**Replication Instructions**
**Consumption and Income Inequality in the U.S. Since the 1960s**
**Bruce Meyer and Jim Sullivan**
**June 2022**

The document provides the guidelines for replicating the results presented in "Consumption and Income Inequality in the U.S. Since the 1960s," *Journal of Political Economy*. Programs are provided to produce these results starting from the raw data made available from the Consumer Expenditure Survey (CE) or the Current Population Survey (CPS). In addition, for the analyses using CE data, an already cleaned version of the data is made available so one can replicate the results without having to start from the raw data. These replication instructions are organized into three parts:

- File and program structure
- Replicating final results
- Producing cleaned CE data files from the raw data

## I. Replication data files and programs

**File Structure**

| | |
|---|---|
| ~\cleaned_data: | cleaned SAS or Stata dataset |
| ~\programs\Consumption results: | all SAS or Stata programs to produce main consumption results |
| ~\programs\Income results: | all SAS or Stata programs to produce main income results |
| ~\demand_system: | all files to replicate our demand system estimates |
| ~\Produce main CE data: | all files to generate the cleaned, main CE dataset |
| ~\programs\invoked_files | Additional code or ancillary data invoked by primary programs |

In addition to these data and programs, we provide an Excel file, Tables_and_Figures.xlsx, that includes the tables and figures from the paper so one can see how the output from these programs is configured to produce the results presented in the paper.

**Cleaned data files:**
all_cons_data_for_stata_80_17.dta.gz (cleaned CE Stata data file for main results, zipped)
cleaned_sas_data.zip (zip folder that contains allYYYY.sas7bdat—cleaned SAS data files for certain years YYYY)

**Raw data files:**
Raw CE data files are available from the Bureau of Labor statistics at https://www.bls.gov/cex/pumd_data.htm, for years 1980-2017. Raw data for earlier years are available through the ICPSR (www.icpsr.umich.edu).

Raw CPS income files are available for recent years at
https://www.census.gov/data/datasets/time-series/demo/cps/cps-asec.html
Raw CPS data files for earlier years are available through the NBER (
https://www.nber.org/research/data/current-population-survey-cps-supplements-annual-demographic-file).

## II. Producing main results

### a. Main consumption results using the CE

<Datasets and Programs Required>
See: ~\programs\Consumption results
Do file: MASTER_61_17_MS.do, cons_inequality_data_61_17_MS.do,
cons_inequality_data_predhome_61_17_MS.do, qregressions_new_17.do
Stata data file: all_cons_data_for_stata_80_17.dta
Excel file: price_indices17_JH.xlsx, cons_inequality_format_MS.xlsx

#### a.1. Producing Data for Columns AV to DJ of sheet "RAW RESULTS OUTPUT"

Step 1. Run MASTER_61_17_MS.do, which in turn runs three other do files (qregressions_new_17.do, cons_inequality_data_61_17_MS.do, cons_inequality_data_predhome_61_17_MS.do)

Step 2. Open the output file "cons_inequality_data_adj.xls" and copy all data (columns B-BC) and paste them into column B2 of sheet "Raw_Data" in "cons_inequality_format_MS.xlsx".

Step 3. Open the output file "cons_inequality_data_updated60.xls", copy data for year 1961, and paste them into Row 2 of sheet "Raw_Data" in "cons_inequality_format_MS.xlsx".

Step 4. Open the output file "cons_inequality_data_updated80.xls", copy all data for years 1980 and 1981, and paste them into Rows 5-6 of sheet "Raw_Data" in "cons_inequality_format_MS.xlsx".

Step 5. Sheet "Formatted Data" of "cons_inequality_format_MS.xlsx" contains data for Columns AV-DJ of sheet "RAW RESULTS OUTPUT".

#### a.2. Producing Data for Columns O to AC of sheet "TA3 Single moms"

Step 1. Run MASTER_61_17_MS.do.

Step 2. The output file "cons5_inequality_by_group.xls" contains data in Columns O to AC of sheet "TA3 Single moms" except for years 1961, 1980, and 1981.

