

### Empirical Project 3 The Creating Moves to Opportunity (CMTO) Experiment

Posted on Thursday, April 4, 2019

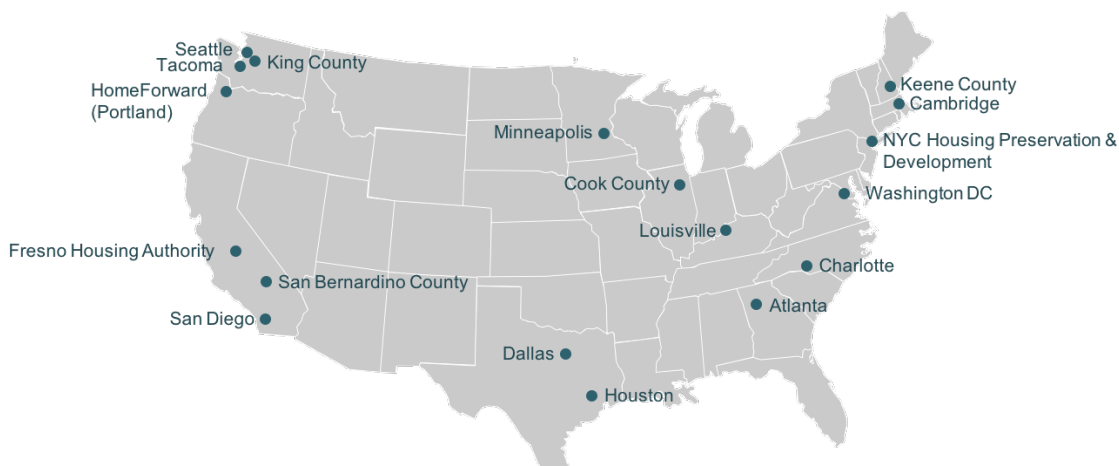
Due at midnight after the lecture on Thursday, April 18, 2019

In this empirical project, you will analyze experimental data from the ongoing [Creating Moves to Opportunity Experiment \(CMTO\)](#).<sup>1</sup> In particular, you will analyze data from pilot studies conducted in partnership with the King County Public Housing Authority (KCHA) and the Seattle Public Housing Authority (SHA).

Each Public Housing Authority (KCHA and SHA) handled recruitment into the experiment and randomization separately, although they followed the same procedure. Families with at least one child below age 15 who submitted applications for Section 8 housing vouchers were asked if they would like to participate in the experiment. Families who said yes were then randomly assigned to two groups: (i) a control group that received the standard services for families with a Section 8 voucher and (ii) an experimental treatment group that was offered an additional suite of services oriented to help families move to high-opportunity areas. These additional services included information about opportunity areas, assistance in identifying available units and in getting in contact with landlords, and some financial assistance for moving expenses. Importantly, treatment group families were not required to use these services and were not required to move to higher opportunity places if they did not want to do so. For more information about the study, see the websites for [Opportunity Insights](#), [Creating Moves to Opportunity](#), and the [Poverty Action Laboratory](#).

Your analysis of these data is at the cutting edge of affordable housing policy. CMTO is currently in the pilot phase in Seattle and King County, but conversations are now underway to expand it to many other cities across the U.S., shown in the map below.

#### Public Housing Authorities Planning to Participate in CMTO



---

<sup>1</sup> For the purposes of this assignment, you will work with a simulated dataset we have created that preserves the key features of the original data, but does not contain actual information from real households to protect their privacy.

## Instructions

Please submit your Empirical Project on Canvas. Your submission should include three files:

1. A 4-6 page research summary as a word or pdf document (double spaced and including references, graphs, and tables)
2. A do-file with your STATA code or an .R script file with your R code
3. A log file of your STATA or R output

## Specific questions to address in your research summary

1. (5 points) Explain why *pha* is the “assignment variable” in this study
2. (5 points) Provide evidence that the housing authorities really did randomly assign families to treatment and control groups. Similar to [Table 1 in Taubam et al. \(2014\)](#), please create a nicely formatted table that reports means of 5 or 6 relevant characteristics for families in the *control group*.

*Note:* Part of this question is to get you to think about which variables should be balanced in a randomized experiment. You need to read carefully through all the variables in Table 1 and decide which 5 or 6 you will summarize.

3. (5 points) For each of the variables you summarized above, calculate:
  - (i) the difference between the mean in the treatment group and the mean in the control group;
  - (ii) the standard error for the difference in means (with unequal variances).

Add these as columns two and three to the table you started in question 2.

*Note:* You should estimate the difference in means and its standard error using a multivariable regression that controls for the “assignment variable” *pha*.

