

# Semiparametric Single Index Models

---

Yasuyuki Matsumura

November 25, 2024

Graduate School of Economics, Kyoto University

- A semiparametric single index model is given by

$$Y = g(X^T \beta_0) + u,$$

where

$Y \in \mathbb{R}$  : a dependent variable,

$X \in \mathbb{R}^q$  : a  $q \times 1$  explanatory vector,

$\beta_0 \in \mathbb{R}^q$  : a  $q \times 1$  vector of unknown parameters,

$u \in \mathbb{R}$  : an error term which satisfies  $\mathbb{E}(u \mid X) = 0$ ,

$g(\cdot)$  : an unknown distribution function.

# Introduction

- Even though  $x$  is a  $q \times 1$  vector,  $x^T \beta_0$  is a scalar of a single linear combination, which is called **a single index**.
- By the form of the single index model, we obtain

$$\mathbb{E}(Y \mid X) = g(X^T \beta_0),$$

which means that the conditional expectation of  $Y$  only depends on the vector  $X$  through a single index  $X^T \beta_0$ .

- The model is semiparametric when  $\beta \in \mathbb{R}^q$  is estimated with the parametric methods and  $g(\cdot)$  with the nonparametric methods.
- Some of the parametric single index models are really familiar with us.

## Examples of Parametric Single Index Model

- If  $g(\cdot)$  is the identity function, then the model turns out to be a linear regression model:

$$Y = g(X^T \beta_0) + u = X^T \beta_0 + u.$$

- If  $g(\cdot)$  is the CDF of Normal(0, 1), then the model turns out to be a probit model.
  - See the textbook for further discussions on a probit model.
- If  $g(\cdot)$  is the CDF of logistic distribution, then the model turns out to be a logistic regression model.

# Identification Conditions

---

# Identification Conditions

## Proposition 8.1 (Identification of a Single Index Model)

For the semiparametric single index model  $Y = g(x^T \beta_0) + u$ , identification of  $\beta_0$  and  $g(\cdot)$  requires that

- (i)  $x$  should not contain a constant/intercept, and must contain at least one continuous variable. Moreover,  $\|\beta_0\|=1$ .
- (ii)  $g(\cdot)$  is differentiable and is not a constant function on the support of  $x^T \beta_0$ .
- (iii) For the discrete components of  $x$ , varying the values of the discrete variables will not divide the support of  $x^T \beta_0$  into disjoint subsets.

## Identification Condition (i)

- Note that the location and the scale of  $\beta_0$  are not identified.
- The vector  $x$  cannot include an intercept because the function  $g(\cdot)$  (which is to be estimated in nonparametric manners) includes any location and level shift.
  - That is,  $\beta_0$  cannot contain a location parameter.

## Identification Condition (i)

- Some normalization criterion (scale restrictions) for  $\beta_0$  are needed.
  - One approach is to set  $\|\beta_0\|$ .
  - The second approach is to set one component of  $\beta_0$  to equal one. This approach requires that the variable corresponding to the component set to equal one is continuously distributed and has a non-zero coefficient.
  - Then,  $x$  must be dimension 2 or larger. If  $x$  is one-dimensional, then  $\beta_0 \in \mathbb{R}^1$  is simply normalized to 1, and the model is the one-dimensional nonparametric regression  $E(Y | x) = g(x)$  with no semiparametric component.



## Identification Conditions (ii) and (iii)

- The function  $g(\cdot)$  cannot be a constant function and must be differentiable on the support of  $x^T \beta_0$ .
- $x$  must contain at least one continuously distributed variable and this continuous variable must have non-zero coefficient.
  - If not,  $x^T \beta_0$  only takes a discrete set of values and it would be impossible to identify a continuous function  $g(\cdot)$  on this discrete support.

## **Estimation: Ichimura's Method**

---

# Ichimura's Method<sup>1</sup>

- Suppose that the functional form of  $g(\cdot)$  were known.
- Then we could estimate  $\beta_0$  by minimizing the least-squares criterion:

$$\sum_{i=1} [Y_i - g(X_i^T \beta)]^2$$

with respect to  $\beta$ .

