PART I

Methods and applications based on kernels

CHAPTER 1

Semiparametric least squares estimation of multiple index models: Single equation estimation

Hidehiko Ichimura and Lung-Fei Lee

1 Introduction

Supported by increasing availability of extensive data sets and computational advancements, a broad class of nonlinear econometric models has been proposed to study more and more realistic empirical situations.[1] Typically these parametric modeling efforts aim to avoid inconsistency of estimates by specifying a model that captures some effects that are not considered in another model. Thus these studies are generally careful in specification of systematic effects. The models are estimated by finite dimensional maximum likelihood method or some computationally simpler method of moment variants, which requires specifying error distribution up to a finite dimensional parameter. Although misspecification of unobserved error distribution as well as misspecification of a systematic effect leads to inconsistent estimates, generally a parametric class of distribution is casually assumed without any justification.

In the first place, Manski (1975) showed that a parametric specification of error distribution is not necessary for consistent estimation in multiple choice models, and thus originated semiparametric estimation literature in econometrics.[2]

Since Manski's study, recognition of inconsistency in maximum likelihood methods under misspecified error distribution has led to many

We are grateful for the research support from the NSF under grant no. SES-8809939. We would like to thank Jim L. Powell, Tom M. Stoker, T. Scott Thompson, and two referees for their valuable comments. Any errors are our own.

[1] Survey articles by Amemiya (1981, 1984) and a book by Maddala (1983) are good sources of these empirical studies.

[2] Stein (1956) originates studies in estimation of a finite dimensional parameter when an infinite dimensional nuisance parameter exists.

studies on estimation methods that do not require specifying a parametric class of error distribution, semiparametric methods. But compared with the rich class of models studied under parametric approaches, models studied under semiparametric methods with known asymptotic distribution have been limited: These are binary choice models, duration models, censored and truncated Tobit models, and selectivity models. Therefore, in order to apply semiparametric methods, researchers have faced a trade-off between inconsistency arising from misspecification in error distributions and inconsistency arising from misspecification in systematic components.

The purpose of this chapter is to extend the applicability of a semiparametric approach. We show that all models that can be represented in multiple index framework (Stoker 1986) can be estimated by the semiparametric least squares method if identification conditions are met.

As we shall see, many econometric models involve multiple indices. But at the same time, we shall see that many examples of multiple index models involve more than one equation. In this chapter we treat single equation estimation. The approach does not account for cross equation parameter restrictions and specific variance-covariance structure implied by each model. Thus when these restrictions are available, the estimator is likely to be inefficient. These issues will be studied in future works.

The estimation method will generalize Ichimura's (1988) semiparametric least squares method originally proposed for single index models to multiple index models. Generalizing his method to multiple index models requires controlling bias terms of higher dimensional conditional expectation estimators. We employ a negative kernel method to handle the problem. For some cases, such as sample selectivity models, it is more natural to allow unbounded dependent variables. Our analysis is more general than Ichimura's original analysis in that it allows both unbounded dependent and independent variables. The estimator is shown to be $\sqrt{n}$-consistent and asymptotically normal. A consistent estimator of the asymptotic variance-covariance matrix is provided.

Section 2 gives the definition and examples of multiple index models. To make our asymptotic analysis transparent, we provide a summary of our analysis in Section 3. Identification issues and consistency of the estimator are discussed in Section 4. After showing sufficient conditions for asymptotic normality of the estimator in Section 5, we construct a consistent estimator of the variance-covariance matrix in Section 6. Section 7 concludes the chapter with some discussion of future works. All the proofs of lemmas and theorems are in the chapter appendix.

2 Multiple index models

We study estimation of multiple index models in single equation context. Let x_i denote all explanatory variables in a model and let x_{li} ($l = 0, 1, \ldots, M$), which are subvectors of x_i, be the explanatory variables in lth index.

Definition 1 (Single equation multiple index model).

1. $y_i = x_{0i}\alpha(\theta_0) + \varphi(x_{1i}\beta_1(\theta_0), \ldots, x_{Mi}\beta_M(\theta_0)) + \epsilon_i$, *where* $i = 1, \ldots, n$.
2. $E(\epsilon_i \mid x_i) = 0$,
3. φ *is not known,*
4. *Functions* $\alpha(\theta), \beta_1(\theta), \ldots, \beta_M(\theta)$ *are all known functions of a basic parameter vector* θ.

