Midterm Exam: Solutions

Gov 51, Spring 2025

March 2025

1 Matching Estimator

One of the key identifying assumption of matching is:

$$\underbrace{\mathbb{E}(Y_i(0)|T_i = 1)}_{\text{unobserved}} = \mathbb{E}(Y_i(0)|T_i = 0, i \in \mathcal{M})$$

Can you explain this equality in several sentences? (5 pts)

answer

On average, outcomes for those treated would have been the same to those in the matched set (not treated), if they were not treated.

2 OLS and Lasso

2.1 Loss function

We know that OLS regression and Lasso have different loss functions. Write down the two loss functions. (5 pts)

$$\widehat{\beta}_{\text{OLS}} = \arg\min_{\widetilde{\beta}} \sum_{i=1}^{N} \left(Y_i - \widetilde{\beta} X_i\right)^2$$

$$\widehat{\beta}_{\text{Lasso}} = \arg\min_{\widetilde{\beta}} \frac{1}{2} \sum_{i=1}^{N} \left(Y_i - \widetilde{\beta} X_i\right)^2 + \lambda|\tilde{\beta}|$$

2.2 Interpret the difference

Give a brief interpretation of the difference between OLS and Lasso. We want you to discuss this from three aspects:

1. When should we choose OLS over Lasso? And when do we prefer Lasso to OLS? (4 pts)

Answer

- OLS - for unbiased estimation of β.
- LASSO for variable selection.

2.3 Toy example

Please consider the toy dataset below.

Unit	Y	X
1	-4	1
2	-8	2
3	-12	3
4	-16	4
5	-20	5

2.3.1 OLS

Assume a linear model: $Y = \beta X + 0$. Can you calculate β_{OLS} for this dataset? (5 pts)

answer

$$\begin{aligned}\beta_{\text{OLS}} &= \frac{\sum_i X_i Y_i}{\sum_i X_i X_i} \\ &= \frac{-4 \cdot 1 - 8 \cdot 2 - 12 \cdot 3 - 16 \cdot 4 - 20 \cdot 5}{1 + 4 + 9 + 16 + 25} \\ &= \frac{-220}{55} \\ &= -4\end{aligned}$$

2.3.2 Lasso

Suppose $\lambda = 110$, can you calculate β_{Lasso} for this dataset? (5 pts)

answer

$$\lambda = 110$$

This makes $\frac{\lambda}{\sum_i^N X_i X_i} = \frac{110}{55} = 2$

$$\begin{aligned}\beta_{\text{Lasso}} &= \left(\frac{\sum_i X_i Y_i}{\sum_i X_i X_i} - -1 \cdot \frac{\lambda}{\sum_i X_i X_i} \right) \cdot \mathbf{1} \left(|\beta_{\text{OLS}}| > 2 \right) \\ &= (-4 - -2) * 1 \\ &= -2\end{aligned}$$

3 Difference in difference and OLS

3.1 Four scenarios

Assume a linear regression to run DID design:

$$Y_i = \beta_0 + \beta_1 \cdot \text{period} + \beta_2 \cdot \text{treatment} + \beta_3 \cdot \text{period} \cdot \text{treatment} + \epsilon_i$$

Can you fill in the table below of the four possible situations? (8 pts)

answer

	Period= 0	Period= 1
Treatment = 0	$Y_i = \beta_0 + \epsilon_i$	$Y_i = \beta_0 + \beta_1 + \epsilon_i$
Treatment = 1	$Y_i = \beta_0 + \beta_2 + \epsilon_i$	$Y_i = \beta_0 + \beta_1 + \beta_2 + \beta_3 + \epsilon_i$

3.2 DID estimator

From the table above, can you tell us the value of DID estimator? Please show the process that you derive the result. (12 pts)

answer

The DID estimator would be:

$$\underbrace{(\beta_0 + \beta_1 + \beta_2 + \beta_3 + \epsilon_i)}_{\text{Period 1, treated}} - \underbrace{(\beta_0 + \beta_2 + \epsilon_i)}_{\text{Period 0, treated}}$$

$$-$$

$$(\underbrace{(\beta_0 + \beta_1 + \epsilon_i)}_{\text{Period 1, control}} - \underbrace{(\beta_0 + \epsilon_i)}_{\text{Period 0, control}})$$

Which yields: β_3

4 Bonus Question

Please show the following equation to be true:

$$\hat{\beta} = \frac{\sum_{i=1}^{N} X_i(Y_i - \bar{Y})}{\sum_{i=1}^{N} X_i(X_i - \bar{X})} = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

Hint: spread out both numerator and denominator of both forms, and take a closer look at the difference. (5 pts)