Semiparametric Single Index Models: Ichimura and Klein and Spady's methods

Isa Marques, Xi Sun, Xueying Liu

Bonn, Germany
January 18, 2017

Outline

- Introduction
- Identification
- Ichimura (1993) method
- Klein and Spady (1993) method
- Monte Carlo simulation

Introduction

Definition of Single Index Model: (following notation in Li and Racine (2007))

$$Y_i = g(X_i'\beta) + \epsilon_i, \tag{1}$$

1. $\{x_i, y_i\}$ for i = 1, ..., n is an i.i.d. sample;
2. Y_i is the dependent variable, $X_i \in \mathbb{R}^q$ is a vector of explanatory variables, β is the q $\times$ 1 vector of unknown parameters;
3. The functional form of the linear index, $X_i'\beta$ is specified and it is a single index because it is a scalar;
4. $g : \mathbb{R} \to \mathbb{R}$ is unspecified;
5. The distribution of ϵ_i is not specified except $E(\epsilon_i | x_i) = 0$.

Advantage of Single Index Model vs. Parametric and Nonparametric Models

- Parametric model: prespecifies functional forms which is assumed to be fully described by a finite set of parameters.
  eg, Binary choice model:

  $$E(Y|x) = 1 - F(-(\alpha + x'\beta))$$

  Misspecification of error distribution leads to inconsistent estimation of parameters.
- Fully nonparametric models:

  $$Y = g(X, \epsilon) \text{ or } Y = g(X) + \epsilon$$

  Curse of dimensionality: as the dimensions of the model increase, the convergence rate of the estimator decreases.

Identification

Identification(Horowitz (2009)): β and $g(\cdot)$ must be uniquely determined by sample data.

$$E(Y|x) = g(x'\beta_0). \tag{2}$$

Proposition (Identification of a Single Index Model)

1. *x should not contain a constant and it must contain at least one continuous variable with nonzero coefficient. Furthermore, one component of β_0 is set to 1.*
2. *The support of $x'\beta_0$ is bounded convex set with at least one interior point. g is differentiable and it is not a constant function on the support of $x'\beta_0$.*
3. *For the discrete components of x, varying the values of the discrete variables will not divide the support of $x'\beta_0$ into disjoint subsets, and g must be nonperiodic.*

Identification

$$E(Y|x) = g(x'\beta_0).$$

Proposition (Identification of a Single Index Model)

- *1. x should not contain a constant and it must contain at least one continuous variable with nonzero coefficient. Furthermore, one component of β_0 is set to 1.*

Intuition:

- $g^*(\gamma + \delta v) = g(v)$, for all v in the support of $X'\beta$.
  restriction on γ: location normalization,
  restriction on δ: scale normalization.
- Suppose finitely many v in support, there may exist infitely many satisfying $g(\cdot)$, indistinguishable from each other.

Identification

$$E(Y|x) = g(x'\beta_0).$$

Proposition (Identification of a Single Index Model)

- *3. For the discrete components of x, varying the values of the discrete variables will not divide the support of $x'\beta_0$ into disjoint subsets, and g must be nonperiodic.*

Intuition: Assume a continuous X_1 with support $\left[0,1\right]$, and a discrete X_2 with support $\{0,1\}$, g is strictly increasing and non periodic and set $\beta_1 = 1$ as a *scale normalization*.

$$E[Y|X = (x_1, 0)] = g(x_1), \text{support of } g(\cdot) : [0,1];$$
$$E[Y|X = (x_1, 1)] = g(x_1 + \beta_2), \text{support of } g(\cdot) : [\beta_2, 1 + \beta_2].$$

Ichimura's (1993) method

For known g, a nonlinear least squares (NLS) method can be used to estimate β_0 by minimizing:

$$S_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left[Y_i - g(X_i'\beta)\right]^2 \tag{3}$$

with respect to β. However, both g and β_0 are unknown. For a given β we estimate instead:

$$G(X_i'\beta) \overset{def}{=} E(Yi|X_i'\beta) = E[g(X_i'\beta_0)|X_i'\beta]. \tag{4}$$

Given this, the weighted NLS problem is as follows:

$$S_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}[Y_i - \hat{G}_{-i}(X_i'\beta)]^2 w(x_i)\mathbf{1}(X_i \in A_n) \tag{5}$$

where $\hat{G}_{-i}(X_i'\beta)$ is a leave-one-out Nadaraya-Watson kernel estimator, $\mathbf{1}(Xi \in A_n)$ is a trimming function and $w(x_i)$ is a weighting function.

