Spring 2023

Gregory Bruich, Ph.D.
Dept. of Economics, Harvard University

Lab 3: The Moving to Opportunity (MTO) Experiment

Methods/concepts: treatment effect estimation, non-compliance, intent-to-treat (ITT) effects, treatment-on-the-treated (TOT) effects, bar graphs to visualize treatment effects

LAB DESCRIPTION

This lab uses data from the Moving to Opportunity (MTO) Experiment called **mto.dta** to estimate the causal effect of neighborhoods on mental health.[1] For more details on the variables included in these data, see Table 1. A list and description of each of the Stata and R commands needed for this lab are contained in Table 2 and Table 3, respectively.

The Moving to Opportunity (MTO) Experiment was implemented by the Department of Housing and Urban Development in 1994-1998. Figure 1 shows the most common housing locations for families in the control group (the Martin Luther King Jr. Towers in Harlem) and the experimental group (Wakefield in the Bronx). Figure 2 shows the recruitment flyer for the experiment.

In today's lab, we will look at the long-run impacts on mental health of individuals who were adults when experiment was conducted in 1994-1998. We will focus on individuals in the Experimental Voucher and Control group only. The outcome variable that we will focus on is the Kessler 6 Psychological Distress Index score measured in interviews conducted between June 2008 and April 2010. The Kessler 6 Psychological Distress Index was developed by Ronald Kessler at Harvard Medical School. (Kessler was a co-author with Ludwig, Duncan, Gennetian, Katz, Kling, and Sanbonmatsu on a long-run follow up of MTO published in *Science* in 2012. The data for this lab are from their paper).

QUESTIONS

1. Briefly summarize the results from the lecture on the Moving to Opportunity Experiment. What were the results for individuals who were young children in 1994-1998? What were the results for older children and adults in 1994-1998?

2. Now turn to the **mto.dta** file. What fraction of individuals in the control group moved? Report the mean of *moved* for observations with *voucher* equal to 0.

3. What fraction of individuals in the experimental group moved? Report the mean of *moved* for observations with *voucher* equal to 1.

4. Estimate a linear regression (`regress` in stata and `lm` in R) of *moved* on an intercept and *voucher*. What is the relationship between the estimated coefficients in the regression and the means that you reported in questions 2-3?

5. What do your results in the previous questions tell us about the *compliance rate* in the Moving to Opportunity Experiment? Is there non-compliance, and if so, is it one-sided or two-sided?

6. Use a linear regression (`regress` in stata and `lm` in R) to estimate the *intent-to-treat effect* of the experimental voucher on the *Kessler 6 Psychological Distress Index score* (the variable *kessler*).

[1] The data are a simulated dataset that preserves the key features of the Moving to Opportunity Experiment, but does not contain actual information from real households to protect their privacy.

7. Use your estimates of the compliance rate and the intent-to-treat effect in the previous questions to estimate the *treatment-on-the-treated effect* of actually using the experimental voucher to move (the variable *moved*) on the psychological distress index (the variable *kessler*). Provide some intuition for the calculation of the TOT estimate.

8. A natural, but incorrect, way of analyzing data from an experiment with non-compliance is to compare outcomes for those who actually received the treatment and those who did not receive the treatment. Imbens and Rubin (2015) refer to this incorrect analysis as an "As-Treated" analysis. Implement this incorrect approach by calculating the difference in means of *kessler* for those who moved and those who did not move

9. Another incorrect way of analyzing data from an experiment with non-compliance is to drop observations in the treatment group that did not receive the treatment (and drop observations from the control group who actually received the treatment if there are any). Imbens and Rubin (2015) refer to this incorrect analysis as a "Per Protocol" analysis. Implement this incorrect approach by calculating the differences in means for *kessler* in the experimental treatment versus control group, after excluding observations in the treatment group that did not move (i.e., `voucher == 1 & moved == 0`).

10. Contrast your (incorrect) "per protocol" and "as treated" estimates with the (correct) *treatment-on-the-treated effect* estimate you calculated earlier. Which method yields the biggest estimate?

11. Explain why the "per protocol" and "as treated" approaches lead to biased estimates, while the TOT leads to valid inference about the impact of MTO.

