

Exam 5106/9106B – Fall 2021 Postponed

IMPORTANT: Always motivate your answers. Show knowledge and understanding of the concepts taught in the course. Your answers should be as short as possible and as long as necessary. Total marks of the examination is 140. Each sub-question carries 10 marks.

1. Suppose y_i $i = 1, 2, \dots, 10$ are i.i.d draw from a unknown continuous distribution where the CDF and PDF are $F(\cdot)$ and $f(\cdot)$. You are supposed to estimate the distribution functions based on the attached Stata output. Suppose that y_i follows the following distribution

$$f(y; \alpha) = \begin{cases} \alpha y^{-(1+\alpha)}, & y \geq 1 \\ 0, & \text{otherwise} \end{cases}$$

- (a) Write down the likelihood function and derive the score.

ANSWER HINT:

$$\begin{aligned} \mathcal{L} &= \sum_{i=1}^n \mathcal{L}_i = \sum_{i=1}^n (\log \alpha - (1 + \alpha) \log y_i) \\ &= n (\log \alpha - (1 + \alpha) \overline{\log y}) \end{aligned}$$

where $\overline{\log y} \equiv \frac{1}{n} \sum_{i=1}^n \log y_i$

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \sum_{i=1}^n \left(\frac{1}{\alpha} - \log y_i \right) = n \left(\frac{1}{\alpha} - \overline{\log y} \right)$$

- (b) Derive the ML estimator of α from the score. Compute it, and derive the standard error of your estimate.

ANSWER HINT:

Solving $\frac{\partial \mathcal{L}}{\partial \alpha} = 0$ gives

$$\hat{\alpha} = \frac{1}{\overline{\log y}}$$

which equals 1.752539 using the data above

Since $\text{Var}(\hat{\alpha}_{MLE}) = -\frac{1}{n} E[\partial^2 \mathcal{L}_i / (\partial \alpha)^2]^{-1}$ and

$$E[\partial^2 \mathcal{L}_i / (\partial \alpha)^2] = -\frac{1}{\alpha^2}$$

the s.e. equals .554 in the data.

- (c) Test for $H_0 : \alpha = 1$ against $H_1 : \alpha \neq 1$ at the 5% significance level using the LR, Wald and LM test.

ANSWER HINT:

LR: $2 * 10 * (\log(1.75) - (1 + 1.75) * .57 - \ln(1) - (1 + 1) * .57) = 2.63$

W: $(1.752539 - 1)^2 / 0.554^2 = 1.843$

LM: $(10(1 - .57))^2 / 10 = 1.843$

none of which are stat sign at the 5% level (3.84).

- (d) Explain how would you use bootstrapping to obtain a standard error estimate for $\hat{\alpha}_{MLE}$. Use the Stata output to compute the bootstrapped standard error of $\hat{\alpha}_{MLE}$.

ANSWER HINT:

Nonparametric bootstrap:

implement the following procedure:

- i. Take an r -th random sample **with replacement** from the data
- ii. Calculate the statistic $\hat{\beta}^{(*r)}$
- iii. Repeat steps 1. and 2. many (R) times

The empirical distribution of $\hat{\beta}^{(*r)}$, $r = 1, \dots, R$ is the approximation to the sampling distribution of $\hat{\beta}$

- used to construct s.e.'s and confidence intervals..

From stata output, we can calculate that that bootstrap se is .6316372

- (e) Use the Stata output to generate the bias corrected estimate of α

ANSWER HINT:

Non-linear estimators are usually biased in small

$$\text{Bias} = E[\hat{\beta}] - \beta$$

From stata output, we have

$$\hat{\beta}^{(*)} = 2.127461$$

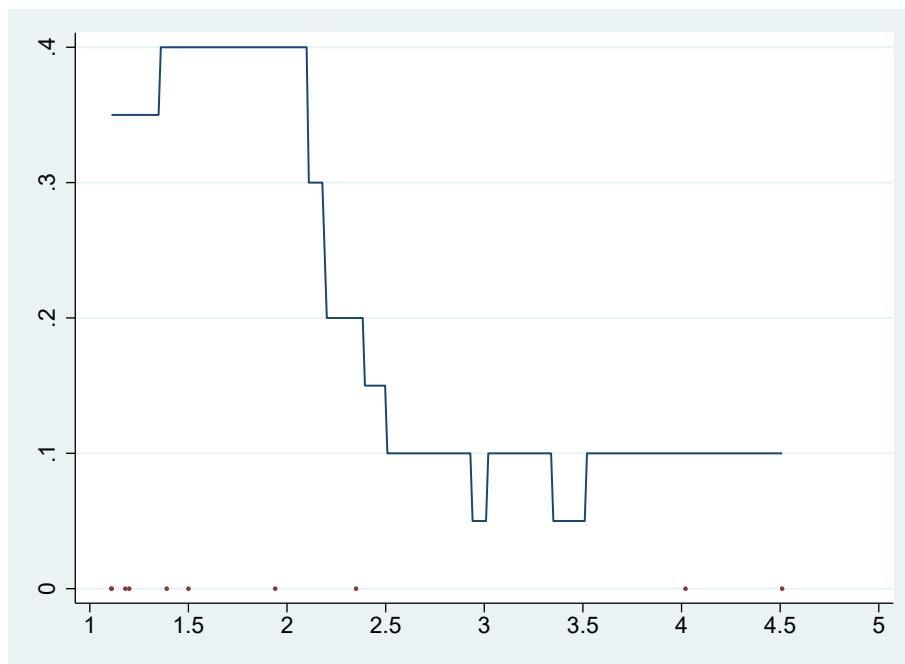
We can estimate the bias with the bootstrap by

$$\text{Bias}^* = \overline{\hat{\beta}^{(*)}} - \hat{\beta}$$

The bias-corrected estimate of β equals

$$\hat{\beta} - \text{Bias}^* = \hat{\beta} - (\overline{\hat{\beta}^{(*)}} - \hat{\beta}) = 2\hat{\beta} - \overline{\hat{\beta}^{(*)}} = 1.3776173$$

- (f) Now suppose that you don't know the distribution family and want to use nonparametric methods to estimate the density. The following figure shows a kernel density estimate based on the same 10 observations. The red dots represent the observations.



One of the following five kernels is used, and the bandwidth takes value in the set $\{0.5, 1, 2\}$. Determine which kernel and bandwidth are used.

Kernel	$K(z)$
Uniform	$\frac{1}{2} \times 1[z < 1]$
Triangular	$(1 - z) \times 1[z < 1]$
Epanechnikov	$\frac{3}{4}(1 - z^2) \times 1[z < 1]$
Biweight	$\frac{15}{16}(1 - z^2)^2 \times 1[z < 1]$
Gaussian	$(2\pi)^{-1/2} \exp(-z^2/2)$

ANSWER HINT:

Uniform with bandwidth 1. piecewise constant fitted curve \rightarrow uniform.
 $f(3)=0.05 \rightarrow h=1$

Let Z be a random variable that is distributed according to pdf $f(\cdot)$. In the following, you are asked to find an estimate for $p = \text{prob}(Z \leq 3)$.

- (g) Compute the estimate of the MLE estimator of p and its standard error.

ANSWER HINT:

Note that $p = F(3)$ and realize that $F(3) = \int_1^3 f(x)dx = 1 - 3^{-\alpha}$

Using result from (b), we have $\hat{p}_{MLE} = 1 - 3^{-\hat{\alpha}} = 0.85$.

Nonlinear function of a MLE estimator. Standard errors can be obtained through either the Delta method

$$d(1 - 3^{-\alpha})/d\alpha = 3^{-\alpha} \ln 3$$

$$\text{so } \hat{se}(\hat{p}_{MLE}) = 3^{-\hat{\alpha}} \ln 3 se(\hat{\alpha}) = 0.089$$

- (h) Propose a non-parametric estimator of p , and compute its estimate and standard error.

