Nonparametric IV estimation of local average treatment effects with covariates

Markus Frölich

University College London, UK
Universität St.Gallen, Bodanstrasse 8, SIAW, 9000 St. Gallen, Switzerland

June 7, 2022

Content

ELSEVIER

Journal of Econometrics

Volume 139, Issue 1, July 2007, Pages 35-75

Nonparametric IV estimation of local average treatment effects with covariates

Markus Frölich [a, b] [1]

Show more

+ Add to Mendeley   Share   Cite

Introduction

- In many applications one wants to uncover the causal relationship between a variable D and an outcome Y, where the variable D is itself endogenous.
- If a variable Z exists that affects only D but not Y, then an exogenous variation in Z induces an exogenous variation in D and thus overcomes the endogeneity of D. Such a variable Z is called an instrumental variable.

Introduction

- The LATE has been introduced by Imbens and Angrist (1994) and further been analysed in several papers. Most of the discussion on LATE focuses on the case where Z is a proper instrumental variable without conditioning on X.
- Nonparametric estimation of LATE with confounding covariates X has not been analysed until this paper was published.

Introduction

- In this paper, nonparametric identification and estimation of LATE with covariates is analysed.
- A nonparametric estimator in the form of a ratio of two matching estimators is shown to be $\sqrt{n}$-consistnet and efficient.

Model

Y	outcome variable
D	treatment variable
Z	instrument variable
X	covariates, dim(X)=k
$Y_{i,Z_i}^{D_i}$	the observed value of Y_i
$Y_{i,z}^{d}$	the potential outcome of Y_i, usually $z \in \{0,1\}, d \in \{0,1\}$
D_{i,Z_i}	the observed value of D_i
$D_{i,z}$	the potential outcome of D_i, usually $z \in \{0,1\}$

Model

According to the reaction of D on an external intervention on Z, the units i can be distinguished into different types. With D and Z binary, four types $\tau \in \{n, c, d, a\}$ are possible:

τ_i=n	if $D_{i,0} = 0$ and $D_{i,1} = 0$	Never-taker
τ_i=c	if $D_{i,0} = 0$ and $D_{i,1} = 1$	Complier
τ_i=d	if $D_{i,0} = 1$ and $D_{i,1} = 0$	Defier
τ_i=a	if $D_{i,0} = 1$ and $D_{i,1} = 1$	Always-taker

The impact of D on Y can at most be ascertained for the subpopulations of compliers and defiers.

Model

Under certain assumptions given below, the LATE γ for the subpopulation of compliers

$$\gamma = \mathbf{E}[\, Y^1 - \, Y^0 \mid \tau = c]$$

is identified.

Define also the treatment effect for the compliers with characteristics x as

$$\gamma(x) = \mathbf{E}[\, Y^1 - \, Y^0 \mid X = x, \tau = c].$$

Model

To identify the LATE γ, we assume that:

Assumption 1(Monotonicity). The subpopulation of defiers has probability measure zero: $\mathbf{P}(\mathbf{D}_{i,0} > \mathbf{D}_{i,1}) = 0$.

Assumption 2(Existence of compliers). The subpopulation of compliers has positive probability: $\mathbf{P}(\mathbf{D}_{i,0} < \mathbf{D}_{i,1}) > 0$.

Assumption 3(Unconfounded type). The relative size of the subpopulations always-takers, never-takers and compliers is independent of the instrument:for all $x \in \mathbf{Supp}(X)$

$$\mathbf{P}(\tau_i = t \mid X_i = x, Z_i = 0) = \mathbf{P}(\tau_i = t \mid X_i = x, Z_i = z) \text{ for } t \in \{a, n, c\}.$$

Assumption 4(Mean exclusion restriction). The potential outcomes are mean independent of the instrumental variable Z in each subpopulation: for all $x \in \mathbf{Supp}(X)$

$\mathrm{E}\left[Y^0_{i,Z_i} \mid X_i = x, Z_i = 0, \tau_i = t\right] = \mathrm{E}\left[Y^0_{i,Z_i} \mid X_i = x, Z_i = 1, \tau_i = t\right]$ for $t \in \{n, c\}$,

$\mathrm{E}\left[Y^1_{i,Z_i} \mid X_i = x, Z_i = 0, \tau_i = t\right] = \mathrm{E}\left[Y^1_{i,Z_i} \mid X_i = x, Z_i = 1, \tau_i = t\right]$ for $t \in \{a, c\}$.