- We could think about replacing  $g(\cdot)$  with a nonparametric estimator  $\hat{g}(\cdot)$ .
- However, since  $g(z)$  is the conditional mean of  $Y_i$  given  $X_i^T \beta_0 = z$ ,  $g(\cdot)$  depends on unknown  $\beta_0$ , so we cannot estimate  $g(\cdot)$  here.

---

<sup>1</sup>Ichimura (1993).

- Nevertheless, for a fixed value of  $\beta$ , we can estimate

$$G(X_i^T \beta) := \mathbb{E}(Y_i \mid X_i^T \beta) = \mathbb{E}(g(X_i^T \beta) \mid X_i^T \beta).$$

- In general  $G(X_i^T \beta) \neq g(X_i^T \beta)$ .
- When  $\beta = \beta_0$ , it holds that  $G(X_i^T \beta_0) = g(X_i^T \beta_0)$ .

- First, we estimate  $G(X_i^T \beta)$  with the leave-one-out NW estimator:

$$\begin{aligned}\hat{G}_{-i}(X_i^T \beta) &:= \hat{\mathbb{E}}_{-i}(Y_i \mid X_i^T \beta) \\ &= \frac{\sum_{j \neq i} Y_j K\left(\frac{X_j^T \beta - X_i^T \beta}{h}\right)}{\sum_{j \neq i} K\left(\frac{X_j^T \beta - X_i^T \beta}{h}\right)}.\end{aligned}$$

# Ichimura's Method

- Second, using the leave-one-out NW estimator  $\hat{G}_{-i}(X_i^T \beta)$ , we estimate  $\beta$  with

$$\begin{aligned}\hat{\beta} &:= \arg \min_{\beta} \sum_{i=1}^n \left[ Y_i - \hat{G}_{-i}(X_i^T \beta) \right]^2 w(X_i) \mathbf{1}(X_i \in A_n) \\ &:= \arg \min_{\beta} S_n(\beta),\end{aligned}$$

which is called **Ichimura's estimator (the WSLs estimator)**.

- $w(X_i)$  is a nonnegative weight function.
- $\mathbf{1}(X_i \in A_n)$  is a trimming function to trim out small values of  $\hat{p}(X_i^T \beta) = \frac{1}{nh} \sum_{j \neq i} K \left( \frac{X_j^T \beta - X_i^T \beta}{h} \right)$ , so that we do not suffer the random denominator problem.
  - $A_\delta = \{x : p(x^T \beta) \geq \delta, \text{ for } \forall \beta \in \mathcal{B}\}$ .
  - $A_n = \{x : \|x - x^*\| \leq 2h, \text{ for } \exists x^* \in A_\delta\}$ , which shrinks to  $A_\delta$  as  $n \rightarrow \infty$  and  $h \rightarrow 0$ .

- Hardle, Hall and Ichimura (1993) suggest picking  $\beta$  and the bandwidth  $h$  jointly by minimization of  $S_n(\beta)$ .
- Further discussions on bandwidth selection follow in Section 8.4.

# Asymptotic Distribution of Ichimura's Estimator

- Let  $\hat{\beta}$  denote the semiparametric estimator of  $\beta_0$  obtained from minimizing  $S_n(\beta)$ .
- To derive the asymptotic distribution of  $\hat{\beta}$ , the following conditions are needed:



# Asymptotic Distribution of Ichimura's Estimator

## Assumption 8.1

The set  $A_\delta$  is compact, and the weight function  $w(\cdot)$  is bounded and positive on  $A_\delta$ . Define the set

$$D_z = \{z : z = x^T \beta, \beta \in \mathcal{B}, x \in A_\delta\}.$$

Letting  $p(\cdot)$  denote the PDF of  $z \in D_z$ ,  $p(\cdot)$  is bounded below by a positive constant for  $\forall z \in D_z$

## Assumption 8.2

$g(\cdot)$  and  $p(\cdot)$  are 3 times differentiable w.r.t.  $z = x^\beta$ . The third derivatives are Lipschitz continuous uniformly over  $\mathcal{B}$  for  $\forall z \in D_z$ .

# Asymptotic Distribution of Ichimura's Estimator

## Assumption 8.3

The kernel function is a bounded second order kernel, which has bounded support; is twice differentiable; and its second derivative is Lipschitz continuous.