ANSWER HINT:

Empirical Distribution Function is a simple nonparametric estimator for the CDF $F()$.

So we have $\hat{F}(3) = \frac{1}{10} \sum_{i=1}^{10} I(y_i \leq 3) = 0.8$

$\hat{Var}[\hat{F}(3)] = \frac{\hat{F}(3)(1-\hat{F}(3))}{10} = 0.016$ se 0.126

2. Using data from a universal child care program aimed at 3-6 years olds, a group of researchers want to better understand the effectiveness of universal child care programs. Their main outcome variable is a measure of overall school readiness y_i at age 7. They also observe individual child care attendance D_i during age 3 to age 6.
- (a) Explain whether you can use the following regression to estimate the effect of child care attendance on school readiness:

$$y_i = a + \gamma D_i + \varepsilon_i \quad (1)$$

What is the key assumption for $\hat{\gamma}$ to have a causal interpretation?

ANSWER HINT:

This is a simple comparison of mean between those who attend and not attend day care. $\hat{\gamma}$ has a causal interpretation as long as CIA holds. However, readiness to school and choice to attend day care can be endogenously determined.

- (b) They suggest using a measure of local child care supply, Z_i , which is the local child care coverage rate at age 3 as the instrument for D_i . Consider first the case where we use a binary version of the instrument, namely $\tilde{Z}_i = Z_i > 0.5$. In IV analysis, it is often useful to classify people into different groups, such as always takers, etc. Define these groups and classify the children who participate in the study into these groups. Name the assumptions underpinning instrumental variables as a strategy for identifying the effect of an endogenous treatment on the outcome for compliers. Suppose that these assumptions hold, based on the Stata output (d: treatment status, z: binary instrument), calculate the share of children in each group.

ANSWER HINT:

- (a) Random assignment:
- (b) Exclusion restriction:
- (c) First-stage relevance:
- (d) Monotonicity

- i. **Compliers:** $d_i(1) = 1$ and $d_i(0) = 0$;
people who take the drug if assigned to treatment group, but not if not assigned.
- ii. **Always-takers:** $d_i(1) = 1$ and $d_i(0) = 1$;
people who always take the drug, regardless of their assignment status
- iii. **Never-takers:** $d_i(1) = 0$ and $d_i(0) = 0$;
people who never take the drug, regardless of their assignment status
- iv. **Defiers:** $d_i(1) = 0$ and $d_i(0) = 1$;
people who don't take the drug if assigned to treatment group, but will take if not assigned..

$$p_a = \Pr(D = 1|Z = 0) = 0.52$$

$$p_n = \Pr(D = 0|Z = 1) = 0.24$$

$$p_c = 1 - p_a - p_n = 0.24$$

- (c) In the following, we use the continuous instrument Z_i . Assume that the child care attendance follows the following latent index model:

$$D_i^* = \beta_0 + \beta Z_i - V_i$$

$$D_i = 1, \text{ if } D_i^* > 0 \quad D_i = 0, \text{ otherwise.}$$

where V_i follows a standard normal distribution. The potential outcome can be specified as $y_i(1) = a + \gamma + \varepsilon_{i1}$ and $y_i(0) = a + \varepsilon_{i0}$ with $E(\varepsilon_{ij}|D_i) = 0$. Explain what the random variables ε_{ij} and V_i represent? What does it imply if ε_{ij} and V_i is correlated?

ε_{ij} unobserved individual characteristics that influence the potential outcome. $\varepsilon_{i1} - \varepsilon_{i0}$ measures heterogeneity in gain from treatment.
 V_i unobserved characteristics that make individual less likely to receive treatment (unobserved cost for treatment)
 If they are correlated, we have “selection in gains”, selection bias if using OLS to regress y on d .

- (d) Suppose that Z_i is a valid instrument for D_i , define the marginal treatment effect(MTE), and outline how you will estimate it.

