Econometrics 2

Sergey Lychagin

Central European University

Winter 2020

Types of Dependent Variables

Continuous A continuous variable is one which can take on infinitely many, uncountable values.

Discrete The number of permitted values is either finite or countably infinite.

- Binary: $Y \in \{0, 1\}$ *qualitative* interpretation, e.g. go to college, take a job, buy a product ...
- Multinomial: outcome of a choice among several alternatives, e.g. college degrees, means of transportation, ...
- Ordered: ordered alternatives, e.g. educational degree, measure of subjective wellbeing, income brackets
- Count: number of telephone calls, number of rooms...

Censored or Truncated observe wage only if person employed, e.g observe wage up to an upper limit

Duration Duration of (un)employment

Binary Response Models

Y is a binary variable. We are interested in the effect of X on Y.
Example: effect of education on the decision to work

1. Linear Probability Model: use linear regression with Y as dependent variable
2. Non-Linear Probability Models: Logit, Probit
3. Estimation of non-linear models
4. Parameter interpretation in non-linear models

Linear Probability Model

$$\begin{aligned} Y_i &= X_i\beta + \varepsilon_i \\ E[Y_i|X_i] &= X_i\beta \\ E[Y_i|X_i] &= P(Y_i = 1|X_i) * 1 + P(Y_i = 0|X_i) * 0 = P(Y_i = 1|X_i) \\ P(Y_i = 1|X_i) &= X_i\beta \end{aligned}$$

Model the probability of working conditional on education

- The distribution of ε_i given X_i can be summarized as:

$$\begin{aligned} 1 - P(Y_i = 1|X_i) &= 1 - X_i\beta \text{ if } Y_i = 1 \\ 0 - P(Y_i = 0|X_i) &= -X_i\beta \text{ if } Y_i = 0 \\ Var(\varepsilon_i|X_i) &= (1 - X_i\beta)^2 X_i\beta + (0 - X_i\beta)^2(1 - X_i\beta) \\ &= X_i\beta(1 - X_i\beta) \end{aligned}$$

Heteroskedastic standard errors, use *robust standard errors.*

Binary outcome and Linear Probability Model

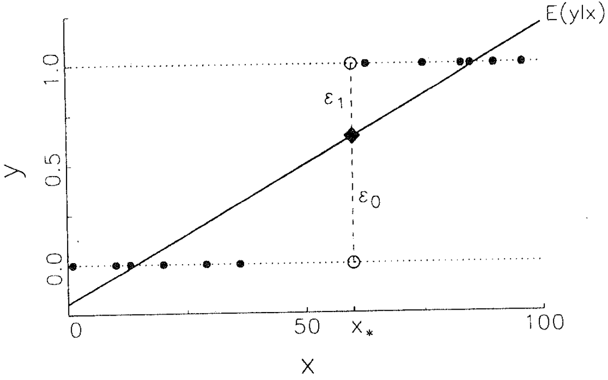


Figure 3.1. Linear Probability Model for a Single Independent Variable

Non-linear Probability Models

LPM is easy to estimate and interpret. However, predicted probabilities may be non-sensical (< 0 or > 1).

Solution: approximate $P(Y_i = 1|X_i)$ with some c.d.f.

$$P(Y_i = 1|X_i) = F(X_i\beta)$$

$F()$ is a cumulative distribution function, $F \in [0, 1]$ by definition.

- $F = \Lambda$ logistic distribution c.d.f *logit model*
- $F = \Phi$ normal distribution c.d.f *probit model*

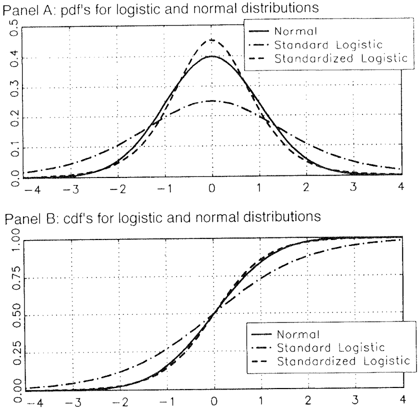


Figure 3.3. Normal and Logistic Distributions

Latent Variable Model

$$Y_i^* = X_i\beta + \varepsilon_i$$

Y_i^* is continuous unobserved variable, "utility" from working.
We only observe the decision whether the individual is working

$$Y_i = \begin{cases} 1 & \text{if } Y_i^* > 0 \\ 0 & \text{if } Y_i^* \leq 0 \end{cases}$$