Assumption 5(Common support). The support of X is identical in both subpopulations:

$$\mathbf{Supp}(X \mid Z = 1) = \mathbf{Supp}(X \mid Z = 0).$$

Model

Theorem 1

(Identification of LATE). Under Assumptions 1-5 and supposing that $\mathbf{E}[Y] < \infty$, the LATE γ for the subpopulation of compliers is nonparametrically identified as

$$\gamma = \mathrm{E}\left[Y^1 - Y^0 \mid \tau = c\right] = \frac{\int (\mathrm{E}[Y \mid X = x, Z = 1] - \mathrm{E}[Y \mid X = x, Z = 0]) f_x(x) \mathrm{d}x}{\int (\mathrm{E}[D \mid X = x, Z = 1] - \mathrm{E}[D \mid X = x, Z = 0]) f_x(x) \mathrm{d}x}. \quad (1)$$

And that for all x with $\mathbf{P}(\tau = c \mid X = x) > 0$ the treatment effect $\gamma(x)$ is identified as

$$\gamma(x) = \mathrm{E}\left[Y^1_{i,Z_i} - Y^0_{i,Z_i} \mid X_i = x, \tau_i = c\right] = \frac{\mathrm{E}[Y \mid X = x, Z = 1] - \mathrm{E}[Y \mid X = x, Z = 0]}{\mathrm{E}[D \mid X = x, Z = 1] - \mathrm{E}[D \mid X = x, Z = 0]}. \quad (2)$$

Estimation

In this section nonparametric estimation of the LATE γ is discussed. Define the conditional mean functions

$$m_z(x) = \mathbf{E}[\,Y \mid X = x, Z = z] \text{ and } \mu_z(x) = \mathbf{E}[D \mid X = x, Z = z],$$

and let $\hat{m}_z(x)$ and $\hat{\mu}_z(x)$ be the corresponding nonparametric regression estimators. A nonparametric imputation estimator of γ is

$$\frac{\sum_i (\hat{m}_1(X_i) - \hat{m}_0(X_i))}{\sum_i (\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i))},$$

where the expected values $\mathbf{E}[\,Y \mid X, Z]$ and $\mathbf{E}[D \mid X, Z]$ are imputed for each observation X_i.

Estimation

Using the observed values Y_i and D_i as estimates of

$$\mathbf{E}[\,Y_i \mid X_i, Z = z] \text{ and } \mathbf{E}[D_i \mid X_i, Z = z],$$

whenever $z = Z_i$.

Gives LATE estimator $\hat{\gamma}$ as

$$\hat{\gamma} = \frac{\sum_{i:Z_i=1} (\,Y_i - \hat{m}_0\,(X_i)) - \sum_{i:Z_i=0} (\,Y_i - \hat{m}_1\,(X_i))}{\sum_{i:Z_i=1} (D_i - \hat{\mu}_0\,(X_i)) - \sum_{i:Z_i=0} (D_i - \hat{\mu}_1\,(X_i))}. \tag{3}$$

The estimator $\hat{\gamma}$ corresponds to a ratio of two matching estimators.

Estimation

To analyse the properties of $\hat{\gamma}$, it is useful to first derive the semiparametric efficiency bound for the estimation of the LATE γ.

Theorem 2

(Efficiency bound). The semiparametric variance bound for γ is

$$\begin{aligned}\mathscr{V} =& \frac{1}{\Gamma^2}\mathrm{E}\left[\frac{\sigma^2_{Y_1}(X) - 2\gamma\sigma^2_{Y_1 D_1}(X) + \gamma^2\sigma^2_{D_1}(X)}{\pi(X)} + \frac{\sigma^2_{Y_0}(X) - 2\gamma\sigma^2_{Y_0 D_0}(X) + \gamma^2\sigma^2_{D_0}(X)}{1-\pi(X)}\right] \\ &+ \frac{1}{\Gamma^2}\mathrm{E}\left[(m_1(X) - m_0(X) - \gamma\mu_1(X) + \gamma\mu_0(X))^2\right],\end{aligned} \tag{4}$$

where $\Gamma = \int(\mu_1(x) - \mu_0(x))f_x(x)dx$ is the denominator in (1).

Estimation

$\sigma^2_{Y_z}(x), \sigma^2_{D_z}(x)and\sigma^2_{Y_z D_z}(x)$are the conditional variances and covariances in the $Z = z$ subpopulation:

$$\sigma^2_{Y_z}(x) = \mathbf{Var}\left[\, Y \mid X = x, Z = 1\right],$$

$$\sigma^2_{Y_z D_z}(x) = \mathbf{Cov}\left[\, Y, D \mid X = x, Z = z\right].$$

Asymptotic Results

Theorem 3

(Asymptotic normality of γ). Suppose that

(i)the LATE γ is identified,

(ii)$\{(Y_i, D_i, Z_i, X_i)\}_{i=1}^n$are iid, $Z_i \in \{0,1\}$ and $X_i \in \mathscr{R}^k$,

