

Semiparametric Single Index Models

Yasuyuki Matsumura

Graduate School of Economics, Kyoto University

November 4, 2024

- 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.

- 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 $\text{Normal}(0, 1)$, 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.

- 1 Identification Conditions
- 2 Estimation: Ichimura (1993)
- 3 Direct Semiparametric Estimators for β
- 4 Bandwidth Selection
- 5 Klein and Spady (1993)
- 6 Lewbel (2000)
- 7 Manski's (1975) Maximum Score Estimator
- 8 Horowitz's (1992) Smoothed Maximum Score Estimator
- 9 Han's (1987) Maximum Rank Estimator
- 10 Multinomial Discrete Choice Models
- 11 Ai's (1997) Semiparametric Maximum Likelihood Approach
- 12 References

- 1 Identification Conditions
- 2 Estimation: Ichimura (1993)
- 3 Direct Semiparametric Estimators for β
- 4 Bandwidth Selection
- 5 Klein and Spady (1993)
- 6 Lewbel (2000)
- 7 Manski's (1975) Maximum Score Estimator
- 8 Horowitz's (1992) Smoothed Maximum Score Estimator
- 9 Han's (1987) Maximum Rank Estimator
- 10 Multinomial Discrete Choice Models
- 11 Ai's (1997) Semiparametric Maximum Likelihood Approach
- 12 References

Proposition 8.1 (Identification of a Single Index Model)

For the semiparametric single index model $Y = g(x^T \beta_0) + u$, identification of β_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 β_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, β_0 cannot contain a location parameter.
- Some normalization criterion (scale restrictions) for β_0 are needed.
 - One approach is to set $\|\beta_0\|$.
 - The second approach is to set one component of β_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.

- 1 Identification Conditions
- 2 Estimation: Ichimura (1993)
- 3 Direct Semiparametric Estimators for β
- 4 Bandwidth Selection
- 5 Klein and Spady (1993)
- 6 Lewbel (2000)
- 7 Manski's (1975) Maximum Score Estimator
- 8 Horowitz's (1992) Smoothed Maximum Score Estimator
- 9 Han's (1987) Maximum Rank Estimator
- 10 Multinomial Discrete Choice Models
- 11 Ai's (1997) Semiparametric Maximum Likelihood Approach
- 12 References

- 1 Identification Conditions
- 2 Estimation: Ichimura (1993)
- 3 Direct Semiparametric Estimators for β**
- 4 Bandwidth Selection
- 5 Klein and Spady (1993)
- 6 Lewbel (2000)
- 7 Manski's (1975) Maximum Score Estimator
- 8 Horowitz's (1992) Smoothed Maximum Score Estimator
- 9 Han's (1987) Maximum Rank Estimator
- 10 Multinomial Discrete Choice Models
- 11 Ai's (1997) Semiparametric Maximum Likelihood Approach
- 12 References

- 1 Identification Conditions
- 2 Estimation: Ichimura (1993)
- 3 Direct Semiparametric Estimators for β
- 4 Bandwidth Selection**
- 5 Klein and Spady (1993)
- 6 Lewbel (2000)
- 7 Manski's (1975) Maximum Score Estimator
- 8 Horowitz's (1992) Smoothed Maximum Score Estimator
- 9 Han's (1987) Maximum Rank Estimator
- 10 Multinomial Discrete Choice Models
- 11 Ai's (1997) Semiparametric Maximum Likelihood Approach
- 12 References

- 1 Identification Conditions
- 2 Estimation: Ichimura (1993)
- 3 Direct Semiparametric Estimators for β
- 4 Bandwidth Selection
- 5 Klein and Spady (1993)**
- 6 Lewbel (2000)
- 7 Manski's (1975) Maximum Score Estimator
- 8 Horowitz's (1992) Smoothed Maximum Score Estimator
- 9 Han's (1987) Maximum Rank Estimator
- 10 Multinomial Discrete Choice Models
- 11 Ai's (1997) Semiparametric Maximum Likelihood Approach
- 12 References

