DATA AND PROGRAMS FOR

Effective Policy for Reducing Inequality:

The Earned Income Tax Credit and Distribution of Income

Hilary W. Hoynes, hoynes@berkeley.edu

Ankur J. Patel, ankur.patel@treasury.gov

Updated June 2022

The data and programs contained within these directories replicate all tables and figures from the main draft and online appendix. The STATA do-files were last run on a desktop machine running Windows 7 64-bit and STATA 13.

Directory structure:

dofiles – STATA programs for generating tables and figures.

data_working – Final datasets and intermediate input datasets.

STATA Programs: \dofiles

1) 1_SingleWomen_ReadInCps.do / .log
    a) Imports IPUMS-CPS into STATA format.
    b) Constructs a sample of single women and their children (if present) with income and demographic information.

c) Calculates federal and state income taxes as well as payroll taxes using NBER-TAXSIM.

d) 1A_housing_merge.do

i) Prepares variables not found in IPUMS CPS (housing subsidy and school lunch value) to be merged into the samples.

2) 2_SingleWomen_simCredit.do / .log

a) Constructs simulated federal and state income tax (total tax, child tax credit, AMT, federal and state EITC) and transfer values (AFDC/TANF cash benefits). Simulated variables are calculated three ways, using a uniform distribution of earnings, and using income and demographic information from the 1983 and 1993 CPS.

3) 3_SingleWomen_regPrep.do / .log

a) Merges sample of single women to simulated variables, AFDC waiver variables, Federal Poverty Line and missing housing and school lunch values. Constructs some of the variables used as controls in following program.

4) 4_SingleWomen_analysis.do / .log

a) Produces results for: Main paper tables 1-4; figures 3-9; Online appendix figure 3a, 5-7, 9; tables 2-4, 6-8.

b) Restricts sample of single women to final sample.

c) Constructs rest of control and outcome variables.

d) Specifies STATA locals that describe various control sets, subsamples and experimental designs.

e) The program regHelper produces a matrix of the reduced form estimate, the 2SLS estimate, the percent impact, the extensive margin elasticity, and the mean of the dependent variable for a specified outcome and experimental design (equation 1 of the

main paper). This program is used to estimate main paper tables 1-4, main table figures 6-8, online appendix table 3, 4, 6, and online appendix figure 8.

   f) The program eventStudy, with the help of the program regHelper3, estimates the event time model specified in equation 2 of the main paper. This program is used to produce main paper figures 3a, 3b, 4a, 4b, 5, and appendix figures 5, 6, 7, 9.

   g) Estimates the numbers removed from poverty presented in figure 9 of the main paper.

5) 5_SingleRestrictions_ReadInCps.do / .log
   a) Constructs sample of single women, as in 1_SingleWomen_ReadInCps.do, with age, disability and school attendance, and education restrictions removed.

6) 6_SingleRestrictions_regPrep.do / .log
   a) Like 3_SingleWomen_regPrep.do with restrictions removed.

7) 7_SingleRestrictions_analysis.do / .log
   a) Estimates online appendix table 5 using similar structure to 4_SingleWomen_analysis.do.

8) 8_MarriedCouples_readInCps.do / .log
   a) Constructs a sample of married couple and their children with demographic and income information. Calculates federal and state income taxes and employment taxes using NBER TAXSIM.

9) 9_MarriedCouples_simCredit.do / .log
   a) Constructs simulated information for the sample of married couples.

10) 10_MarriedCouples_regPrep.do / .log
   a) Performs a similar function as 3_SingleWomen_regPrep.do.

11) 11_MarriedCouples_analysis.do / .log

a) Estimates online appendix figure 3b, table 8 using a procedure similar to 4_SingleWomen_analysis.do.

Additional datasets and programs: \data_working

1) The program 7za.exe is a file archive program (https://www.7-zip.org/).
2) The STATA dataset afdcparameters.dta contains AFDC and TANF parameters from Hoynes and Luttmer (2011), see online appendix for more details.
3) The STATA dataset afdcwaivers.dta contains information regarding state AFDC waivers, see online appendix for more details (http://aspe.hhs.gov/hsp/isp/waiver2/waivers.htm; http://aspe.hhs.gov/hsp/waiver-policies99/policy_CEA.htm).
4) The program taxsim9 and ankurs_taxsim.ado are collectively the NBER TAXSIM program (http://users.nber.org/~taxsim/).
5) The STATA dataset cpi.dta contains CPI-U from BLS (https://www.bls.gov/cpi/).
6) The STATA dataset state_level_variables_from_UKPRC.dta contains state level unemployment and population variables from the University of Kentucky Center for Poverty Research National Welfare Data (http://www.ukcpr.org/data).
7) The STATA dataset FPL.dta contains Census Poverty Thresholds by family size (https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html).

IPUMS CPS

The primary data used in the analysis is publicly available and was pulled from IPUMS CPS (https://cps.ipums.org/cps/).