## Assumption 8.4

$\mathbb{E}(|Y^m|) < \infty$  for  $\exists m \geq 3$ .  $\text{var}(Y | x)$  is bounded and bounded away from zero for  $\forall x \in A_\delta$ .  $\frac{q \ln(h)}{nh^{3+\frac{3}{m-1}}} \rightarrow 0$  and  $nh^8 \rightarrow 0$  as  $n \rightarrow \infty$ .

# Asymptotic Distribution of Ichimura's Estimator

**Theorem 8.1.** Under assumptions 8.1 through 8.4,

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

with

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

$$V = \mathbb{E}\{w(X_i)(g_i^{(1)})^2 \\ \times (X_i - E_A(X_i | X_i^T \beta_0))(X_i - E_A(X_i | X_i^T \beta_0))^T\},$$

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

where

- $(g_i^{(1)}) = \frac{\partial g(v)}{\partial v} \big|_{v=X_i^T \beta_0},$
- $\mathbb{E}_A(X_i | v) = \mathbb{E}(X_i | x_A^T \beta_0 = v),$
- $x_A$  has the distribution of  $X_i$  conditional on  $X_i \in A_\delta.$

# Asymptotic Distribution of Ichimura's Estimator

- See Ichimura (1993); and Hardle, Hall and Ichimura (1993) for the proof of **Theorem 8.1**.
  - Horowitz (1998) provides an excellent heuristic outline for proving **Theorem 8.1**, using only the familiar Taylor series methods, the standard LLN, and the Lindeberg-Levy CLT.
- We can consistently estimate  $\text{Avar} \left( \sqrt{n}(\hat{\beta} - \beta) \right) = \Omega_I$  with its sample analogue.

# Optimal Weight under the Homoscedasticity Assumption

- Here we introduce the following homoscedasticity assumption:  $\mathbb{E}(u_i^2 \mid X_i) = \sigma^2$ .
- Under this assumption, the optimal choice of  $w(\cdot)$  is  $w(X_i) = 1$ .
- In this case,

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n (Y_i - \hat{G}_{-i}(X_i^T \beta)^2) \mathbf{1}(X_i \in A_n)$$

is **semiparametrically efficient** in the sense that  $\Omega_I$  is **the semiparametric variance lower bound** (conditional on  $X \in A_\delta$ ).

# Optimal Weight under Heteroscedasticity

- In general,  $\mathbb{E}(u_i^2 \mid X_i) = \sigma^2(X_i)$ .
  - The semiparametrically efficient estimator  $\hat{\beta}_{opt}$  has a complicated structure.
- **An infeasible case:** If one assumes that  $\mathbb{E}(u_i^2 \mid X_i) = \sigma^2(X_i^T \beta_0)$ , that is, the conditional variance depends only on the single index  $X_i^T \beta_0$ , the choice of  $w(X_i) = \frac{1}{\sigma^2 X_i^T \beta_0}$  can lead to a semiparametrically efficient estimation.

# Direct Semiparametric Estimators for $\beta$

---

# Bandwidth Selection

---



## **Klein and Spady (1993)**

---

**Lewbel (2000)**

---

## Manski's (1975) Maximum Score Estimator

---

## **Horowitz's (1992) Smoothed Maximum Score Estimator**

---

## **Han's (1987) Maximum Rank Estimator**

---

# Multinomial Discrete Choice Models

---

## **Ai's (1997) Semiparametric Maximum Likelihood Approach**

---

## References

---



## References (1)

- Hardle, W, P. Hall and H. Ichimura (1993) “Optimal Smoothing in Single-Index Models,” *Annals of Statistics*, 21, 157-178.
- Horowitz, J.L. (1998) *Semiparametric Methods in Econometrics*, Springer.
- Ichimura, H. (1993) “Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models,” *Journal of Econometrics*, 3, 205-228.
- Li, Q. and J. S. Racine, (2007). *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.
- 末石直也 (2024) 『データ駆動型回帰分析：計量経済学と機械学習の融合』日本評論社.
- 西山慶彦, 人見光太郎 (2023) 『ノン・セミパラメトリック統計解析 (理論統計学教程：数理統計の枠組み)』共立出版.

## References (2)

Useful references also include some lecture notes of the following topic courses:

- ECON 718 NonParametric Econometrics (Bruce Hansen, Spring 2009, University of Wisconsin-Madison),
- セミノンパラメトリック計量分析（末石直也，2014 年度後期，京都大学）.