- 1 Identification Conditions
- 2 Estimation: Ichimura (1993)
- 3 Direct Semiparametric Estimators for β
- 4 Bandwidth Selection
- 5 Klein and Spady (1993)
- 6 Lewbel (2000)**
- 7 Manski's (1975) Maximum Score Estimator
- 8 Horowitz's (1992) Smoothed Maximum Score Estimator
- 9 Han's (1987) Maximum Rank Estimator
- 10 Multinomial Discrete Choice Models
- 11 Ai's (1997) Semiparametric Maximum Likelihood Approach
- 12 References

- 1 Identification Conditions
- 2 Estimation: Ichimura (1993)
- 3 Direct Semiparametric Estimators for β
- 4 Bandwidth Selection
- 5 Klein and Spady (1993)
- 6 Lewbel (2000)
- 7 Manski's (1975) Maximum Score Estimator**
- 8 Horowitz's (1992) Smoothed Maximum Score Estimator
- 9 Han's (1987) Maximum Rank Estimator
- 10 Multinomial Discrete Choice Models
- 11 Ai's (1997) Semiparametric Maximum Likelihood Approach
- 12 References

- 1 Identification Conditions
- 2 Estimation: Ichimura (1993)
- 3 Direct Semiparametric Estimators for β
- 4 Bandwidth Selection
- 5 Klein and Spady (1993)
- 6 Lewbel (2000)
- 7 Manski's (1975) Maximum Score Estimator
- 8 Horowitz's (1992) Smoothed Maximum Score Estimator**
- 9 Han's (1987) Maximum Rank Estimator
- 10 Multinomial Discrete Choice Models
- 11 Ai's (1997) Semiparametric Maximum Likelihood Approach
- 12 References

- 1 Identification Conditions
- 2 Estimation: Ichimura (1993)
- 3 Direct Semiparametric Estimators for β
- 4 Bandwidth Selection
- 5 Klein and Spady (1993)
- 6 Lewbel (2000)
- 7 Manski's (1975) Maximum Score Estimator
- 8 Horowitz's (1992) Smoothed Maximum Score Estimator
- 9 Han's (1987) Maximum Rank Estimator
- 10 Multinomial Discrete Choice Models
- 11 Ai's (1997) Semiparametric Maximum Likelihood Approach
- 12 References

- 1 Identification Conditions
- 2 Estimation: Ichimura (1993)
- 3 Direct Semiparametric Estimators for β
- 4 Bandwidth Selection
- 5 Klein and Spady (1993)
- 6 Lewbel (2000)
- 7 Manski's (1975) Maximum Score Estimator
- 8 Horowitz's (1992) Smoothed Maximum Score Estimator
- 9 Han's (1987) Maximum Rank Estimator
- 10 Multinomial Discrete Choice Models**
- 11 Ai's (1997) Semiparametric Maximum Likelihood Approach
- 12 References

- 1 Identification Conditions
- 2 Estimation: Ichimura (1993)
- 3 Direct Semiparametric Estimators for β
- 4 Bandwidth Selection
- 5 Klein and Spady (1993)
- 6 Lewbel (2000)
- 7 Manski's (1975) Maximum Score Estimator
- 8 Horowitz's (1992) Smoothed Maximum Score Estimator
- 9 Han's (1987) Maximum Rank Estimator
- 10 Multinomial Discrete Choice Models
- 11 Ai's (1997) Semiparametric Maximum Likelihood Approach**
- 12 References

- 1 Identification Conditions
- 2 Estimation: Ichimura (1993)
- 3 Direct Semiparametric Estimators for β
- 4 Bandwidth Selection
- 5 Klein and Spady (1993)
- 6 Lewbel (2000)
- 7 Manski's (1975) Maximum Score Estimator
- 8 Horowitz's (1992) Smoothed Maximum Score Estimator
- 9 Han's (1987) Maximum Rank Estimator
- 10 Multinomial Discrete Choice Models
- 11 Ai's (1997) Semiparametric Maximum Likelihood Approach
- 12 References

References (1)

- 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）.