Ichimura's (1993) method

Theorem Ichimura

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, \Omega_I),$$

where $\Omega_I = V^{-1}\Sigma V^{-1}$, and

$$\Sigma = E\{w(X_i)^2\sigma^2(X_i)(g_i^{(1)})^2(X_i - E_A(X_i|X_i'\beta_0)) \times (X_i - E_A(X_i|X_i'\beta_0))'\},$$

with $g_i^{(1)} = [\partial g(v)/\partial v]|_{v=X_i'\beta_0}$, $E_A(X_i|v) = E(X_i|x_A'\beta_0 = v)$ with x_A having the distribution of X_i conditional on $Xi \in A_\delta$, and

$$V = E[w(X_i)(g_i^{(1)})^2(X_i - E_A(X_i|X_i'\beta_0))(X_i - E_A(X_i|X_i'\beta_0))'].$$

Assumptions: $\beta_n - \beta_0 = O(n^{-\frac{1}{2}})$, $\hat{\beta}_n - \beta_0 = O_p(n^{-\frac{1}{2}})$; $\hat{G}_{-i}(X_i'\beta_n) = G(X_i'\beta_n) + o_p(1)$; $\hat{G}_{-i}(X_i'\beta_0) = g(X_i'\beta_0) + o_p(1)$.

1. It can be shown $S_n(\beta_n) = \frac{1}{n}\sum_i\{g(X_i'\beta_0) - E[g(X_i'\beta_0)|X_i'\beta_n] + \epsilon_i\}^2 + o_p(1)$.
2. With two Taylor expansions we have $g(X_i'\beta_0) - E[g(X_i'\beta_0)|X_i'\beta_n)] = g^{(1)}(X_i'\beta_n)(X_i - E[X_i'|X_i'\beta_n)(\beta_0 - \beta_n) + o_p(1)$
3. Minimize S_n in order to β_n.

Ichimura's (1993) method

- **Bandwidth Selection**: Ichimura requires $h_n = O(n^{-\frac{1}{5}})$. Haerdle et al.(1993) suggest an empirical way of selecting the bandwidth for optimal smoothing of both g and β.
- **Weight Function**: In case data is heteroskedastic, use analogue of Feasible Generalized Least Squares. Under certain regularity conditions, the efficiency bound for the single index model with unknown g and using only data for which $X \in A_\delta$ is Ω_I with $w(x) = \frac{1}{\sigma^2(x)}$. The efficiency bound is then

$$\Omega_{SI} = \left\{ E\left[\frac{1}{\sigma^2(x)} \frac{\partial}{\partial \beta} G(X'\beta, \beta) \frac{\partial}{\partial \beta} G(X'\beta, \beta) \right] \right\}^{-1}. \qquad (6)$$

  For unknown $\sigma^2(x)$, a consistent estimator is used that follows a two-step procedure.
- **Main Problem**: Uses iterative method, particularly difficult if the objective function is multimodal or nonconvex.

Klein and Spady's (1993) method

The model is defined as $Y_i = \mathbf{1}(X_i'\beta \geq \epsilon_i)$. Let g denote the distribution of ϵ_i. Thus, the *log − likelihood* objective function is as follows:

$$\mathcal{L}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \tau_i \{(1 - Y_i) ln[1 - \hat{G}_{-i}(X_i'\beta)] + Y_i ln[\hat{G}_{-i}(X_i'\beta)]\} \tag{7}$$

where τ_i is a simplified trimming function $\tau_i = \mathbf{1}(X_i \in A_\delta)$.

Theorem Klein and Spady

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, \Omega_{KS}),$$

where $\Omega_{KS} = \left\{ E\left[\frac{\partial}{\partial\beta} G(X_i'\beta) \frac{\partial}{\partial\beta} G(X_i'\beta)' \frac{1}{g(X_i'\beta_0)(1-g(X_i'\beta_0))} \right] \right\}^{-1}$ and $\Omega_{KS} = \Omega_{SI}$, i.e., the estimator is asymptotically efficient.

Klein and Spady's (1993) method

- **Bandwidth Selection**: Klein and Spady require $n^{-\frac{1}{6}} < h_n < n^{-\frac{1}{8}}$. Härdle et al.'s (1993) solution can potentially be applied here.
- **Main Problem**: Computation is difficult.
- **Comparison between Ichimura's and Klein and Spady's model**: Klein and Spady's model seems more appropriate for the binary choice model case, from a theoretical perspective. Ichimura's model uses a weight function to correct for heteroskedasticity. However, Klein and Spady's model is efficient in the sense that it reaches the semiparametric efficiency bound.