12. The most natural way to visualize estimates from a randomized experiment is using a bar graph, with one bar representing the control group and a second bar representing the treatment group. The height of the bar for the treatment group equals the sum of the control group mean and the ITT or TOT estimate, allowing one to easily judge the magnitude of the treatment effects. Construct three bar graphs (and include them in your lab write up) to visualize:

    a. The fraction of the control group that moved and the fraction of the treatment group that moved from question 4.

    b. The mean of *kessler* in the control group and the mean of *kessler* in the treatment group, corresponding to the intent-to-treat (ITT) effect estimate from question 6

    c. The mean of *kessler* in the control group and the *adjusted* mean of *kessler* in the treatment group, corresponding to the treatment-on-the-treated (TOT) effect from question 7

13. The files to submit for this lab are:

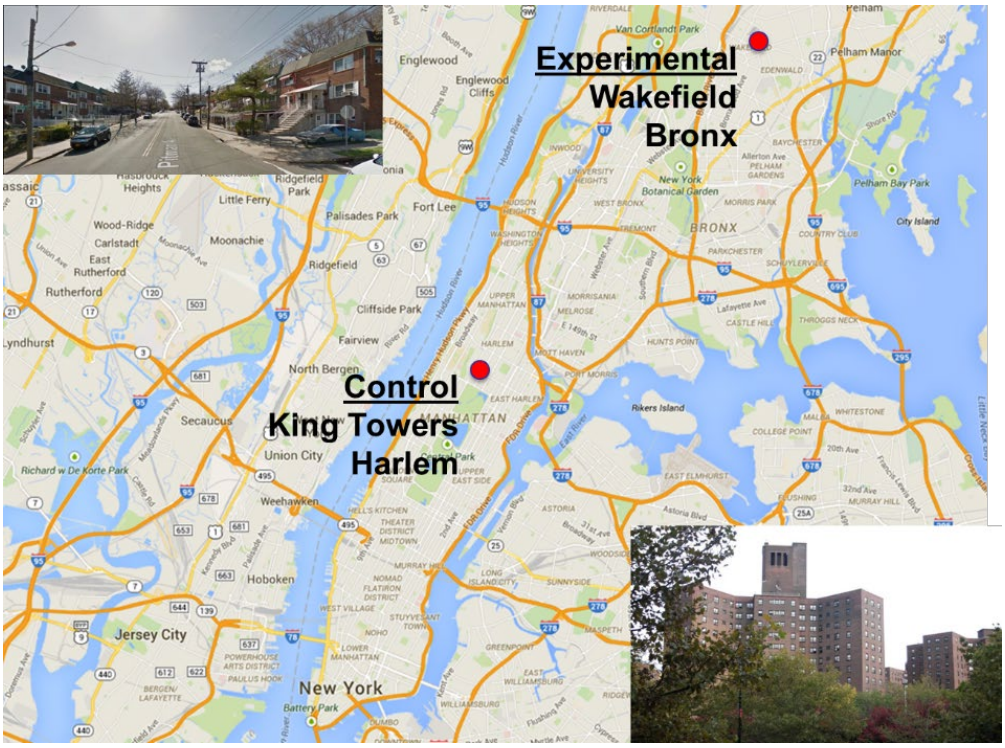
    a. Your well annotated do-file/.R file replicating all your analyses above (with enough comments that a principal investigator on a research project would be able to follow and understand what each step of the code is doing). You can submit this to Gradescope.

    b. For Stata users, a log-file with the log showing the output generated by your final do-file. You can submit this file to the same Gradescope assignment as the do-file.

    c. A PDF version of the solutions to the above questions. For graphs, you can save them as .png files and insert them into the document. You can submit this file to the same

Gradescope assignment as the .R script/do-file and log-file. (Please do not submit a word document: we can only read PDFs in Gradescope).

Figure 1
Most Common MTO Residential Locations in New York, NY



Note: Figure shows the most common housing locations for families in the control group (the Martin Luther King Jr. Towers in Harlem) and the experimental group (Wakefield in the Bronx) in the Moving to Opportunity Experiment.

Figure 2
Recruitment Flyer for the Moving to Opportunity Experiment

Note: Figure shows the recruitment flyer that was distributed in public housing units to recruit participants for the Moving to Opportunity Experiment.