(iii)$\mathbf{E}\left[Y^2\right] < \infty$,

(iv) the nonparametric estimator $\hat{m}_1$ of $m_1(x) = \mathbf{E}[Y \mid X = x, Z = 1]$ is asymptotically linear

$$\hat{m}_1(x) - m_1(x) = \frac{1}{n_1} \sum_{j: Z_j = 1} \xi_1^m(Y_j, X_j, x) + b_1^m(x) + R_1^m(x),$$

where $n_z = \sum_{i=1}^n 1(Z_i = z)$, with the properties:

Asymptotic Results

Theorem 3

(A)$\mathbf{E}\left[\xi_1^m\left(Y_j, X_j, x\right) \mid X=x, Z_j=1\right]=0$;

(B)$\mathbf{E}\left[\xi_1^m\left(Y_j, X_j, x\right)^2 \mid Z_j=1, Z_i=0\right]=o(n)$;

(C)$\frac{1}{\sqrt{n_0}} \sum_{i: Z_i=0} b_1^m(X_i)=o_p(1)$;

(D)$\frac{1}{\sqrt{n_0}} \sum_{i: Z_i=0} R_1^m(X_i)=o_p(1)$;

(E)$\mathrm{E}\left[\xi_1^m\left(Y_j, X_j, X_i\right) \mid Y_j, X_j, Z_j=1, Z_i=0\right]=\left(Y_j-m_1\left(X_j\right)\right) \frac{f_{x \mid z=0}\left(X_j\right)}{f_{x \mid z=1}\left(X_j\right)}+\mathrm{o_p}(1)$,

and analogously for $\hat{m}_0, \hat{\mu}_1, \hat{\mu}_0$. Then the estimator (3) of the LATE γ is asymptotically normally distributed

$$\sqrt{n}(\hat{\gamma}-\gamma) \rightarrow N(0, \mathscr{V}), \tag{5}$$

where $\mathscr{V}$ is given by (4).

Asymptotic Results

Theorem4 shows that local polynomial regression satisfies the conditions of **Theorem3**.

Theorem 4

(Efficiency of local polynomial regression). Consider the assumptions:

(1)$m_1(x), m_0(x), \mu_1(x)$ and $\mu_0(x)$ are $\bar{p}$-times continuously differentiable with $\bar{p}$th derivative Hölder continuous, where $\bar{p} > k$ and $k = \dim(X)$,

(2)the bandwidth sequence h_{n_1}satisfies $n_1 h_{n_1}^k / \ln n_1 \to \infty$ and $n_1 h_{n_1}^{2\bar{p}} \to 0$,

(3)the Kernel function K is symmetric, commpact and Lipschitz continuous,

(4)the density $f_{x|z=1}$ of X is $\bar{p}$-times continuously differentiable with its $\bar{p}$th derivative Hölder continuous,

(5a)the Kernel function K has moments of order 1 through $\bar{p}$ equal to zero,
(5b)the Kernel function K has moments of order 1 through $\bar{p}-1$ equal to zero,

(6)$m_1(x), m_0(x), \mu_1(x)$ and $\mu_0(x)$ are estimated at an interior point of $\mathbf{Supp}(X \mid Z = 1)$,

Asymptotic Results

Theorem 4a.

If Assumptions 1-4 and 5a are satisfied, local polynomial regression of order $\bar{p}$ satisfies the conditions of **Theorem3**.

Theorem 4b.

If Assumptions 1-4 and 5b and 6 are satisfied, local polynomial regression of order $0 \leq p < \bar{p}$ satisfies the conditions of **Theorem3**.

LATE with propensity score

There are two alternative approaches to matching on X: matching on the propensity score and weighting by the propensity score.

Propensity scored based estimators can also be constructed for the estimation of γ, with $\pi(x) = \mathbf{P}(Z = 1 \mid X = x)$ as the propensity score. By noting that

$$\mathbf{E}\left[YZ/\pi(x)\right] = \int (1/\pi(x))\mathbf{E}\left[YZ \mid X = x\right] f_x(x)\, dx = \int m_1(x) f_x(x)\, dx.$$

LATE with propensity score

The LATE can be written as

$$\gamma = \frac{\mathbf{E}\left[YZ/\pi(x) - Y(1-Z)/(1-\pi(x))\right]}{\mathbf{E}\left[DZ/\pi(x) - D(1-Z)/(1-\pi(x))\right]}, \tag{6}$$

and estimated by the propensity score weighting estimator

$$\hat{\gamma}_{\pi w} = \frac{\sum_i (Y_i Z_i/\pi(X_i) - Y_i(1-Z_i)/(1-\pi(X_i)))}{\sum_i (D_i Z_i/\pi(X_i) - D_i(1-Z_i)/(1-\pi(X_i)))}. \tag{7}$$