Model

$$\begin{aligned} P(Y_i = 1|X_i) = P(Y_i^* > 0|X_i) &= P(X_i\beta + \varepsilon_i > 0|X_i) \\ &= P(\varepsilon_i > -X_i\beta|X_i) \end{aligned}$$

Probit as Latent Variable Model

Assumption about the distribution of ε_i:

$$\begin{aligned}
\varepsilon_i | X_i &\sim N[0,1] \\
P(Y_i = 1 | X_i) &= P(\varepsilon_i > -X_i\beta | X_i) = \\
&= 1 - \Phi(-X_i\beta) = \Phi(X_i\beta) \quad \text{Probit model}
\end{aligned}$$

- Estimated parameters in the Probit model β
- Fixed error variance $\sigma_\varepsilon^2 = 1$, less flexible than OLS.

Estimation of Binary Response Models

linear model $\min S(\beta) = \sum (Y_i - X_i\beta)^2$
We can solve the least squares problem analytically

non-linear model $\min S(\beta) = \sum (Y_i - F(X_i\beta))^2$
Non-linear least squares

Maximum likelihood joint probability distribution of the data is treated as function of the unknown parameters

$$\begin{aligned} L(\beta) &= P(\text{observe}(y_1, x_1)\text{and}(y_2, x_2)\text{and}...(y_n, x_n)|\beta) \\ \hat{\beta}_{ML} &= \max_{\beta} L(\beta) \end{aligned}$$

Chooses the most plausible parameter $\hat{\beta}_{ML}$ given the data

Parameter Interpretation

Compare 2 models

linear model

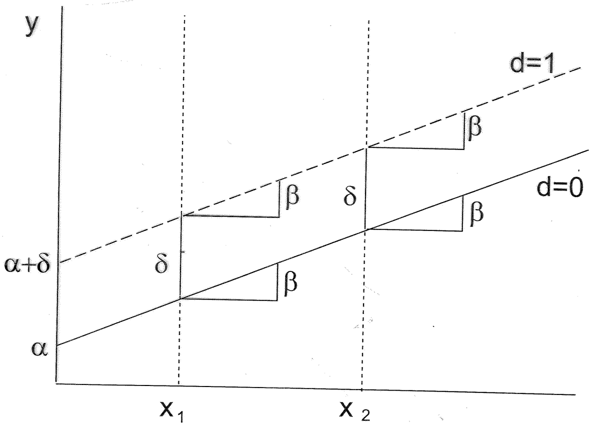
$$Y_i = \alpha + X_i\beta + \delta D_i$$

non-linear model

$$Y_i = F(\alpha + X_i\beta + \delta D_i)$$

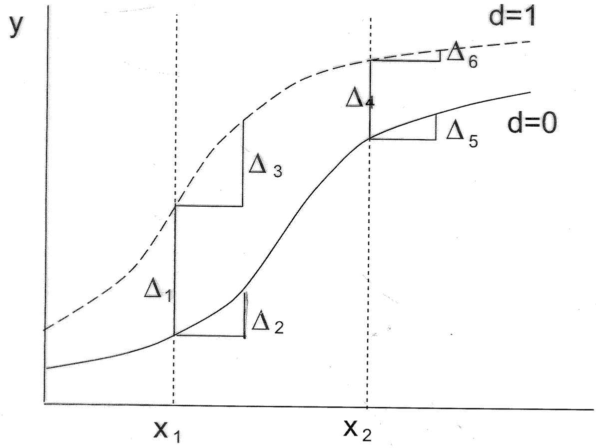
Linear Model

Figure 6: Linear Model



Non-linear Model

Figure 7: Nonlinear Model



Partial Change, Marginal Effect

- $Y = \alpha + X\beta + \delta D$, $\frac{\partial Y}{\partial X} = \beta$
- Partial effect in the non-linear model

$$\frac{\partial P(Y_i = 1|X)}{\partial X} = \frac{\partial F(X\beta)}{\partial X} * \frac{\partial X\beta}{\partial X} = f(X\beta)\beta$$

