This README documents how to reproduce the results in Raven Molloy’s and Abigail Wozniak’s paper “Labor Reallocation over the Business Cycle: New Evidence from Internal Migration,” published in the Journal of Labor Economics (2011, 29(4): 697-739).

**Repository Structure**

*The repository is structured as follows:*

* **\Table\_#** or **\Figure\_#**
  + These folders contain the datasets and do-files for the specific Table(s) and/or Figure(s) referenced in the folder name.
* **\Data\_Setup**
  + This folder contains the materials needed to replicate the datasets used in constructing Table 1 and Table 5, 6, and A3.

**Dataset Construction**

*Steps for replicating the main datasets:*

* **Table 1 (table1.dta):** See Appendix for construction of rows 1 and 2.
* **Tables 2-4:** Not provided.
* **Tables 5 & 6 (marchcps\_mig.dta)**
  1. Extract desired variables from IPUMS CPS site. Note that the version of the IPUMS CPS March microdata used in the paper is called ipums6.do, which contains the complete list of variables downloaded from the IPUMS site. The site was accessed on June 10, 2010. There was no case or flag selection applied to the extract.
  2. From IPUMS downloads site, copy the do file (renamed ipums#.do) to this folder.
  3. ipums#.do is referenced in make\_marchcps.do, which creates marchcps\_mig.dta and marchcps\_mig\_test.dta.
  4. makestatecycle\_ipums#.do creates statecycle\_ipums.dta from the latest versions of Raven's data (bartik\_s10.dta and stateemp3.dta) by merging it into the appropriate state codes (ipumscps\_statecodes.dta) for matching with IPUMS data by year from 1962 on.
  5. If adding new cycle information, run makestatecycle\_ipums#.do first. Then rerun make\_marchcps.do.
  6. If adding new CPS variables, just run ipums#.do within make\_marchcps.dta.

**Tables 1-6, Figures 1-3, and Appendix Materials Replication**

*Note that the coefficients for the relative and average house price covariates are not included in the Tables 2-4 because the data are proprietary (CoreLogic). Also note that because the CoreLogic data are not included in these analyses, the results will not exactly match the published tables.*

* **Table 1 and Figures 1 and 2**
  + table1\_figs1and2.do – uses table1.dta to produce the results in Table 1, Figure 1 (g1.gph), and Figure 2 (g2.gph).
* **Table 2**
  + table2.do – uses table2.dta to produce the results in Table 2.
* **Table 3 & Table A1**
  + table3.do – uses table3.dta to produce the results in Table 3 and Appendix Table A1.
* **Table 4 and Table A2**
  + table4.do – uses table4.dta to produce t4results.dta and the results in Table 4 and Appendix Table A2.
* **Figure 3**
  + fig3.do – uses fig3.dta to produce Figure 3 (g3.gph).
* **Table 5**
  + lp\_lt35\_2.do– uses marchcps\_mig.dta to the produce the “18-35” results in Table 5.
  + lp\_gt35\_2.do – uses marchcps\_mig.dta to the produce the “36-65” results in Table 5.
* **Table 6**
  + interactions\_lt35\_2.do – uses marchcps\_mig.dta to produce the “18-35” results in Table 6.
  + interactions\_gt35\_2.do – uses marchcps\_mig.dta to produce the “36-65” results in Table 6.
* **Table A3**
  + firstage2.do – uses marchcps\_mig.dta to produce the first stage estimates in Appendix Table A3. Note that the titles "Heads 18 to 35" and "Heads 36 to 65” should be swapped.

**Appendix**

*Steps to replicate rows (1) and (2) of Table 1 from Saks and Wozniak (Journal of Labor Economics 2011).*

1. To create migration data:
   1. Data for 1948-1976 are were collected from tables based on the Current Population Survey (CPS) published by the Bureau of Labor Statistics. Tables by year are posted here: <https://www.census.gov/topics/population/migration/data/tables.html>. Table numbers vary by year. Tables report the number of non-migrants, within-county migrants, across-county migrants, and across-state migrants.
   2. Data for 1980-2009 were calculated from CPS microdata published by IPUMS (<https://cps.ipums.org/cps/>).
   3. The annual time series for the two sample periods are combined. We focus on the migration rate defined as the number of inter-county migrants age 14+ divided by population age 14+. The population used in the denominator is defined as the sum of within-county migrants, across county migrants, and non-migrants. People who moved into the US or emigrated from the US are excluded from the population used in the denominator. We estimate a trend in this series using a Hodrick-Prescott filter using a smoothing parameter of 100 (standard for annual data). Then we calculate a detrended series as the difference between the migration rate and its trend.
2. To create the two independent variables (national employment gap and standard deviation of state unemployment gaps)
   1. We take national and state-level employment data from the Bureau of Labor Statistics Current Employment Statistics program (<https://www.bls.gov/ces/>).
   2. The state-level data are not seasonally adjusted, so first we seasonally adjust using the X11 procedure, then we calculate an HP filter on the seasonally-adjusted levels using the smoothing parameter 14400 (standard for monthly data).
   3. Next we calculate the employment gap as the log of the ratio of the seasonally-adjusted series to its trend. To convert these monthly series to annual series, we take the April values of each year.
   4. We calculate the standard deviation of these gaps across states in each year.
   5. Calculation of the national employment gap is similar, except that the BLS publishes seasonally-adjusted national data, so we do not have to do the seasonal adjustment ourselves.
   6. Finally, we standardize the national employment gap and standard deviation of state employment gaps to have a mean equal to zero and standard deviation equal to 1.
3. We estimate the coefficients reported in Table 1 using an OLS regression:
   1. Column (1) regresses the detrended inter-county migration rate on the national employment gap. Sample period is 1948-2009.
   2. Column (2) regresses the detrended inter-country migration rate on the national employment gap and the standard deviation of the state-level employment gaps.
   3. Columns (3) and (4) are similar regressions, except the sample period is limited to 1948-2005. For this analysis, the underlying data are different in that the detrended migration rate and employment gaps are estimated using HP filters that only use data through 2005 (using the same smoothing parameters as above).

Programs and files:

* cps4880.do is a Stata do-file that reads in the CPS tables from 1948 to 1980 and calculates annual time series of migrants and non-migrants. It reads annual .csv files derived from the PDF files. It saves the data in cps4880.dta.
* cpsmicro.do is a Stata do-file that reads in CPS microdata downloaded from IPUMS (cps\_00007.dat) and creates annual time series of migrants and non-migrants. Data are saved in cpsmigann\_jole.dta.
* The Fame input file statedata2.inp reads in the employment data from an FRB database, does seasonal adjustment for the state-level data, and calculates an HP filter for the state and national data. It saves the data in statedata2.xls, which was then converted to data2.csv.
* The Stata do-file stateemp.do reads in the data from data2.csv, calculates the employment gaps, calculates the standard deviation of the state-level gaps, and saves the results in stateemp2.dta.
* cpsgraphs\_final.do is a Stata do-file that combines the annual series from the cps4880.dta and cpsmigann\_jole.dta. It also merges in the national employment gap and standard deviation of state-level employment gaps from stateemp2.dta. It calculates the migration rate and calls the RATS program cpsdata.prg to run the HP filter. The RATS program creates output named cpsdata.out, which is then read into Stata by this same program. The RATS program cpsdata05 runs the HP filter through 2005 for the detrended migration rate used in columns (3) and (4) of Table 1. The program then calculates the detrended migration series and saves the detrended migration data and employment gap data in the Stata dataset table1.dta.
* The Stata-do file table1\_figs1and2.do runs the regressions reported in Table 1.