ANSWER HINT:

$U \sim \text{Uniform}(0, 1)$

The MTE is defined as (f.e. Heckman & Vytlacil, 1999):

$$MTE(p) \equiv E[Y^1 - Y^0 | U = p]$$

the average treatment effect for those who are on a given margin p of participation, and who will participate if $p(Z) > p$

Let

$$k_j(p) \equiv E[\varepsilon_{ij} | U = p]$$

and

$$k(p) \equiv k_1(p) - k_0(p) = E[\varepsilon_{i1} - \varepsilon_{i0} | U = p]$$

so that

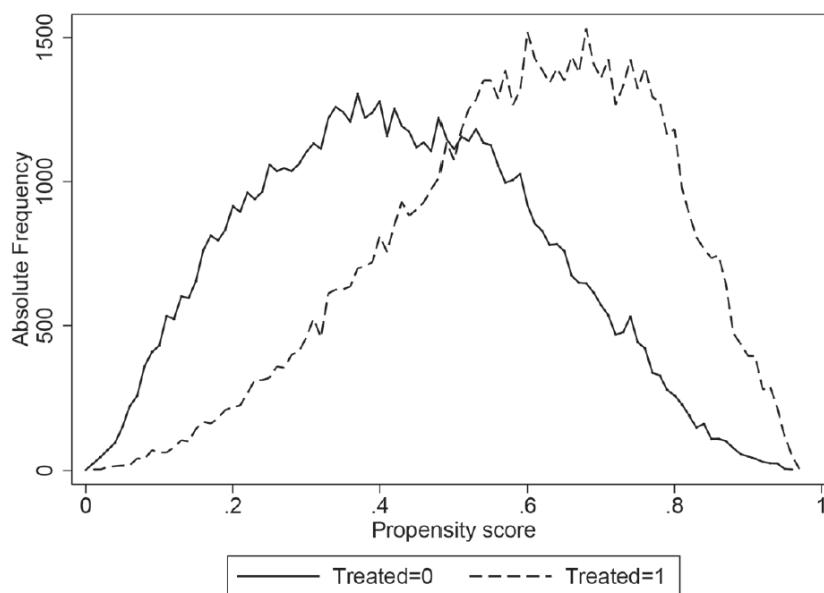
$$MTE(p) \equiv E[Y_1 - Y_0 | U = p] = \gamma + k(p)$$

$k(p)$ is the average unobserved gain of treatment for those whose unobserved characteristics make them at the participation margin at $P(Z) = p$.

To estimate the MTE, one can use the following steps:

1. Estimate and predict the propensity score using a probit.
2. run a OLS on y_i using a polynomial of p_i or use nonparametric methods
3. the MTE is defined as the marginal effect of p_i on y_i (or using numerical differentiation method in case of nonparametric estimation in step 2)

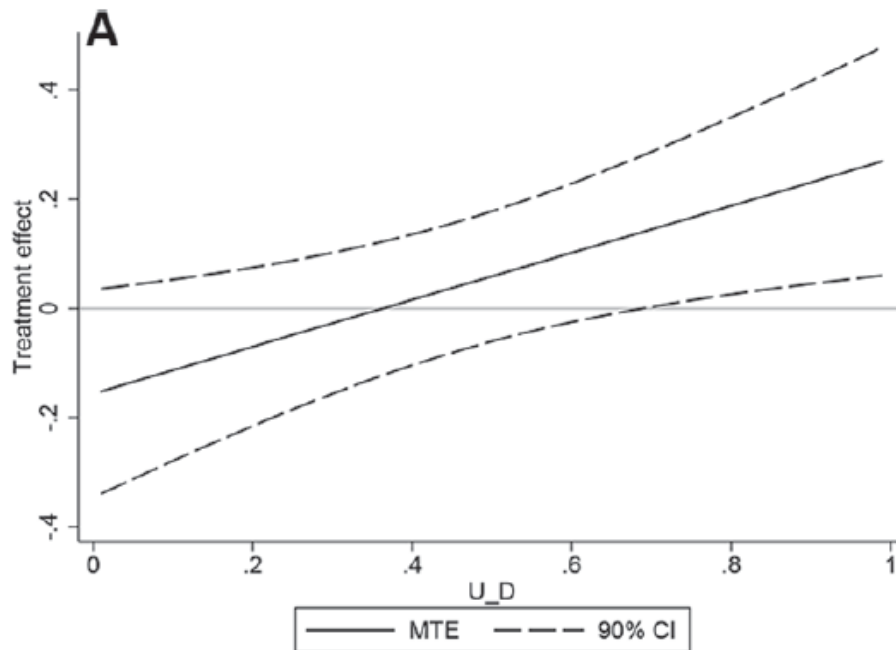
- (e) The following figure shows the distribution of the predicted propensity score by treatment status. What does it imply for the MTE estimation?