- Marginal effect
  - mean $\left(\frac{\partial P(Y=1|X_i)}{\partial X_i}\right)$
  - $\frac{\partial P(Y=1|X_i)}{\partial X_i}\Big|_{X_i=\bar{X}}$
- Discrete change (effect of dummy variable D)
  - $\frac{\Delta P(Y=1|D)}{X,\Delta D} = P(Y_i = 1|X, D = 1) - P(Y_i = 1|X, D = 0)$

Parameter Interpretation

- In the non-linear model the effect of X_i or D_i on Y_i is not constant across different values of X_i and D_i
- We can always interpret the sign and significance level of β and δ
- Scale of estimated parameters. Probability distributions are maximized at zero.
    - Logit: $f(0) = \frac{exp(0)}{[1+exp(0)]^2} = 0.25$
    - Probit: $f(0) = \frac{1}{\sqrt{2\pi}} \approx 0.4$
    - $\beta_l \approx 1.6\beta_p$
- Investigate predicted probabilities: plots, $\min \hat{p}(Y_i = 1|X_i)$, $\max \hat{p}(Y_i = 1|X_i)$

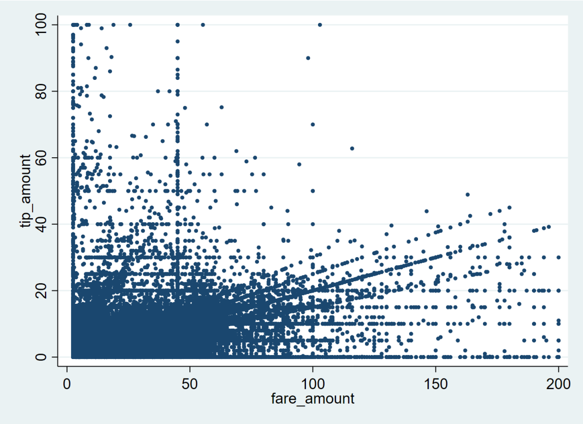
Other questions to consider

1. What if ε_i in the latent utility is $N[0, \sigma]$, $\sigma \neq 1$? Can we identify β and σ?
2. Heteroskedasticity: what if ε_i in the latent utility is $N[0, \sigma(X)]$?

Tobit

Suppose the outcome of interest has a corner solution.

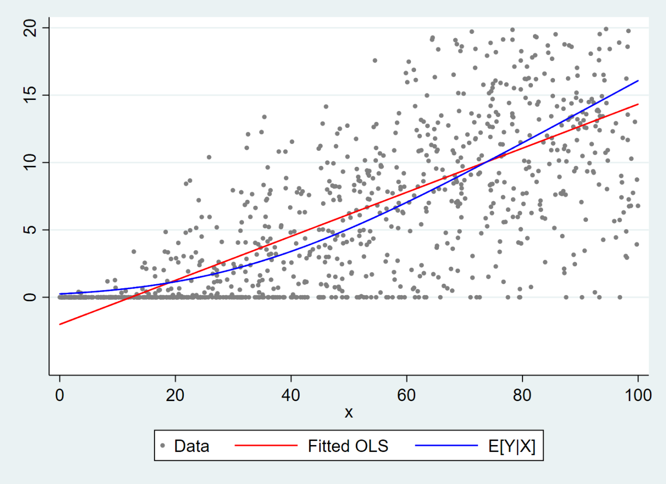
Tipping cab drivers in NYC:



Many zeros; some passengers don't tip.

True relationship vs what OLS estimates

How does X (total fare) affect expected tip, $E[Y|X]$?



True relationship vs what OLS estimates

What's wrong with OLS here?

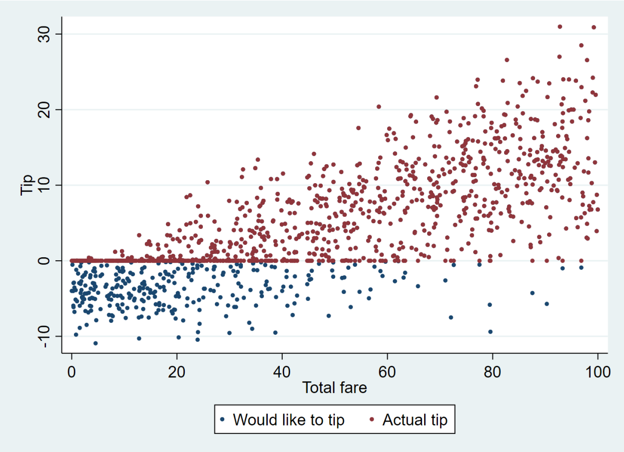
1. OLS overestimates the effect for short trips.
2. Vice versa for long trips.
3. Issues with prediction (negative tips?).