DATA DESCRIPTION, FILE: mto.dta

The data consist of $N = 2{,}595$ individuals in the Moving to Opportunity Experiment who were adults (not children) in 1994-1998 when the randomization was conducted. (Individuals in the Section 8 group have been dropped from the data.) The Psychological Distress Index was measured in interviews conducted between June 2008 and April 2010. For more information about these data, see Jens Ludwig, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, and Lisa Sanbonmatsu. 2012. "Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults," *Science* 331(6101): 1505-1510.

TABLE 1
Variable Definitions

	Variable (1)	Description (2)	Obs. (3)	Mean (4)	St. Dev. (5)	Min (6)	Max (7)
1	*site*	Site in 1994-1998: 1=Baltimore 2=Boston 3=Chicago 4=Los Angeles 5=New York City	2,595	3.159	1.341	1	5
2	*moved*	Mover/Treatment Compliance Flag: 1 = moved with experimental voucher in 1994-1998, 0 = did not move using the experimental voucher in 1994-1998	2,595	0.242	0.428	0	1
3	*voucher*	1 = randomly assigned to experimental voucher group in 1994-1998, 0 = randomly assigned to the control group in 1994-1998	2,595	0.561	0.496	0	1
4	*kessler*	Psychological Distress Index from 10-15 year follow-up interviews with families Measured in interviews conducted between June 2008 and April 2010.	2,595	6.551	3.750	0	24

Note: Table describes variables in mto.dta.

TABLE 2
Stata Commands

STATA command	Description
*clear the workspace clear all version 17 *change working directory and open data cd "C:\Users\gbruich\Ec 50\Lab 3\" use mto.dta, clear *Display all variables in the data describe *Report detailed information on all variables codebook	This code shows how to clear the workspace, change the working directory, and open a Stata data file. To change directories on either a mac or windows PC, you can use the drop down menu in Stata. Go to file -> change working directory -> navigate to the folder where your data is located. The command to change directories will appear; it can then be copied and pasted into your .do file. The describe and codebook commands will report information on what is included in the data set loaded into memory.
*Summary stats for one variable sum yvar *Observations with treatment_group equal to 1 sum yvar if treatment_group == 1 *Observations with treatment_group equal to 0 sum yvar if treatment_group == 0	We used these commands in Lab 1. These commands report means and standard deviations for *yvar*. The first line calculates these statistics across the full sample. The other lines illustrate how to calculate these statistics for observations meeting certain criteria: when another variable in the data is equal to 1, or equal to 0.
*Estimate linear regression regress yvar treatment_group	We used these commands in Lab 2. These commands report estimated regression coefficients from a regression of *yvar* on an intercept and a variable *treatment_group*. The intercept is always included by default, which is usually what you want.
* Opportunity Insights Style Bar Graphs clear all set obs 2 gen treat = 0 replace treat = 1 in 2 *Control group mean gen y = 0. 1 in 1 *Treatment group mean replace y = .4 in 2 #delimit ; twoway (bar y treat if treat == 0, barwidth(0.4) color(red)) (bar y treat if treat == 1, barwidth(0.4) color(blue)) , legend(off) xlab(0 "Control Group" 1 "Treatment Group") xtitle("") ytitle("Moved Using Experimental Voucher" " ") xsc(range(-0.3 1.3)) ylab(0(.2).5,nogrid) graphregion(color(white)) bgcolor(white) ; #delimit cr graph export fig1_compliance.png, replace	These commands show how to draw an Opportunity Insights style bar graph. We start by clearing the workspace. We will make a data set with two observations using `set obs 2`. Then we create a variable treat that equals 0 for the first observation and 1 for the second. Next we create a variable `y` that equals the height of the bars that we want to plot. For the first observation, we fill in what the height should be for the control group bar: 0.1. For the second observation, we fill in what we want for the treatment group bar: 0.4. Then we use the #delimit command to reset the character that marks the end of a command to a semi colon ; and later set it back to a carriage return `cr`. We do this because the options for the graph are quite complicated and spill over onto multiple lines. Everything from `twoway` through the semi colon in red is one command. We create the graph by overlaying two `bar` type twoway graphs, one for the control group bar and one for the treatment group bar. Most of the other lines are just formatting options. The ones in blue are what you might want to change: the colors for the bars, the y-axis label, and the range and increments on the y-axis of the graph. Everything else can stay the same. The `graph export` command saves the graph.

