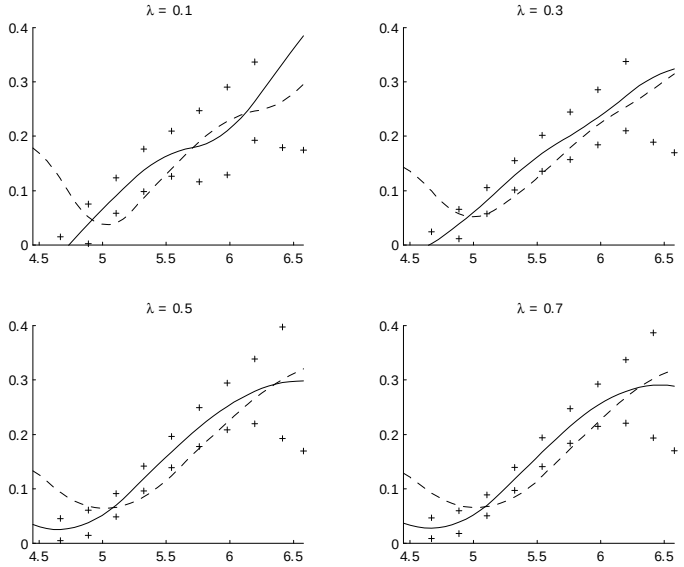

Figure 10: Engel curve for leisure - IV estimate sensitivity to λ, $k_n = 9$.

Footnotes

[1]We thank the co-editor Whitney Newey, several (at least three) anonymous referees, C. Ai, D. Andrews, V. Chernozhukov, I. Crawford, J. Hahn, J. Hausman, B. Honore, J. Horowitz, A. Lewbel, J. Powell and J. 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, Toulouse, UPenn, Harvard/MIT, UCLA, Virginia, Yale and Columbia. This study is part of the program of research of the ESRC Centre for the Microeconomic Analysis of Fiscal Policy at IFS. The authors gratefully acknowledge financial supports from ESRC/UK (Blundell and Chen) and NSF/USA (Chen). The usual disclaimer applies.

[2]See Newey, Powell and Vella (1999) for a nonparametric CF approach and also the reviews by Blundell and Powell (2003), Florens (2003) and Florens, Heckman, Meghir and Vytlacil (2007). Blundell, Duncan and Pendakur (1998) allow for endogeneity of total expenditure using a parametric additive CF approach within the context of a kernel regression framework.

[3]In the mathematics and statistics literature, there are many 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 δ. See, e.g., Kress (1999), Korostelev and Tsybakov (1993), Donoho (1995), Engl, Hanke and Neubauer (1996), Cavalier, Golubev, Picard and Tsybakov (2002) and Cohen, Hoffmann and Reiss (2004). However, the nonparametric IV regression in econometrics corresponds to an ill-posed inverse problem in which both the conditional expectation operator T and the g are unknown; see Carrasco, Florens and Renault (2006) for a recent review.

[4]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 $h_l(Y_{2i} - \phi(X_{1i}'\theta_1)), l = 1, ..., L$, the SMD estimators we propose will be invariant to the commodity omitted.

[5]This is nested within the fully nonparametric specification (1). Blundell, Duncan and Pendakur (1998) compare this specification (2) with the fully nonparametric alternative and find that it provides a good representation of demand behavior for households in the FES.

[6]Let $\hat{h}_j$ be the estimate of h_o from the $j-$th simulated data set, and $\overline{h}(y) = \sum_{j=1}^{100} \hat{h}_j(y)/100$. The pointwise squared bias is $[\overline{h}(y) - h_o(y)]^2$ and the pointwise variance is $100^{-1}\sum_{j=1}^{100}[\hat{h}_j(y) - \overline{h}(y)]^2$. All the tables report the integrated squared bias as $\int_{\underline{y}}^{\overline{y}}[\overline{h}(y) - h_o(y)]^2 dy$, where $\underline{y}$ and $\overline{y}$ are respectively the 2.5th and 97.5th empirical percentiles of Y_2 from the no-kids subsample data $\{Y_{2i}\}_{i=1}^{n=628}$; the reported integrated variance and MSE are computed in a similar way. Another way to report Monte Carlo integrated squared bias is to use $628^{-1}\sum_{i=1}^{628}[\overline{h}(Y_{2i}) - h_o(Y_{2i})]^2$; similarly to report the integrated variance and MSE as sample average across the no-kids subsample data. We have tried both, although they give different numbers (the ones computed using numerical integration against $y \in [\underline{y}, \overline{y}]$ are generally bigger), the qualitative patterns are the same.

[7]Alternatively we could fix λ 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.

[8]Newey (1997) also supplies us with a consistent estimator of the asymptotic variance of $\widehat{h}(y_2)$ 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 $\widehat{h}(y_2)$ in the endogenous case, still assuming that it is asymptotically normal. This gives confidence bands very similar to the ones obtained by the bootstrap, both in the exogenous and endogenous cases.

[9]In general, a smaller penalization is 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.

[10]Inspection of the associated plots for the Engel curves h show that the overall shapes and turning points are maintained for these alternative sieve basis approximations. These findings are consistent with our Monte Carlo results.