ANSWER HINT:

It shows a large common support for the propensity score $P(Z)$, which ranges from 0.01 to 0.96. Note that the MTE is only identified on the common support, This means that the MTE identified for the interval 0.01 to 0.96. It might be hard for the student to know the exact cutting points so it is ok if they claim full support.

- (f) The following figure shows the estimated MTE function with 90% confidence interval. What do the axes represent? Does the figure tell us anything about the selection mechanism? Based on the MTE function, arrange ATE, ATT, and ATNT in the order of increasing magnitude.



ANSWER HINT:

x axis shows the unobserved resistance to attend the daycare.

y axis shows the treatment effects

The MTE is upward sloping: Reverse selection on gains.

In this case: $ATNT > ATE > ATT$

- ATE is an equally weighted average over the MTE curve
- ATT is a weighted average, more heavily weighting individuals to the left ("low resistance = high propensity")
- ATU is a weighted average, more heavily weighting individuals to the right ("high resistance = low propensity")

Stata Output

Problem 1.

```
. describe
```

Variable name	Storage type	Display format	Value label	Variable label
x	float	%9.0g		x
lnx	float	%9.0g		natural log of x

```
. list
```

	x	lnx
1.	1.2	.1823216
2.	1.5	.4054651
3.	4.02	1.391282
4.	1.94	.662688
5.	1.11	.10436
6.	1.11	.10436
7.	4.51	1.506297
8.	1.18	.1655144
9.	2.35	.8544153
10.	1.39	.3293037

```
. su
```

Variable	Obs	Mean	Std. dev.	Min	Max
x	10	2.031	1.247918	1.11	4.51
lnx	10	.5706007	.5242123	.10436	1.506297

```
. * program to compute MLE estimates of alpha
. * it returns a scalar alpha_hat
```

```
. program ml_est, rclass
.     version 17
.     code skipped
.     return scalar alpha_hat
. end
```

```
. bootstrap r(alpha_hat), reps(200) saving(mybs, replace): ml_est
```

```
. use mybs, clear
. (bootstrap: ml_est)
```

```
. use mybs, clear
. (bootstrap: ml_est)
```

```
. list
```

	_bs_1
1.	2.216439
2.	1.536116
3.	1.769239
4.	1.499552
5.	2.741992
6.	1.577018
7.	3.28103
8.	2.306023
9.	2.749356
10.	1.597841

Critical Values of Chi-square

df	.50	.25	.10	.05	.025	.01	.001
1	0.45	1.32	2.71	3.84	5.02	6.63	10.83
2	1.39	2.77	4.61	5.99	7.38	9.21	13.82
3	2.37	4.11	6.25	7.81	9.35	11.34	16.27
4	3.36	5.39	7.78	9.49	11.14	13.28	18.47
5	4.35	6.63	9.24	11.07	12.83	15.09	20.52
6	5.35	7.84	10.64	12.59	14.45	16.81	22.46
7	6.35	9.04	12.02	14.07	16.01	18.48	24.32
8	7.34	10.22	13.36	15.51	17.53	20.09	26.12
9	8.34	11.39	14.68	16.92	19.02	21.67	27.88
10	9.34	12.55	15.99	18.31	20.48	23.21	29.59

Problem 2:

```
. tab z d
```

z	d		Total
	0	1	
0	2600	2800	5400
1	1800	5600	7400
Total	4400	8400	12800