*close any possibly open log-files cap log close *start a log file log using lab3.log, replace *commands go here *close and save log file log close	These commands show how to start and close a log file, which will save a text file of all the commands and output that appears on the command window in stata. The first line is short for "capture log close" which will close any open log files, and otherwise just proceed to the next step. Then the "log using lab3.log replace" starts the log file and changes the default in two ways. First, it changes the file type to have a .log file extension, which creates a plain text log file (which is readable in Gradesope so is important!). Second, it also adds the ", replace" option which will save over any other log file that has the same name. This is usually what you want. The rest of your lab code can go below the "log using lab3.log, replace" line. At the end of your do-file you can include the last line which is "log close" which will close and save the log-file.

TABLE 3
R Commands

R command	Description

#clear the workspace
rm(list=ls()) # removes all objects from the environment
cat('\014') # clears the console
 	
#Install and load haven package
if (!require(haven)) install.packages("haven"); library(haven)
 	
#Change working directory and load stata data set
setwd("C:/Users/gbruich/Ec 50/Lab 3")
mto <- read_dta("mto.dta")
	This sequence of commands shows how to open Stata datasets in R. The first block of code clears the work space. The second block of code installs and loads the "haven" package. The third block of code changes the working directory to the location of the data and loads in mto.dta. To change the working directory in R Studio, you can also use the drop down menu. Go to session -> set working directory -> choose working directory.

#Summary stats for one variable
mean(mto$yvar, na.rm=TRUE)
 	
#Summary stats for observations with treatment_group == 1
#Subset data
new_df <- subset(mto, treatment_group == 1)
 	
#Report mean
mean(new_df$yvar, na.rm=TRUE)
 	
#Alternatively, do it all at once using the with() function
with(subset(mto, treatment_group == 1), mean(yvar, na.rm=TRUE))
 	
#Summary stats for observations with treatment_group == 0
with(subset(mto, treatment_group == 0), mean(yvar, na.rm=TRUE))
 	
#Alternatively, get both means using tapply()
tapply(mto$yvar, mto$treatment_group, mean)
 	
#Alternatively, get both means using by()
by(mto$yvar, list(mto$treatment_group), mean)
	We used these commands in Lab 1. These commands report means for *yvar*. The first line calculates these statistics across the full sample.
	
	The other lines illustrate how to calculate these statistics for observations meeting certain criteria: when another variable in the data is equal to 1, or equal to 0.
	
	The first few examples use the subset() function to pick out only the observations in a data frame that meet certain criteria. We can combine this with the with() function. We also have seen how to use the tapply() function to report the mean of yvar grouped by another variable treatment_group. We can also use the by() function to do the same thing.

#Estimate linear regression
mod1 <- lm(yvar ~ treatment_group, data=mto)
mod1
	We used these commands in Lab 2. These commands report estimated regression coefficients from a regression of *yvar* on an intercept and a variable *treatment_group*. The intercept is always included by default, which is usually what you want.

#Bar graph
#Load tidyverse library
if (!require(tidyverse)) install.packages("tidyverse"); library(tidyverse)
 	
#Create a data frame with two columns
#Column 1 is the height of the two bars (in blue)
#Column 2 is the group names (in red)
df <- data.frame(c(0.1, 0.4),
c("Control group", "Treatment group"))
 	
# Change name of 1st column of df to "Moved" 	
names(df)[1] <- "Moved"
 	
# Change name of 2nd column of df to "Group" 	
names(df)[2] <- "Group"
 	
# Bar graph displaying results 	
ggplot(data=df, aes(x=Group, y=Moved, fill=Group)) +
geom_bar(stat="identity", show.legend = FALSE, width=.6) +
scale_fill_manual(values=c("red", "blue")) +
labs(y = "Moved Using Experimental Voucher", x = "")
 	
ggsave("fig1.png")
	These commands show how to draw and save a bar graph. We start by loading the tidyverse package.
	
	We will make a data frame with two observations and two columns. Column 1 is the height of the two bars (in blue font). Column 2 is the group names (in red font). For the first observation, fill in what the height should be for the control group bar: 0.1. For the second observation, we fill in what we want for the treatment group bar: 0.4.
	
	Then we give the first column the name "Moved" and the second column the name "Group" so that we can refer to these in the ggplot command.
	
	We use the geom_bar plot type in ggplot. The "identity" option says to plot the numbers in the data frame as is, as opposed to plotting some statistic computed for the data frame. The scale_fill_manual() code changes the color of the bars.