Implementation

What we have done

- Klein and Spady function, Ichimura function
  - Kernel (arbitrary choice: Gaussian)
  - function of G_{-i} and loss function S_n and $\mathcal{L}_n$
  - Minimisation using grid-search (arbitrary choice of grid)

What we are going to do

- Trimming Function in the minimization process $\mathbf{1}(X_i \in A_n)$
- Weighing fuction for heteroskedasticity: $w(x_i)$

What we leave out

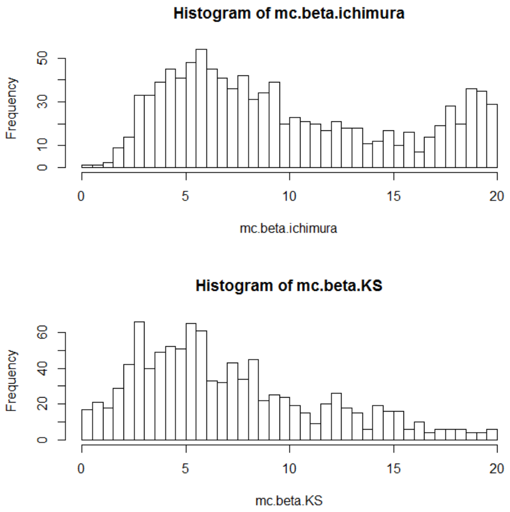
- Bandwidth selection, computationally difficult
  - Candidate for bandwidth: 0.2

Monte Carlo Simulation

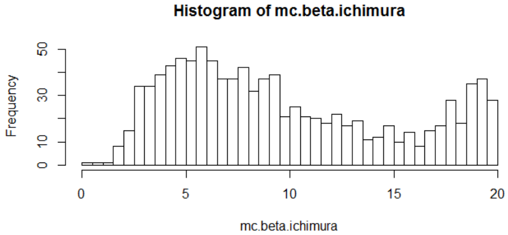
What we have done

- Generated binary dataset, specifying true function g: $e \overset{i.i.d.}{\sim} N(0,1)$
- Define loss function $\frac{1}{m}\sum_{i=1}^{m}(\hat{\beta}_i - \beta_0)^2$
- Vary n (sample size), m (number of Monte Carlo simulation) is set to 1000
- $\beta_0 = 7.6$ (arbitrary choice)

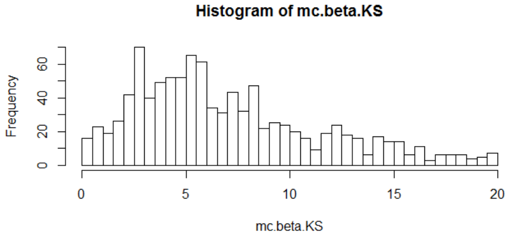
n = 30


Histogram of mc.beta.ichimura
Frequency
0
10
30
50
0
5
10
15
20
mc.beta.ichimura
Histogram of mc.beta.KS
Frequency
0
20
40
60
0
5
10
15
20
mc.beta.KS

n = 100



Histogram of mc.beta.ichimura
Frequency
0
10
30
50
0
5
10
15
20
mc.beta.ichimura



Histogram of mc.beta.KS
Frequency
0
20
40
60
0
5
10
15
20
mc.beta.KS

Monte Carlo Simulation: comparison based on loss function

Number observations	**Ichimura**	**Klein and Spady**
30	68.16583	38.60356
100	67.99048	38.59288

Table: Results for loss function

Monte Carlo Simulation

What we are going to do

- Continuous data (for comparison of Ichimura vs OLS)
- Binary data (for comparison of Probit, Ichimura, Klein and Spady)
- Vary the true distribution g (for comparison of Ichimura and Klein and Spady)

What we leave out

- Comparsion between g and G_{-i}

Application on Real Data

Data Features

- Binary outcome: (Male or Female Voice)
- Include at least one continuous variable
- Features of voice: frequency, spectral flatness, peak frequency etc

Steps in Estimation

- Separate data into:
  - Training sample
  - Prediction sample
- Same range of independent variable
- Training sample to estimate $\hat{\beta}$ and G_{-i}
- Use those estimates on prediction sample
- Define accuracy rate for evaluation

References I

Horowitz, Joel L. (2009). *Semiparametric and Nonparametric Methods in Econometrics*. Springer Science+Business Media.

Ichimura, Hidehiko (1993). "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models". In: *Journal of Econometrics* 58.1, pp. 71–120.

Klein and Spady (1993). "An Efficient Semiparameteric Estimator For Binary Response Models". In: *Econometrica* 61.2, pp. 287–421.

Li, Qi and Jeffrey Scott Racine (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.