Step 3. The output file "cons5_inequality_by_group_61_16_updated60.xls" contains the 1961 data, while the output file "cons5_inequality_by_group_61_16_updated80.xls" contains the 1980-81 data for Columns O to AC of sheet "TA3 Single moms"

### a.3. Producing Decomposition Results for Consumption

Run data_for_decomp.do
Run melly_decomp_cons6.do

### a.4. Producing Elasticity Estimates

To replicate the elasticity estimates in Table 1, run the files below.

a. cons_data_for_elasticities.sas (This code invokes the SAS dataset cons_final60_12, which is provided in the subdirectory, invoked_files. Unzip this SAS dataset first.)
b. MASTER_readme_first_updated_v4.do
c. elasticities_w_qreg_pred.do

### a.5. Producing Estimates for Consumption Growth by Asset Quintile and Consumption Shares

Running cons_inequality_data_61_17_MS.do (which is done automatically when you run MASTER_61_17_MS.do) generates an excel file "quintile_sort_on_asset" which contains data for columns P-T in "T6 Cons chgs by asset quint" (Table 6 in Tables_and_Figures.xlsx). It also generates an excel file "quintile_sort_on_cons6" which contains data for columns K-O in "TA.1 means_core_total" (for Appendix Tables A.1 and A.2 in Tables_and_Figures.xlsx), except for year 1980. Running "cons_inequality_data_predhome_61_17_MS.do" generates an excel file "quintile_sort_on_cons6_80" which contains data for year 1980.

## b. Main Income Results using the CPS

Note: YYYY represents the four-digit year
<Datasets and Programs Required>
See: ~\programs\Income results
Raw data files (the names for the raw CPS data files used):
2006-2018 CPS ASEC data (asecYYYY_pubuse.dat), mar64.raw, mar65.raw, mar66.raw, mar67.raw, 07559-0001-Data.txt, 07560-0001-Data.txt, 07561-0001-Data.txt, 07562-0001-Data.txt, 07563-0001-Data.txt, all1973.sas7bdat, …, all2005.sas7bdat. Note all1973.sas7bdat, …, all2005.sas7bdat are slightly modified versions of raw CPS data. These data files are included in ~\cleaned_data.

SAS file: *CPS_1992-2018_JH.sas, create_for_taxsim8817.sas, poverty_clean92_18.sas, mkdat6418.sas, inequality_64_18.sas, CPS 1964-1967 (68-72) Poverty Read (CPS Utilities)_JH.sas, CPS 1964-1967 (68-72) Taxsim - Federal Taxes (CPS Utilties)_JH.sas, merge_st_tax6477_JH.sas, create_for_taxsim7375_JH.sas, create_for_taxsim7677_JH.sas, create_for_taxsim7887_JH.sas, poverty_clean64_67.sas, poverty_clean68_72.sas, poverty_clean73_75_JH.sas, poverty_clean76_77_JH.sas, poverty_clean78_79_JH.sas, poverty_clean80_87_JH.sas, poverty_clean88_91_JH.sas*

There are five output files (*ineq_6479.dat, ineq_8089.dat, ineq_9099.dat, ineq_0018.dat, and ineq_6417_by_grp.dat)* that contain data for income inequality results. We first describe how to produce the output files, and then explain what data each output file contains.

**b.1 Road Map for producing the five output files**
1) Construct *povertyYYYY.sas7bdat* from the raw or cleaned CPS ASEC data
2) Construct *inequality_YYYY.sas7bdat* from *povertyYYYY.sas7bdat*
3) Generate output files *ineq_YYYY.dat and ineq_6417_by_grp.dat* from *inequality_YYYY.sas7bdat*

**Construct povertyYYYY.sas7bdat for survey years 2006-2018**

Step 1. Download the CPS ASEC data "asec*YYYY_pubuse.dat*" from the US Census website. Save the data in the folder "C:\Poverty\cpsdata"

Step 2. Run the SAS program *CPS_1992-2018.sas* to obtain the data file *allYYYY.sas7bdat*.

Step 3. Run the SAS program *create_for_taxsim8817.sas* to obtain *taxsYYYY.sas7bdat*.

Step 4. Run the SAS program *poverty_clean92_18.sas* to obtain *povertyYYYY.sas7bdat*

**Constructing *povertyYYYY.sas7bdat* for survey years 1964-1972**

Step 1. Run the following SAS programs in the folder "SAS Programs CPS 64-67 (68-72)"
a. CPS 1964-1967 (68-72) Poverty Read (CPS Utilities)_JH
b. CPS 1964-1967 (68-72) Taxsim - Federal Taxes (CPS Utilties)_JH

Step 2. Run "merge_st_tax6477_JH.sas"

Step 3. Run "poverty_clean64_67", "poverty_clean68_72" to obtain *poverty.YYYY.sas7bdat*

**Constructing *povertyYYYY.sas7bdat* for survey years 1973-2005**

Note: For these years, the input files *allYYYY.sas* are provided in ~\cleaned_data. See Section III for details.