Tobit

Latent outcome — tip one would like to pay

$$Y_i^* = X_i\beta + \varepsilon_i, \quad \varepsilon_i \sim N[0, \sigma]$$

Observed outcome — actual tip: $Y_i = \max\{0, Y_i^*\}$. Parameters to estimate: β and σ.



Estimating Tobit parameters

Use ML; loglikelihood function:

$$\ell(\beta, \sigma; X, Y) = \sum_i \ell_i(\beta, \sigma; Y_i, X_i)$$

$$\text{where } \ell_i(\beta, \sigma; Y_i, X_i) = \begin{cases} \log P(Y_i = 0|X_i), & \text{if } Y_i = 0, \\ \log f_\varepsilon(Y_i - X_i\beta), & \text{otherwise.} \end{cases}$$

In Stata, run

```
tobit Y X, ll(0)
```

`ll` stands for lower limit; can be different from zero.

Interpreting the estimates

We want to know $\dfrac{\partial E[Y|X]}{\partial X}$. But what is $E[Y|X]$?

$$
\begin{aligned}
E[Y|X] &= P(Y=0|X) \times 0 + P(Y>0|X)E[Y|Y>0,X] \\
&= \Phi\left(\frac{X\beta}{\sigma}\right) E[Y|Y>0,X] \\
&= \Phi\left(\frac{X\beta}{\sigma}\right) \{X\beta + E[\varepsilon|\varepsilon > -X\beta]\}
\end{aligned}
$$

Since ε is truncated normal, $E[\varepsilon|\varepsilon > -X\beta] = \sigma\phi\left(\frac{X\beta}{\sigma}\right) \Big/ \Phi\left(\frac{X\beta}{\sigma}\right)$

$$E[Y|X] = \Phi\left(\frac{X\beta}{\sigma}\right) X\beta + \sigma\phi\left(\frac{X\beta}{\sigma}\right)$$

Marginal effects: $\dfrac{\partial E[Y|X]}{\partial X} = \Phi\left(\dfrac{X\beta}{\sigma}\right)\beta$

Tobit. Concluding remarks

- Beware of rigid assumptions: error is normal r.v. with constant variance.
- Constant variance assumption can be relaxed
- Truncated and censored regression — very similar concepts, slight differences in interpretation.

Poisson regression

Two use cases:

1. Count outcomes: childbirths, arrests, number of cars in the family.
2. **Continuous** non-negative dependent variable with skewed distribution and a mass point at zero.

Poisson regression

The outcome Y_i counts events (e.g., number of times woman i gave birth). Let $\theta_i = \exp(X_i\beta)$ approximate the "rate of arrival". Higher $\theta_i \rightarrow$ high draws of Y_i are more likely:

$$P(Y_i = 0|\theta_i) = \exp(-\theta_i),$$
$$P(Y_i = 1|\theta_i) = \theta_i \exp(-\theta_i),$$
$$\dots$$
$$P(Y_i = k|\theta_i) = \frac{\theta_i^k}{k!} \exp(-\theta_i),$$
$$\dots$$
$$E[Y_i|\theta_i] = \theta_i = \exp(X_i\beta)$$

Poisson regression – How to Estimate and Interpret

- **Estimation** — ML: $\max_\beta \sum_i [Y_i X_i \beta - \exp(X_i\beta)]$.
  In Stata:

  ```
  poisson Y X, vce(robust)
  ```

  Robust errors — in case the model is slightly misspecified.
- **Interpretation:**

$$\frac{\partial E[Y_i|X_i]}{\partial X_i} = \exp(X_i\beta)\beta$$

  Marginal effects vary across i, but we know how to deal with this: use average marginal or evaluate the effect at $\bar{X}$.

Surprising application of Poisson regression

Gravity equation — favorite tool of trade economists:

$$\log EXPORTS_{od} = \alpha + \rho_1 \log GDP_o + \rho_2 \log GDP_d \\ - \beta TARIFF_{od} - \gamma \log DIST_{od} + \varepsilon_{od}$$

Two issues:

1. $EXPORTS_{od} = 0$ for many country pairs. Typical shortcut: use a probit to model whether exports are non-zero.
2. We are typically interested in the aggregate: $E[\sum_d EXPORTS_{od}]$.

