Data and Methods Appendix for "Can Variation in Subgroups' Average Treatment Effects Explain Treatment Effect Heterogeneity? Evidence from a Social Experiment"

Marianne P. Bitler
University of California, Davis and NBER

Jonah B. Gelbach
University of Pennsylvania Law School

Hilary W. Hoynes
University of California, Berkeley and NBER

February 23, 2017

1 Files referenced in this document

All files referenced in this document, other than the raw MDRC datasets (see Section **??**) are included in the zip file "bgh_subgroups.zip". The contents are as follows:

1. **Stata do file "bgh_replication_master.do."** This file builds all numerical results referenced in the paper, as well as files in the dofiles folder needed to produce:

   (a) Figure 1
   (b) Figures 2 and 3
   (c) Table 1
   (d) Table 2
   (e) Online Appendix Tables 1, 2, 3, and 4

   In order to replicate our numbers in Table 2, the user needs to use the same seed (`set seed 10203`) and must run through all of the tests in the loop (that is, the results are sensitive to the order the tests are carried out as the seed is updated by Stata as it runs through the loops). These procedures do not make much substantive difference, but are necessary for exact replication.

2. **data** This folder contains "createdata.do", the Stata do file necessary to create the final data from the original MDRC data. Before running this file, you must convert the ctadmrec.sas7bdat file from MDRC into dta form. We created the Stata dta file using Stat/Transfer.

3. **ado** This folder contains the following ado files:

   (a) *graphsharecdfsumsto.ado* This file graphs QTEs by group, and share for both treatment and control groups. This file is used to produce online Appendix Figures 1 and 2.
   (b) *graphsubqte.ado* This file graphs a CDF, and is used to produce Figure 1 and online Appendix Figure 3.
   (c) *jcumul.ado.* This file provides a slight modification of Stata's *cumul* command. This file is called by *qte.ado*, but this option is not used.
   (d) *qte.ado.* This file uses *jcumul* to estimate the CDF of a variable by value a dummy variable, *category*. It then computes quantiles of the estimated distributions, and reports these quantiles with Stata equation name "c" (for *category* == 0) and "t" (for *category* == 1). To produce QTE estimates, it differences the quantiles, and reports these QTE estimates with Stata equation name "d".
   (e) *sgnullsubsampleznoweight.ado.* This file simulates earnings by subgroup. This file is used to create the graphs by actual and synthetic earnings (Figures 2, 3, online Appendix Figure 4).

4. **dofiles** With the exception of "createdata.do", all do files called by "bgh_replication_master.do" are included in this folder. "createdata.do" must be called before creating any figure or table. In addition, "bgh_fig2and3.do" must be called before creating online Appendix Table 4.

5. **graphs** After running "bgh_replication_master.do," this folder will contain the output for Figures 1-3 and online Appendix Figures 1-4. Figure 1 and the online appendix figures are produced in both black and white and color; Figures 2 and 3 are produced in color. Graphs are saved to the "graphs" directory in gph, ps, and eps form. With ps2pdf, graphs are also saved as pdfs.

6. **results** After running "bgh_replication_master.do," this folder will contain the main results, including a log file of the full file "bgh_replication_master.do". All values in Table 1 and Table 2 are included in the log file.

2 Data Access

We cannot directly provide the data used: MDRC requires an application process for researchers who wish to gain access to the public-use Jobs First data. In this document and the associated files, we provide Stata code that allows researchers to replicate our results given access to these data. Details of MDRC's policy and the public-use data are available from `www.mdrc.org/public-use-data-files`.

After creating the datafile "ctadmrec.dta", running the *bgh_replication_master.do* file in this folder will produce all results, including figures and tables.

2.1 Key variables

Below are some key variable names and formats.

1. *Race:* "white" and "black" indicate race for non-Hispanic recipients; "hispanic" indicates Hispanic ethnicity, and "othmsrac" indicates neither white nor black race and non-Hispanic ethnicity.

2. *Age:* Age is divided into three categories from the categorical variables "agelt20", "age2024", "age2534", "age3544", and "agege45" in the MDRC file. We create "agelt25" (younger than 25) and "agegt34" (older than 34) from these variables.

3. *Assignment to treatment:* The dummy variable "e" equals one if the respondent was assigned to treatment.

4. *Education:* "nohsged" indicates less than a high school education or GED; "hsged" indicates no more than a high school diploma or equivalent; "mthsgrad" indicates some education beyond high school (including college graduates). "missnohsged" and "misshsged" indicate these variables are missing.

5. *Marital status:* "marnvr" indicates never married, "marapt" indicates divorced, widowed, or separated respondents.

6. *Number of children:* "kidctgt2" indicates a respondent has more than two children; "kidctle2" indicates a respondent has two or fewer children.

7. *AFDC and Food Stamp receipt:* "adcpqX" and "fstpqX" denote AFDC and Food Stamp receipt prior to quarter X, respectively. These variables are defined for quarters 1 through 7 before random assignment. "anyadcpqX" and "anyfstpqX" denote whether the respondent received any AFDC or Food Stamp benefits X quarters prior to random assignment.

8. *Earnings:*. Akin to measures of AFDC and Food Stamp receipt, "ernpqX" denotes earnings prior to quarter X. We also define "anyernpqX" as whether the respondent had any earnings prior to quarter X.

3 Other results

The do file "bgh_sec2_3.do" produces the results used to split the sample into three groups (no earnings seven quarters prior to random assignment, earnings above the median, and earnings below the median). Results in Section 3.1 are displayed in Table 1, and results in Section 5.4 are displayed in Table 2.