The definition allows cross index parameter restrictions by allowing functions $\alpha, \beta_1, \ldots, \beta_M$ to be known functions of a basic parameter vector θ.

The multiple index model as defined here has some similarity with the projection pursuit model studied in statistics literature. For example Friedman and Stuetzle (1981) discuss a model

$$E(y \mid x) = \sum_{m=1}^{M} \phi_m(x_m\beta_0).$$

There are at least two differences between the projection pursuit model and the multiple index model. First, M, the number of indices, is not known in the projection pursuit model but that in the multiple index model is assumed to be known. Second, the indices in the projection pursuit model are additively separable, but those in the multiple index model are not. Given these two differences the results in this chapter are not best suited to the projection pursuit model. The first difference raises the question of whether we can treat M in the multiple index model as an additional parameter. The second difference raises an efficiency question of using an estimation method that does not account for the separability in indices for projection pursuit models. These points are beyond our scope of the chapter.

As we shall see, many econometric models can be regarded as multiple index models with known number of indices. But at the same time, we shall note that many examples of multiple index models involve more than one equation. In this chapter we treat estimation of a single equation. The model does not account for cross-equation parameter restrictions and specific variance-covariance structure implied by each model.

The sample selection model of Gronau (1974) and Heckman (1974) is

$$y_{0i} = x_{0i}\alpha + \nu_{0i},$$
$$y^*_{1i} = x_{1i}\beta + \nu_{1i},$$

where (y_{0i}, x_{0i}) is observed if and only if $y^*_{1i} > 0$. Then as Heckman (1974) shows,[3]

$$y_{0i} = x_{0i}\alpha + \varphi(x_{1i}\beta) + \epsilon_i$$

with $E(\epsilon_i \mid x_i) = 0$ for some unknown function φ. Thus their model is a single index model. Cosslett (this volume) initiated semiparametric estimation of this model. Powell (1987) obtained the first semiparametric estimator of α with known asymptotic distribution for the model assuming existence of a $\sqrt{n}$-consistent estimator of β up to a multiplicative constant. Our approach allows selection to depend on any finite number of equations. That is, models can be

$$y_{0i} = x_{0i}\alpha + \nu_{0i}$$
$$y^*_{1i} = x_{1i}\beta_1 + \nu_{1i}$$
$$\vdots$$
$$y^*_{mi} = x_{Mi}\beta_M + \nu_{Mi},$$

where (y_{0i}, x_{0i}) can be observed if and only if $y^*_{1i} > 0, \ldots,$ and $y^*_{Mi} > 0$, and joint distribution of $(\nu_{0i}, \nu_{1i}, \ldots, \nu_{Mi})$ depends only on $(x_{1i}\beta_1, \ldots, x_{Mi}\beta_M)$. Thus many simultaneous Tobit models and disequilibrium models are multiple index models. See Maddala (1983) to review simultaneous Tobit models and Quandt (1988) to review disequilibrium models.

Another multiple index single equation model is the bivariate choice model with partial observability in Poirier (1980):

$$y^*_{1i} = x_{1i}\beta_1 - \nu_{1i},$$
$$y^*_{2i} = x_{2i}\beta_2 - \nu_{2i},$$

where the binary indicator I_{1i} is 1 if $y^*_{1i} > 0$ and 0 otherwise and I_{2i} is 1 if $y^*_{2i} > 0$ and 0 otherwise. The model is partially observable as the indicators I_{1i} and I_{2i} cannot be observed separately but only a single binary indicator $I_i = I_{1i} \cdot I_{2i}$ is observed. If joint distribution of (ν_{1i}, ν_{2i}) depends only on $x_{1i}\beta_1$ and $x_{2i}\beta_2$, then $E(I_i \mid x_i) = \varphi(x_{1i}\beta_1, x_{2i}\beta_2)$ for some unknown function φ.[4] Thus by defining E_i to be

[3] Heckman assumes (ν_{0i}, ν_{1i}) to be bivariate normal, but the same analysis applies for any distribution of (ν_{0i}, ν_{1i}) if it only depends on $x_{1i}\beta$.