To find how tariffs affect exports: combine results from the export participation probit, the above gravity equation. Allow for the error in probit to correlate with the error in gravity, heteroskedasticity.

Things get clunky very fast. Silva, Tenreyro (2006) — use Poisson regression instead.

Gravity and Poisson

Write gravity equation in levels rather than in logs:

$$E[EXPORTS_{od}] = \exp\left(\alpha + \rho_1 \log GDP_o + \rho_2 \log GDP_d \right.$$
$$\left. - \beta TARIFF_{od} - \gamma \log DIST_{od}\right)$$

No logs — no need to worry about zeros.
Use standard estimation routine; interpreting results is easy: effect of tariffs on total exports is

$$\beta \sum_d \exp\left(\alpha + \rho_1 \log GDP_o + \rho_2 \log GDP_d \right.$$
$$\left. - \beta TARIFF_{od} - \gamma \log DIST_{od}\right)$$

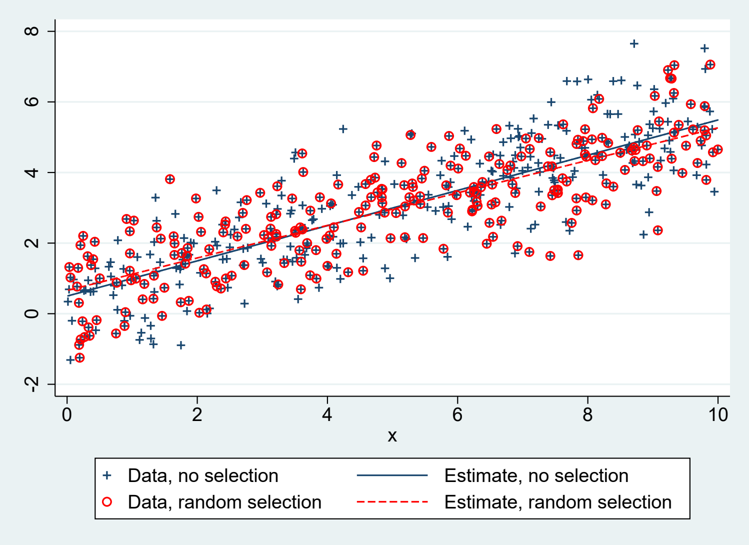
This works **even though exports are not discrete**. ML f.o.c. holds if $E[Y|X] = \exp(X\beta)$. Y doesn't have to be Poisson r.v.!

Sample Selection

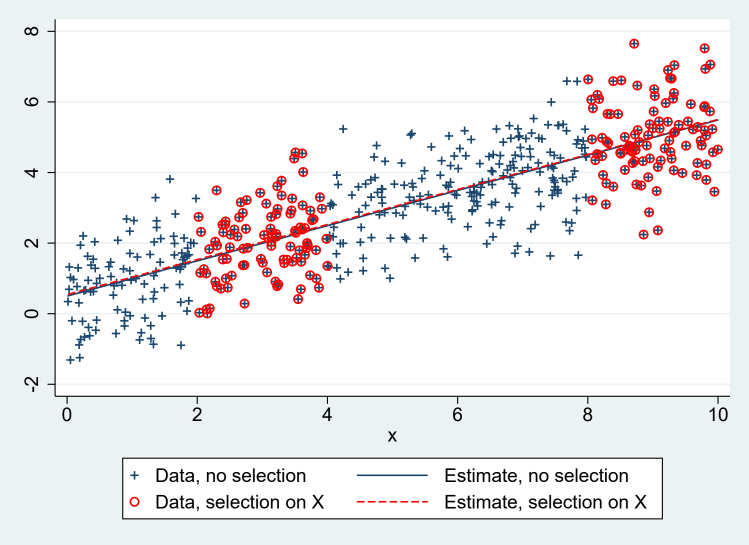
Sample selection — one of the major threats to identification.
What types of selection cause trouble?