LATE with propensity score

To derive the propensity score matching estimator, note that the LATE can also be written as

$$\gamma = \frac{\int (m_{\pi 1}(\rho) - m_{\pi 0)}(\rho)) \cdot f_\pi(\rho) d\rho}{\int (\mu_{\pi 1}(\rho) - \mu_{\pi 0)}(\rho)) \cdot f_\pi(\rho) d\rho} \tag{8}$$

where

$m_{\pi 1}(\rho) = \mathbf{E}\left[Y \mid \pi(X) = \rho, Z = 1 \right]$ and $\mu_{\pi 1}(\rho) = \mathbf{E}\left[D \mid \pi(X) = \rho, Z = 1 \right]$,

f_π is the density function of $\pi(x)$ in the population.

LATE with the propensity score

The propensity score matching estimator is

$$\hat{\gamma}_{\pi m} = \frac{\sum_i (\hat{m}_{\pi 1}(\pi_i) - \hat{m}_{\pi 0}(\pi_i))}{\sum_i (\hat{\mu}_{\pi 1}(\pi_i) - \hat{\mu}_{\pi 0}(\pi_i))}, \tag{9}$$

where $\pi_i = \pi(X_i)$.

LATE with the propensity score

Theorem 5

(Efficiency bound with knowledge of propensity score). Suppose that the propensity score $\pi(x)$ is known. The semiparametric variance bound for γ is $\mathscr{V}$, where $\mathscr{V}$ is given by (4).

LATE with propensity score

For the propensity score matching estimator, to acess efficiency consider first the situation where the propensity score $\pi(x)$ is known. Provided that the conditions of **Theorem3** is satisfied with X replaced by $\pi(x)$, the asymptotic variance of propensity score matching with respect to the known propensity score π is

$$V_{\pi m} = \frac{1}{\Gamma^2} \mathrm{E}\left[\frac{\sigma^2_{\pi\, Y_1}(\pi) - 2\gamma\sigma^2_{\pi\, Y_1 D_1}(\pi) + \gamma^2\sigma^2_{\pi D_1}(\pi)}{\pi} + \frac{\sigma^2_{\pi\, Y_0}(\pi) - 2\gamma\sigma^2_{\pi\, Y_0 D_0}(\pi) + \gamma^2\sigma^2_{\pi D_0}(\pi)}{1-\pi} \right] + \frac{1}{\Gamma^2} \mathrm{E}\left[\left(m_{\pi 1}(\pi) - m_{\pi 0}(\pi) - \gamma\mu_{\pi 1}(\pi) + \gamma\mu_{\pi 0}(\pi) \right)^2 \right], \tag{10}$$

where

$$\sigma^2_{\pi\, Y_1}(\rho) = \mathbf{Var}\left[Y \mid \pi(X) = \rho, Z = 1 \right],$$

$$\sigma^2_{\pi\, Y_1 D_1}(\rho) = \mathbf{Cov}\left[Y, D \mid \pi(X) = \rho, Z = 1 \right].$$

LATE with propensity score

Theorem 6

(Inefficiency of propensity score matching). The difference between the asymptotic variance when matching on the known propensity score and the asymptotic variance when matching on X is non-negative and given by

$$
\begin{aligned}
V_{\pi m}-\mathscr{V}= & \frac{1}{\Gamma^2} \mathrm{E}\left[\frac{1-\pi}{\pi} \operatorname{Var}\left(m_1(X)-\gamma \mu_1(X) \mid \pi\right)+\frac{\pi}{1-\pi} \operatorname{Var}\left(m_0(X)-\gamma \mu_0(X) \mid \pi\right)\right] \\
& +\frac{2}{\Gamma^2} \mathrm{E}\left[\operatorname{Cov}\left(m_1(X)-\gamma \mu_1(X), m_0(X)-\gamma \mu_0(X) \mid \pi\right)\right] \geqslant 0 .
\end{aligned}
\tag{11}
$$

LATE with propensity score

Generally, $V_{\pi m} - \mathscr{V}$ is strictly positive unless the support of the propensity score $\pi(x)$ contains only values where both variances are zero or where

$$\sqrt{(\mathrm{Var}\left(m_1(X) - \gamma\mu_1(X) \mid \pi)\right)/(\mathrm{Var}\left(m_0(X) - \gamma\mu_0(X) \mid \pi)\right)}$$
$$- \pi(1 + \sqrt{(\mathrm{Var}\left(m_1(X) - \gamma\mu_1(X) \mid \pi)\right)/(\mathrm{Var}\left(m_0(X) - \gamma\mu_0(X) \mid \pi)\right)}) = 0. \tag{12}$$