# Labor Economics I: Midterm Exam

Spring 2023

## 1 Roy Model Setup (30%)

1. Write down a Roy model with two alternatives. Use $Y$ as your outcome variable. Let the mean effect (payoff) be different but homogeneous across individuals. Allow for individual idiosyncratic preferences in your model. Mark the subscripts carefully.
2. Use a figure to illustrate the Roy model (like the one we plotted in class). Explain the x-axis, y-axis, and mark the region where each alternative is optimal.
3. Write down an application of Roy model. How does the above notations correspond to the example you gave?

## 2 Roy Model with 3 Alternatives (20%)

Consider the following Roy model:

$$\begin{aligned}
Y_{i1} &= \mu_1 + \epsilon_{i1} \\
Y_{i2} &= \mu_2 + \epsilon_{i2} \\
Y_{i3} &= \mu_3 + \epsilon_{i3} \\
D_i &= \arg\max_l \{Y_{il}\}, D_i \in \{1, 2, 3\}
\end{aligned}$$

1. Come up with an example and map that into this framework.
2. We assume that $\epsilon_{il}$ follows iid type one extreme value (T1EV) distribution (CDF $F(\epsilon) = e^{-e^{-\epsilon}}$). A famous property of this distribution is that
$$P_{ij} = P(V_{ij} + \epsilon_{ij} > V_{ik} + \epsilon_{ik} \forall j \neq k) \Rightarrow P_{ij} = \frac{e^{V_{ij}}}{\sum_{q \in J} e^{V_{iq}}},$$
where $J$ is the set of all alternatives Please use this property to derive the expression of $P(D = 1)$.
3. Suppose you collect 10 data points, $i = \{1, 2, 3, ..., 10\}$. Write down the likelihood function.
4. How will you find the maximum likelihood estimates? Don't do any differentiation, just write your estimator as arg max of what over what. Think about what are observable and what are unobservable.

## 3 IV (20%)

Judge IV is a popular design to estimate the causal effect of incarceration (putting people in jail). The idea is that there are strict judges and lenient judges. If the judges are randomly assigned, being assigned to a strict judge makes you more likely to end up in jail. Let $Z_i = 1$ stand for being assigned to a strict judge and $Z_i = 0$ stand for being assigned to a lenient judge. Let's say we want to estimate the effect of incarceration on one's health.

1. Use the Roy model or the potential outcome framework to write down the above example. Write down the notations and define them carefully.

2. Write down one scenario where the LATE assumption is not satisfied.

3. Suppose the LATE assumptions are satisfied, write down the estimator for estimating LATE.

4. Who are the compliers in this setup? Explain both formally and intuitively.

## 4 Control for Observables (20%)

Suppose you want to study the effect of having an economics Ph.D. on your salary. You are thinking about using the scholarship as an instrument variable. You also collect one's sex and age as potential covariates.

1. Define some notations and write down the formal definition of random assignment. Explain the example and discuss when the random assignment will be violated.

   A: Denote $Y_i(1)$ as the salary when one has an econ Ph.D., and $Y_i(0)$ as the salary when one does not have an econ Ph.D. Let $Z_i$ be the dummy variable of whether one receives a scholarship. $X_i$ denotes one's sex and age.

   The definition of random assignment is that $(Y_i(1), Y_i(0)) \perp D_i$. For example, when brighter students tend to get a Ph.D., then the random assignment may fail.

2. Define some notations and write down the formal definition of the conditional independence assumption. Explain the example and discuss when the conditional independence assumption will be violated.

   A: Conditional independence assumption is $(Y_i(1), Y_i(0)) \perp D_i | X_i$. I asked for control for observables here, so you can also treat $Z_i$ as an observable. You could also discuss the conditional independence of the treatment given the instrument. $(D_i(1), D_i(0)) \perp Z_i | X_i$.

3. Write down the formal definition of the propensity score and explain it intuitively.

   A: The propensity score is the probability of receiving the treatment. You can define it as either $P(D_i = 1 | X_i, Z_i)$ or $P(D_i = 1 | X_i)$.

4. In class, we specify this estimator:

$$\hat{\tau}^{\text{IP}} = \frac{1}{n}\sum_{i:D_i=1}\frac{Y_i}{\hat{p}(X_i)} - \frac{1}{n}\sum_{i:D_i=0}\frac{Y_i}{1-\hat{p}(X_i)}$$
$$= \frac{1}{n}\sum_{i}^{n}\frac{D_iY_i}{\hat{p}(X_i)} - \frac{1}{n}\sum_{i}^{n}\frac{(1-D_i)Y_i}{1-\hat{p}(X_i)}.$$

Explain how you will estimate this with a logistic regression and a linear regression.

A: Under this specific estimator, there is nothing to do with the instrument. You just run a logistic regression of $D$ on $X$ to obtain $\hat{p}$, and then run a linear regression of $Y$ on $D$ weighted by $\frac{1}{\hat{p}}$ for the treatment group and by $\frac{1}{1-\hat{p}}$ for the control group.

## 5 NBER Working Paper Series (10%)

1. Have you signed up for the mailing list? What day of the week do NBER papers for the week usually arrive your mailbox? If you don't remember, describe a recent NBER working paper instead (and name one author at least).