4. (5 points) Is the balance table consistent with families having been randomly assigned to treatment and control groups by the Housing Authorities? Why or why not?
5. (5 points) Estimate the compliance rate for the CMTO experiment. That is, what is the effect of being assigned to the treatment group on the probability of receiving CMTO services?

*Hint:* For this question and question 7, you can use the same regression as in question 3, just changing the dependent variable.

6. (10 points) 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. 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. Explain why these two approaches would lead to biased estimates.

7. (5 points) What is the intent-to-treat (ITT) effect of the CMTO services on signing a lease in a high opportunity area? Estimate it in the data.
8. (10 points) What is the “treatment on the treated” effect (TOT) of the CMTO services on signing a lease in a high opportunity area, i.e. the effect among those who participated? Estimate it in the data and provide some intuition for the calculation of this estimate.
9. (10 points) Use the variables *forecast\_kravg30\_p25* and *origin\_forecast\_kravg30\_p25* to calculate a measure of the predicted change in childhood environment for each family in the data set. Produce a graph of the distribution of this new variable, plotting the data for treatment group and the control group separately. Explain what you observe.
10. (20 points) Is there evidence of heterogeneous treatment effects? Provide 4 different TOT estimates, splitting the sample into:
  - (i) family income greater than the median in the sample and less than (or equal to) the median in the sample;
  - (ii) Each Public Housing Authority separately (KCHA and SHA)

Compare these estimates. Why might the CMTO services differ for each of these groups?

11. (20 points) Suppose that you are submitting these results to a general interest journal such as *Science* for publication. Write an abstract of 100 or fewer words describing what you have found in your analysis of the CMTO data, similar to the abstract in [Taubam et al. \(2014\)](#).

## DATA DESCRIPTION, FILE: *cmto.dta*

The data consist of  $n = 820$  families in a simulated dataset that preserves the key features of the Creating Moves to Opportunity Experiment, but does not contain actual information from real households to protect their privacy.

**Table 1**  
**Variable Definitions in *cmto.dta***

Variable (1)	Definition (2)
<i>household ID</i>	Household identifier
<i>pha</i>	0 = King County Public Housing Authority 1 = Seattle Public Housing Authority
<i>treatment_group</i>	1 = if randomly assigned to treatment group 0 = if randomly assigned to control group
<i>received_cmto_services</i>	=1 if received at least some CMTO Services, = 0 if did not receive any CMTO Services
<i>total_time_meetings</i>	Hours spent in CMTO services (e.g., visiting housing locations)
<i>leased_up</i>	=1 if household signed lease, = 0 otherwise
<i>leased_up_opp</i>	Opportunity areas are census tracts with high upward mobility rates; tracts were chosen based on estimates of mean household income of children from low income families, as well as practical considerations for PHA operation. =1 if household signed lease in high opportunity area. =0 otherwise
<i>forecast_kravg30_p25</i>	Long-run earnings for low-income children in destination Census tract
<i>origin_forecast_kravg30_p25</i>	Long-run earnings for low-income children in origin tract
<i>hoh_age</i>	Head of household's age (at baseline)
<i>child_count</i>	Number of children (at baseline)
<i>child_age</i>	Average age of children in the household (at baseline)
<i>speaks_english</i>	= 1 if head of household speaks English (at baseline), = 0 otherwise
<i>born_abroad</i>	= 1 if head of household was born abroad, = 0 if born in the U.S.
<i>working</i>	= 1 if head of household working (at baseline), = 0 otherwise
<i>homeless</i>	1 = if homeless at baseline, 0 otherwise
<i>hh_income</i>	Household income (\$) at baseline
<i>origin_pop2010</i>	Origin tract population in 2010 Census
<i>black</i>	1= race is African American/Black, 0 = otherwise
<i>white</i>	1= race is White, 0 = otherwise
<i>asian</i>	1= race is Asian, 0 = otherwise
<i>latino</i>	1= race is Latino, 0 = otherwise
<i>race_other</i>	1= race is other, 0 = otherwise
<i>less_hs</i>	1= Less than a high school education, 0 = otherwise, . = missing
<i>college_plus</i>	1 = education is college degree or above 0 = otherwise, and . = missing education

*Note:* This table describes the variables included in *cmto.dta*.