[4] If (ν_{1i}, ν_{2i}) and (x_{1i}, x_{2i}) are independent, then φ is a cumulative distribution function of (ν_{1i}, ν_{2i}).

$$E_i = I_i - \varphi(x_{1i}\beta_1, x_{2i}\beta_2)$$

results in a double index model. As Poirier discusses, his model is applicable to situations where a binary decision is a consequence of decisions made by two individuals. For example, a worker's leaving a firm may be a result of quitting (the worker's decision) or a result of being fired (the firm's decision). Thus Poirier's model applies to cases where an econometrician cannot distinguish quitting from being fired. Another example discussed by Poirier is a closed vote, where unanimity of a committee is required for an approval of a motion. Because our model is not restricted to two indices, a binary decision can involve more than two individuals. A partial observability model also arises because of data limitations. Suppose in multiple choice models we only observe whether one chose a specific item or not; then a partial observability model results.

Many other models can be put into multiple index form. We emphasize here that in order to apply the estimation method one needs to verify identification conditions stated in Section 4 after transforming a model to a multiple index model.

3 Summary of basic analysis, and results

Since φ is an unknown function nonlinear least squares method is not applicable. Ichimura (1988) observed that, in single index context, in order to estimate θ_0 we do not need to know φ as a function of θ but only need to know $E(y_i \mid x_{1i}\beta_1(\theta))$ as a function of θ,[5] which in turn is estimable by any nonparametric regression estimator. Ichimura used a kernel regression estimator to estimate $E(y_i \mid x_{1i}\beta_1(\theta))$. We follow his method and apply it to multiple index models.[6] The estimator minimizes

$$Q_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} I_{Xi}[y_i - x_{0i}\alpha(\theta) - E_n(x_i, \theta)]^2, \tag{1}$$

where I_{Xi} is 1 if $x_i \in X$ and 0 otherwise,

$$E_n(x_i, \theta) = \frac{\sum_{j \neq i}^{n} [y_j - x_{0j}\alpha(\theta)] K\left(\frac{x_i\beta(\theta) - x_j\beta(\theta)}{a_n}\right)}{\sum_{k \neq i}^{n} K\left(\frac{x_i\beta(\theta) - x_k\beta(\theta)}{a_n}\right)} \tag{2}$$

and

$$x_i\beta(\theta) = (x_{1i}\beta_1(\theta), \ldots, x_{Mi}\beta_M(\theta)).$$

[5] This statement holds, of course, subject to identification conditions.
[6] For motivation of the estimation method see Ichimura (1988).

Summations in equation (2) are taken over the n sample except the ith one. Moreover the function K in equation (2) is a function on R^M satisfying the conditions:

$$\int_{R^M} |K(s)|\, ds < \infty \quad \text{and} \quad \int_{R^M} K(s)\, ds = 1.$$

A priori chosen positive sequence $\{a_n\}$ is known as bandwidth or window width and satisfies

$$\lim_{n\to\infty} a_n = 0.$$

We make additional assumptions on a function K and a sequence $\{a_n\}$ as needed.

$E_n(x_i, \theta)$ is known as a kernel regression estimator of

$$E[y_i - x_{0i}\alpha(\theta) \mid x_{1i}\beta_1(\theta), \ldots, x_{Mi}\beta_M(\theta)].$$

See Parzen (1962), Nadaraja (1964), and Watson (1964) for original expositions and Bierens (1987) for a survey. The set X is chosen to be a compact set so that

$$E_n(x_i, \theta) \xrightarrow{p} E_\infty(x_i, \theta)$$

uniformly over $X \times \Theta$, where

$$E_\infty(x_i, \theta) = E[y_i - x_{0i}\alpha(\theta) \mid x_i\beta(\theta)].$$

Here we allow x_i to have unbounded support and allow density of $x_i\beta(\theta)$ to be zero at boundaries for some $\theta \in \Theta$.