Step 1. Run "create_for_taxsim7375_JH", "create_for_taxsim7677_JH", "create_for_taxsim7887_JH", and "create_for_taxsim88##_JH"

Step 2. Run "merge_st_tax6477_JH.sas"

Step 3. Run "poverty_clean73_75_JH.sas", "poverty_clean78_79_JH.sas", "poverty_clean80_87_JH.sas", "poverty_clean88_91_JH.sas", and "poverty_clean92_18_JH.sas" to obtain *poverty.YYYY.sas7bdat*

**Constructing *inequality_YYYY.sas7bdat***

Run the SAS program mkdat6418.sas. The resulting data file will be saved as *inequality_6479, inequality_8089, inequality_9099,* and *inequality_0018*.

**Generating *ineq_YYYY.dat and ineq_6417_by_grp.dat***

Run the SAS program *inequality_64_18.sas*. The resulting data file will be saved as *ineq_6479, ineq_8089, ineq_9099,* and *ineq_0018, and ineq_6417_by_grp*

**Producing Data for Columns BA to FF of sheet "FA.2 inc_90_10"**

The output file "*ineq_6479.dat"* contains data for years 1963-1978 (Rows 4-19)
The output file "*ineq_8089.dat"* contains data for years 1979-1988 (Rows 20-29)
The output file "*ineq_9099.dat"* contains data for years 1989-1998 (Rows 30-39)
The output file "*ineq_0018.dat"* contains data for years 1999-2017 (Rows 40-58)

**Producing Data for Columns AG to AR of sheet "TA3 Single moms"**

The output file "*ineq_6417_by_grp.dat"* contains data for Columns AG-AR

**b.2 Producing Decomposition Results for Income**

Run melly_decomp_inc5.do, which generates the income decomposition results reported in Table 5 of Tables_and_Figures.xlsx. Note that this program uses the data set melly6317.dat, which was created when you ran inequality_64_18_JH.sas above.

**c. Demand system estimates**

The do-files in folder ~\demand_system are modifications of the analysis in Aguiar and Bils (2015), publicly available at https://www.aeaweb.org/articles?id=10.1257/aer.20120599. Analysis uses the .dta file "CE_data" which is also available at this link.

The primary modification is to estimate income inequality using Aguiar and Bils' methods, but restricting the data to only include certain consumption categories.

Do-Files:

Master.do
Define_Extensions.do
Bootstrap_Extensions.do
Bootstrap_Mismeasurement.do

Output:

Note, aguiarbils_tables_master.xlsx links to all the Excel output generated by these .do files, which shows how these output files feed into the results presented in the paper. This Excel file has three main, formatted tabs, and a few tabs with raw data.

The first tab, "Consumption Category Summary," links to the tab "Raw Summary" which is also in the master file. "Raw Summary" in turn links to an excel sheet called CatSum.xlsx. CatSum is generated using the do-file Summary_Statistics.

The second tab, "Table 3 Replication & Extension," links to the 8 excel files created by the do-file Bootstrap_Extensions.do which are named, e.g, bs_col1.xls

The third tab, "Income-Specific Mismeasurement," is like the first tab in that it links to a raw data tab called "Raw Mismeasurement" which is in the master file. "Raw Mismeasurement," in turn, links to 8 excel files named, e.g, bs_mismes_col1.xls, which are created by the do-file Bootstrap_Mismeasurement.do

Running the do-file "Master.do" will generate all necessary raw output files to produce the demand system estimates in Table 4 of the paper.

Running the do-file " Summary_Statistics.do" will generate the statistics reported in Appendix Table A.6 of the paper.

## III. Producing cleaned CE data files from the raw data

Several steps are taken to process the raw CE data files from the BLS and create the master data set from which our results are produced. We outline those steps here. The programs are available in ~\Produce main CE data. For more details on procedures see methods describe at the ICPSR data warehouse.

1. Download raw from the CE or ICPSR. Edit the directory paths in each of the programs to match the location of these raw data.

2. Run the following programs in the following order. Note some of these programs will invoke other code or ancillary data. These invoked files are located in ~\invoked_files
    a. famstep7273.sas
    b. famstep6061.sas
    c. fix_food_shelt_trans84_86.sas
    d. fix_owndwell_8486.sas
    e. fix_fstamps84.sas
    f. fix_retire85.sas
    g. wgt_agg_veh_cloth80_14.sas
    h. famstep80_00.sas
    i. famstep01_17.sas
    j. health80_17.sas
    k. mkconsdat80_17.sas

l. mkmembdata80_17.sas
m. predict_rent2017.do
n. Master Data80_17.sas
o. regs_predict80_17.sas
p. merge_all_to_master80_17.sas
q. debt80_13.sas
r. debt_housing80_17.sas
s. utilities80_17.sas
t. poverty_trends_health60_17.sas
u. inequality60_17.sas

This last program creates all_cons_data_for_stata_80_17.dta, which is the main dataset used to produce consumption inequality results as explained above.