**Table 2a**  
**STATA Commands**

STATA command	Description
<code>sum xvar1 xvar2 xvar3 if treatment_group == 0</code>	Reports summary statistics for <i>xvar1</i> , <i>xvar2</i> , and <i>xvar3</i> for observations with <i>treatment_group</i> equal to 0
<code>reg yvar zvar wvar, r</code>	Estimates multivariate regression of <i>yvar</i> on an intercept, <i>zvar</i> , and <i>wvar</i> , with heteroskedasticity-robust standard errors.
<pre>*Method 1 reg xvar zvar wvar, r local pi = _b[zvar]  reg yvar zvar wvar, r local alpha = _b[zvar]  display `alpha' / `pi'  *Method 2 ivregress 2sls yvar (xvar = zvar) wvar, r</pre>	<p>These commands show how to estimate the ratio between the coefficient on <i>zvar</i> from two different regressions that have all the same right-hand side variables, and different dependent variables.</p> <p>The first method stores the coefficients as local variables called <i>pi</i> and <i>alpha</i> and then “displays” the ratio between them.</p> <p>The second method uses the <code>ivregress 2sls</code> command to estimate the ratio in one step.</p>
<code>sum xvar, d</code>	Reports detailed summary statistics for <i>xvar1</i> , including the mean, median, and other quantiles.
<code>gen difference = xvar1 - xvar2</code>	Generates a new variable that equals the difference between <i>xvar1</i> and <i>xvar2</i>
<pre>reg yvar zvar wvar if dvar &lt;= 5, r ivregress 2sls yvar (xvar = zvar) wvar if dvar &lt;= 5, r</pre>	These commands show how to estimate the previous regressions, but now restricting the sample to observations with <i>dvar</i> less than or equal to 5
<pre>*Command all goes on one line twoway (kdensity xvar if treatment_group == 1) (kdensity xvar if treatment_group == 0)  *save graph graph export density.png, replace</pre>	<p>These commands show how to graph a <a href="#">kernel density</a> estimate of the density of <i>xvar</i> for observations with <i>treatment_group</i> equal to 1 and <i>treatment_group</i> equal to 0. See <a href="#">this tutorial</a> on graphs in stata for more guidance on how to use the drop down menu to produce graphs. The last line saves the graph as a <code>png</code> file.</p>

**Table 2b**  
**R Commands**

<b>R command</b>	<b>Description</b>
<pre>*Subset data cmto_cntrl &lt;- subset(cmto, treatment_group==0, select = c(xvar1, xvar2, xvar3))  *Report summary statistics summary(cmto_cntrl)</pre>	Subsets the data to the variables <i>xvar1</i> , <i>xvar2</i> , and <i>xvar3</i> for observations with <i>treatment_group</i> equal to 0. Reports summary statistics for this data frame.
<pre>#Install and load sandwich and lmtest packages install.packages("sandwich") install.packages("lmtest") library(sandwich) library(lmtest)  #Regression with homoskedasticity-only standard errors mod1 &lt;- lm(yvar~zvar1+wvar, data = cmto) summary(mod1)  #Report heteroskedasticity robust standard errors coeftest(mod1, vcov = vcovHC(mod1, type="HC1"))</pre>	Estimates multivariate regression of <i>yvar</i> on an intercept, <i>zvar</i> , and <i>wvar</i> , with heteroskedasticity-robust standard errors.
<pre>*Method 1 mod1 &lt;-lm(yvar ~ zvar+wvar, data = cmto ) mod2 &lt;-lm(xvar ~ zvar+wvar , data = cmto )  mod1\$coef[2]/mod2\$coef[2]  *Method 2 install.packages("AER") library(AER) tot &lt;- ivreg(yvar ~ xvar + wvar               zvar + wvar ,data=cmto)  summary(tot, vcov = sandwich,         df = Inf, diagnostics = TRUE)</pre>	<p>These commands show how to estimate the ratio between the coefficient on <i>zvar</i> from two different regressions that have all the same right-hand side variables, and different dependent variables.</p> <p>The first method refers to the elements of the vector <code>coef</code> that contains the relevant coefficients and displays the ratio between them.</p> <p>The second method uses the <code>ivreg</code> command to estimate the ratio in one step.</p>
<pre>cmto\$difference &lt;- cmto\$xvar1 - cmto\$xvar2</pre>	Generates a new variable that equals the difference between <i>xvar1</i> and <i>xvar2</i>
<pre>library(ggplot2)  ggplot(data = cmto, aes(x=xvar)) + geom_line(stat = "density", aes(group = treatment_group, colour=as.factor(treatment_group)), size = 1) +   labs(color = "Treatment Status")  ggsave("density.png")</pre>	These commands show how to graph an kernel density estimate of the density of <i>xvar</i> for observations with <i>treatment_group</i> equal to 1 and <i>treatment_group</i> equal to 0. The last line saves the graph as a png file.