In the following we sketch our analysis and some of the results. Rigorous analysis is left for the subsequent sections.

With continuous kernel functions, the objective function $Q_n(\theta)$ is continuous. A standard procedure to establish consistency of an extremum estimator is: First show that the objective function converges in probability to a limit function $Q_\infty(\theta)$ uniformly in $\theta \in \Theta$ and second, show that θ_0 is the unique minimizer of $Q_\infty(\theta)$.[7] We take a similar approach. A difference is that each of the ith summand in the objective function in (1) involves all data and therefore summands are not mutually independent. We show that

$$E_n(x_i, \theta) \xrightarrow{p} E_\infty(x_i, \theta)$$

uniformly in $(x_i, \theta) \in X \times \Theta$. Together with some regularity conditions it implies the uniform convergence of the objective function.

[7] See, for example, Amemiya (1985).

To obtain asymptotic distribution of the estimator we assume that the objective function $Q_n(\theta)$ is a smooth function of θ. Then asymptotic distribution of the estimator can be analyzed by Taylor's expansion.

The estimator $\hat{\theta}_n$ satisfies the first-order conditions:

$$\frac{1}{n}\sum_{i=1}^{n} I_X(x_i)[y_i - x_i\alpha(\hat{\theta}_n) - E_n(x_i,\hat{\theta}_n)]\left[\frac{\partial\alpha'(\hat{\theta}_n)}{\partial\theta}x_{0i}' + \frac{\partial E_n(x_i,\hat{\theta}_n)}{\partial\theta}\right] = 0.$$

By a Taylor's expansion at $\theta = \theta_0$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_{Xi}[y_i - x_i\alpha(\theta_0) - E_n(x_i,\theta_0)]\left[\frac{\partial\alpha'(\theta_0)}{\partial\theta}x_{0i}' + \frac{\partial E_n(x_i,\theta_0)}{\partial\theta}\right] \tag{3}$$

$$= \left\{\frac{1}{n}\sum_{i=1}^{n} I_{Xi}\left[\frac{\partial\alpha'(\bar{\theta}_n)}{\partial\theta}x_{0i}' + \frac{\partial E_n(x_i,\bar{\theta}_n)}{\partial\theta}\right]\left[x_{0i}\frac{\partial\alpha(\bar{\theta}_n)}{\partial\theta'} + \frac{\partial E_n(x_i,\bar{\theta}_n)}{\partial\theta'}\right]\right. \tag{4}$$

$$-\frac{1}{n}\sum_{i=1}^{n} I_{Xi}[y_i - x_i\alpha(\bar{\theta}_n) - E_n(x_i,\bar{\theta}_n)]$$

$$\left.\times\left[\frac{\partial^2 x_{0i}\alpha(\bar{\theta}_n)}{\partial\theta\,\partial\theta'} + \frac{\partial^2 E_n(x_i,\bar{\theta}_n)}{\partial\theta\,\partial\theta'}\right]\right\}$$

$$\times\sqrt{n}(\hat{\theta}_n - \theta_0), \tag{5}$$

where $\bar{\theta}_n$ lies between $\hat{\theta}_n$ and θ_0.

To derive the asymptotic distribution we proceed in three steps:

Step 1: Show that equation (4) converges to a positive definite matrix.

Step 2: Show that equation (5) converges to zero in probability.

Step 3: Show that equation (3) converges in distribution to a multivariate normal random vector.

These steps depend on the uniform convergence of conditional expectations and their first and second derivatives.

To show Step 1, we first show

$$\frac{\partial E_n(x_i,\theta)}{\partial\theta} \xrightarrow{p} \frac{\partial E_\infty(x_i,\theta)}{\partial\theta}.$$

uniformly in $(x_i,\theta)\in X\times\Theta$. Thus difference between equation (4) and

$$\frac{1}{n}\sum_{i=1}^{n} I_{Xi}\left[\frac{\partial\alpha'(\bar{\theta}_n)}{\partial\theta}x_{0i}' + \frac{\partial E_\infty(x_i,\bar{\theta}_n)}{\partial\theta}\right]\left[x_{0i}\frac{\partial\alpha(\bar{\theta}_n)}{\partial\theta'} + \frac{\partial E_\infty(x_i,\bar{\theta}_n)}{\partial\theta'}\right] \tag{6}$$