1. Random selection
2. Selection on X_i
3. Selection on Y_i or ε_i

Random selection



Selection on controls



Selection on unobservables/outcome

More formal treatment

One way to think of selection:

1. Draw random sample $i = 1, .., N$.
2. Apply random treatment X_i. Outcome $Y_i = X_i\beta + \varepsilon_i$
3. Nature tosses a coin, $s_i \in \{0, 1\}$. If $s_i = 1$, we observe $\{X_i, Y_i\}$. **Selected sample**: $\mathcal{S} = \{i : s_i = 1, i = 1, .., N\}$

Run OLS to estimate β from

$$Y_i = X_i\beta + \varepsilon_i, \quad i \in \mathcal{S}$$

or, use OLS for

$$s_iY_i = s_iX_i\beta + s_i\varepsilon_i, \quad i = 1, .., N$$

The results would be **identical**, but the latter is easier to study.

Selection on controls

OLS is a consistent estimator for β in

$$s_iY_i = s_iX_i\beta + s_i\varepsilon_i, \quad i = 1, .., N$$

if $E[s_i\varepsilon_i s_iX_i] = 0$.

Is this the case under selection on controls?

$$\begin{aligned}
E[s_i\varepsilon_i s_iX_i] &= E[s_iX_i\varepsilon_i] \\
&= E\{E[E(s_iX_i\varepsilon_i|\varepsilon_i, X_i)|X_i]\} \\
&= E\{X_iE[\varepsilon_iE(s_i|\varepsilon_i, X_i)]\} \\
&= E\{X_iE[\varepsilon_ip(X_i)|X_i]\} \\
&= E\{X_ip(X_i)E[\varepsilon_i|X_i]\} \\
&= E\{X_ip(X_i) \cdot 0\} = 0
\end{aligned}$$

Selection on unobservables. Heckman's method

What if selection is endogenous? Heckman's model:

$$
\begin{aligned}
Y_i &= X_i\beta + \varepsilon_i, \\
s_i &\sim Probit : s_i = I[Z_i\gamma + \eta_i > 0] \\
(\varepsilon_i, \eta_i)|X_i, Z_i &\sim N\left[0, \left[\begin{array}{cc} \sigma_\varepsilon^2, & \sigma_{\varepsilon\eta} \\ \sigma_{\varepsilon\eta} & \sigma_\eta^2 \end{array}\right]\right]
\end{aligned}
$$

- In the canonical model, if $s_i = 0$, we observe (X_i, Z_i), but not Y_i.
- s_i correlates with ε_i via η_i. If $\sigma_{\varepsilon\eta} = 0$, selection is exogenous.
- Shocks are independent of (X_i, Z_i) — brave assumption!

Selection bias in Heckman's model

If you run Y_i on X_i for $i \in \mathcal{S}$, you are estimating

$$
\begin{aligned}
E[Y_i|X_i, Z_i, s_i = 1] &= X_i\beta + \underbrace{E[\varepsilon_i|s_i = 1, X_i, Z_i]}_{\text{selection bias}} \\
&= X_i\beta + E[\varepsilon_i|\eta_i > -Z_i\gamma, X_i, Z_i] \\
&\text{...tedious algebra...} \\
&= X_i\beta + \frac{\sigma_{\varepsilon\eta}}{\sigma_\eta^2} E[\eta_i|\eta_i > -Z_i\gamma, X_i, Z_i] \\
&\text{...more tedious algebra...} \\
&= X_i\beta + \frac{\sigma_{\varepsilon\eta}}{\sigma_\eta} \frac{\phi\left(\frac{Z_i\gamma}{\sigma_\eta}\right)}{\Phi\left(\frac{Z_i\gamma}{\sigma_\eta}\right)} \\
&= X_i\beta + \rho\lambda\left(\frac{Z_i\gamma}{\sigma_\eta}\right)
\end{aligned}
$$

where $\lambda(t) = \phi(t)/\Phi(t)$ — inverse Mills ratio, ρ — constant.

Heckman's two-step method

1. Run probit, s_i on Z_i, get $\widehat{\gamma}$. Find predicted $\widehat{\lambda}_i = \lambda(Z_i\widehat{\gamma})$.
2. Run OLS, Y_i on X_i and $\widehat{\lambda}_i$; $\widehat{\lambda}_i$ absorbs selection bias.

Notes:

- Std. errors in the second step have to be corrected in a special way (since $\widehat{\lambda} \neq \lambda$). Use built-in commands, they do necessary corrections.
- Although $X_i = Z_i$ is allowed, find a control affecting s_i, but not Y_i. Otherwise λ_i and X_i may be highly collinear.
- But do not exclude controls from X_i just because you want Z_i to have more variables than X_i! You risk violating $\varepsilon_i \perp Z_i$