IPUMS Variable List

Variable	Label
year	Survey year 1982 – 2013
serial	Household serial number
numprec	Number of person records following
hwtsupp	Household weight, Supplement
gq	Group Quarters status
hhintype	Type of household
region	Region and division
statefip	State (FIPS code)
metro	Metropolitan central city status
metarea	Metropolitan area
county	FIPS county code
ownershp	Ownership of dwelling
hhincome	Total household income
pubhous	Living in public housing
rentsub	Paying lower rent due to government subsidy
heatsub	Received energy subsidy
heatval	Value of energy subsidy
foodstmp	Food stamp recipiency
stampval	Total value of food stamps
lunchsub	Government school lunch food subsidy
nfams	Number of families in household
nmothers	Number of mothers in household
nfathers	Number of fathers in household
month	Month
pernum	Person number in sample unit
spmpov	SPM unit's poverty status
spmwt	SPM unit's weight
spmfedtaxac	SPM unit's federal tax
spmeitc	SPM unit's federal EITC
spmfedtaxbc	SPM unit's federal tax before the EITC
spmtotres	Total SPM resources for SPM unit
spmthresh	SPM unit's threshold
spmfamunit	SPM unit's ID

wtsupp	Supplement Weight
perwt04	Person weight 2004
lineno	Line number on original form
momloc	Mother's location in the household
stepmom	Probable step/adopted mother
momrule	Rule for linking mother
poploc	Father's location in the household
steppop	Probable step/adopted father
poprule	Rule for linking father
sploc	Spouse's location in household
sprule	Rule for linking spouse
famsize	Number of own family members in hh
nchild	Number of own children in household
nchlt5	Number of own children under age 5 in hh
famunit	Family unit membership
eldch	Age of eldest own child in household
yngch	Age of youngest own child in household
nsibs	Number of own siblings in household
aspouse	Spouse line number (self-reported)
relate	Relationship to household head
age	Age
sex	Sex
race	Race
marst	Marital status
popstat	Adult civilian, armed forces, or child
yrimmig	Year of immigration
mbpl	Mother's birthplace
fbpl	Father's birthplace
nativity	Foreign birthplace or parentage
hispan	Hispanic origin
educ	Educational attainment recode
higrade	Highest grade of school
educ99	Educational attainment, 1990
schlcoll	School or college attendance
empstat	Employment status
labforce	Labor force status
occ	Occupation
ind	Industry
occly	Occupation last year
indly	Industry last year
wkswork1	Weeks worked last year
hrswork	Hours worked last week
uhrswork	Usual hours worked per week (last yr)

wksunem1	Weeks unemployed last year
absent	Absent from work last week
durunem2	Continuous weeks unemployed, intervalled
fullpart	Worked full or part time last year
nwlookwk	Weeks looked for work last year (didn't work)
hourwage	Hourly wage
paidhour	Paid by the hour
pension	Pension plan at work
union	Union membership
whyunemp	Reason for unemployment
firmsize	Number of employees
whyabsnt	Reason for absence from work
wantjob	Want regular job now
whyptly	Reason for working part-time last year
whyptlwk	Reason for working part time last week
usftptlw	Usually work full time (part time last week)
numemps	Number of employers last year
wnlwnilf	When last worked for pay (NILF last week)
whynwly	Reason not working last year
actnlfly	Activity when not in labor force last year (part-year workers)
ptweeks	Weeks working part time last year
ftotval	Total family income
inctot	Total personal income
incwage	Wage and salary income
incbus	Non-farm business income
incfarm	Farm income
incss	Social Security income
incwelfr	Welfare (public assistance) income
incgov	Income from other govt programs
incaloth	Income from alimony, contributions, other
incretir	Retirement income
incssi	Income from SSI
incdrt	Income from dividends, rent, trusts
incint	Income from interest
incunemp	Income from unemployment benefits
incwkcom	Income from worker's compensation
incvet	Income from veteran's benefits
incsurv	Income from survivor's benefits
incdisab	Income from disability benefits
incdivid	Income from dividends
incrent	Income from rent
inceduc	Income from educational assistance

incchild	Income from child support
incalim	Income from alimony
incasist	Income from assistance
incother	Income from other Source not specified
earnweek	Weekly earnings
incdisa1	Disability income from first source
incdisa2	Disability income from second source
inclongj	Earnings from longest job
increti1	Retirement income from first source
increti2	Retirement income from second source
incsurv1	Survivor benefits income from first source
incsurv2	Survivor benefits income from second source
mthwelfr	Number of months received welfare income
offpov	Official Poverty Status (IPUMS constructed)
offpovuniv	Official Poverty Rate Universe
offtotval	Total Family Income for Replicating Official Poverty Rates
offcutoff	Official Poverty Rate Cutoff
offreason	Reason for Discrepancy between OFFPOV and POVERTY
poverty	Original poverty status (PUMS original)
cutoff	Cutoff for original poverty status, in dollars
vetstat	Veteran status
vetlast	Veteran's most recent period of service
qage	Data quality flag for AGE
qmarst	Data quality flag for MARST
qrace	Data quality flag for RACE
qrelate	Data quality flag for RELATE
qsex	Data quality flag for SEX
qhispan	Data quality flag for HISPAN
qeduc	Data quality flag for EDUC
qhigrade	Data quality flag for HIGRADE
ftype	Family Type
qincwage	Data quality flag for INCWAGE