converges in probability to zero. Using usual uniform law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} I_{Xi}\left[\frac{\partial\alpha'(\theta)}{\partial\theta}x_{0i}' + \frac{\partial E_\infty(x_i,\theta)}{\partial\theta}\right]\left[x_{0i}\frac{\partial\alpha(\theta)}{\partial\theta'} + \frac{\partial E_\infty(x_i,\theta)}{\partial\theta'}\right]$$

converges uniformly to

$$E\left\{I_{Xi}\left[\frac{\partial \alpha'(\theta)}{\partial \theta}x'_{0i}+\frac{\partial E_\infty(x_i,\theta)}{\partial \theta}\right]\left[x_{0i}\frac{\partial \alpha(\theta)}{\partial \theta'}+\frac{\partial E_\infty(x_i,\theta)}{\partial \theta'}\right]\right\}.$$

Since consistency of $\hat{\theta}_n$ implies consistency of $\bar{\theta}_n$ to θ_0, equation (6) converges to a matrix,

$$E\left\{I_{Xi}\left[\frac{\partial \alpha'(\theta_0)}{\partial \theta}x'_{0i}+\frac{\partial E_\infty(x_i,\theta_0)}{\partial \theta}\right]\left[x_{0i}\frac{\partial \alpha(\theta_0)}{\partial \theta'}+\frac{\partial E_\infty(x_i,\theta_0)}{\partial \theta'}\right]\right\}.$$

Thus together with identification conditions Step 1 is proven.

To show Step 2, we show

$$\frac{\partial^2 E_n(x_i,\theta)}{\partial\theta\,\partial\theta'}\xrightarrow{p}\frac{\partial^2 E_\infty(x_i,\theta)}{\partial\theta\,\partial\theta'}$$

uniformly in $(x_i,\theta)\in X\times\Theta$. Thus the difference between equation (5) and

$$\frac{1}{n}\sum_{i=1}^{n} I_{Xi}[y_i-x_i\alpha(\bar{\theta}_n)-E_\infty(x_i,\bar{\theta}_n)]\left[\frac{\partial^2 x_{0i}\alpha(\bar{\theta}_n)}{\partial\theta\,\partial\theta'}+\frac{\partial^2 E_\infty(x_i,\bar{\theta}_n)}{\partial\theta\,\partial\theta'}\right] \tag{7}$$

converges to zero. To see that equation (7) converges to zero first note that by a usual uniform law of large number,

$$\frac{1}{n}\sum_{i=1}^{n} I_{Xi}[y_i-x_i\alpha(\theta)-E_\infty(x_i,\theta)]\left[\frac{\partial^2 x_{0i}\alpha(\theta)}{\partial\theta\,\partial\theta'}+\frac{\partial^2 E_\infty(x_i,\theta)}{\partial\theta\,\partial\theta'}\right]$$

converges uniformly to

$$E\left\{I_{Xi}[y_i-x_i\alpha(\theta)-E_\infty(x_i,\theta)]\left[\frac{\partial^2 x_{0i}\alpha(\theta)}{\partial\theta\,\partial\theta'}+\frac{\partial^2 E_\infty(x_i,\theta)}{\partial\theta\,\partial\theta'}\right]\right\}.$$

As noted previously, $\bar{\theta}_n$ converges in probability to θ_0. Therefore equation (7) converges to

$$E\left\{I_{Xi}[y_i-x_i\alpha(\theta_0)-E_\infty(x_i,\theta_0)]\left[\frac{\partial^2 x_{0i}\alpha(\theta_0)}{\partial\theta\,\partial\theta'}+\frac{\partial^2 E_\infty(x_i,\theta_0)}{\partial\theta\,\partial\theta'}\right]\right\}. \tag{8}$$

But because

$$E_\infty(x_i,\theta_0)=\varphi(x_i\beta_0)$$

and $E(\epsilon_i\,|\,x_i)=0$, the expected value in equation (8) is zero and thus Step 2 follows.

To show Step 3, first note that