README

Replication instructions for "Measuring Employer-to-Employer Transitions"
American Economic Journal: Macroeconomics
by Shigeru Fujita, Giuseppe Moscarini, and Fabien Postel-Vinay

1   Overview

This replication package consists of this readme file and of the following five folders:

1. `CPSData`: This folder includes the zipped file (`CPSMonthly-Raw.zip`) of all monthly CPS micro data used in our analysis and the zipped final micro data `ee_micro_final.dta.zip` which one obtains when the imputation process is completed. In addition, it also contains three dta files that includes seasonally adjusted time series data as explained below.

2. `Stata`: This folder has two subfolders: `Imputation` and `AdditionalAnalysis`. The former includes Stata codes to generate our final EE series, as described in Section 5.1 below. The latter includes Stata codes for additional analysis (Figures and Tables) presented throughout the paper.

3. `Matlab`: This folder contains Matlab codes that are written mostly for data visualization purposes.

4. `ExcelFiles`: This folder contains Excel datasets and other auxiliary files that are either directly downloaded from the original data sources or produced during the imputation process.

5. `SIPP`: This folder contains codes to extract raw SIPP data files for the 1996, 2000, 2004, and 2008 panels (1996-2013 data), which are not included in this package, merge them with data dictionaries, save them in Stata format, and construct the aggregate time series of the EE transition rate, plotted on Figure 16. The folder also contains two data zipped folders. First, `SIPP_1996-2013_data.zip`. Expanding this file will create the folder that includes said SIPP data files and an additional auxiliary, but essential, data file FFseries.dta that are used by the Stata code that produces the final EE time series data. Second, `SIPP_2014_data.zip` contains raw data from the 2014 SIPP panel, already available from the Census Bureau website in Stata format, with Stata code that derives the inline numbers reported on page 39 of the manuscript.

   The first and main step to replicate the results of the paper is to run `0_master_imputation.do`, which executes nine Stata programs, as described below. Once the folder structure is adjusted within the master code, the remaining programs should run without any changes. During this process, various auxiliary data files and results files are generated and saved.

These files are used later for further analysis or visualized later in Matlab. The Stata programs for further analysis are included in the subfolder `AdditionalAnalysis` noted above. The production of the SIPP EE series is a separate process.

2   Data Availability and Provenance Statements

2.1   Statement about Rights

We certify that the authors of the manuscript have legitimate access to and permission to use the data used in this manuscript.

2.2   Summary of Availability

All data are publicly available.

2.3   Details on each Data Source

The data used in our paper are obtained from the following sources. All the data used in the paper are public use data.

- CPS basic monthly micro data files: Census Bureau (2022a). All files are downloaded from:

  `https://www.census.gov/data/datasets/time-series/demo/cps/cps-basic.html`.

  On this page, each Basic Monthly CPS file is posted separately for each year. One can go back to 1994 and download all monthly files since 1994 on this page. For each month, download the file posted under "DOS/Windows." The name of the downloaded file identifies month and year and the file has an extension `dat`. The file name needs to be manually changed to the following format `cpsbyyyymm.raw`, where `yyyy` and `mm` represent year and month. For the user's convenience, we have included all monthly files, already renamed, in a zipped file `CPSMonthly-Raw.zip`, as mentioned above.

- Longitudinal Employer-Household Dynamics Job-to-Job Flows: Census Bureau (2022b). The data are downloaded from:

  `https://lehd.ces.census.gov/data/`.

  On this page, there are drop-down menus for Version, Region, and Format. Select United States, j2jr, XLSX for Region, Type, and Format, respectively, and download `j2jr_us_all_aggregations_s.xlsx` for the aggregate level analysis (as in Figure 17). For the group level analysis, select United States, j2j, and csv, download `j2j_us_all.csv.gz`, and unzip it. Note that for the series plotted in Figure 17, the version R2022Q4 of the data is used, while for the group-level analysis associated with Figures 19, OA7, OA8, the version 2021Q2 is used. They are the latest versions available at the time of the respective analysis.

- Current Population Survey (Employment Situation Summary): Bureau of Labor Statistics (2022a). We use the number of unemployed (seasonally adjusted). It is readily available at:

  `https://data.bls.gov/cgi-bin/surveymost?bls`

  On this page, download Civilian unemployment (LNS13000000). The original monthly series is converted into a quarterly average series and then copied into `tightness.xlsx` available in the `ExcelFiles` folder of this package.

- Job Openings and Labor Turnover Survey (JOLTS): Bureau of Labor Statistics (2022b). We use the quit rate and the number of job openings (both seasonally adjusted) for the total nonfarm sector. It is readily available at:

  `https://data.bls.gov/cgi-bin/surveymost?jt`.

  On this page, download Job openings level, Total nonfarm (JTS000000000000000JOL) and Quits rate, Total nonfarm (JTS000000000000000QUR). Both series are downloaded as monthly series but are converted into quarterly average series and then copied into `tightness.xlsx` and `joltsQ.xlsx`, respectively, available in the `ExcelFiles` folder of this package.

- US Business Cycle Expansions and Contractions: NBER (2023). We use the business cycle dates to draw the recession bars in some figures, as explained below (see Table 1 below). The business cycle dates are obtained at:

  `https://www.nber.org/research/data/us-business-cycle-expansions-and-contractions`.

- Survey of Income and Program Participation (SIPP): Moscarini and Postel-Vinay (2016) and Census Bureau (2014). We plot the time series data based on the SIPP data in Figure 16, which uses four SIPP panels: 1996, 2001, 2004, and 2008. The series is originally produced for Moscarini and Postel-Vinay (2016). All datasets are downloaded from the following page:

  https://www.nber.org/research/data/survey-income-and-program-participation-sipp

  On this page, one can download Core files for all waves for the above mentioned four panels. Those files are posted under under `.zip`. One also needs associated dictionary files posted under `.dct`. Our package does not include these raw files, but rather includes the extraction code `extract_NBER.do` and the already-extracted datasets in Stata format (`SIPP_1996-2013_data.zip`). Expanding this zipped folder will create a folder that includes all datasets.

  We also use the SIPP 2014 panel for the calculation discussed in Section V.A. These datasets are downloaded, already in Stata, from the following location.

  `https://www.census.gov/programs-surveys/sipp/data/datasets.2014.html#list-tab-1064665487`.

  On this page, clicking on the link to each wave takes one to the page where the dataset for that wave is available in Stata format. Download the dataset for all four waves for the SIPP 2014 panel.

3 Dataset List

Table 1 lists all the datasets that are included in this package. During the imputation and the additional analysis, various Stata dta files are generated but they are not included in this package. One exception is the final micro data `ee_micro_final.dta` that is generated when the imputation process is completed. The other exceptions are the following three dta files, `EE_grp_SA.dta`, `EEr_M_SA.dta`, and `ENQ_Q_SA.dta`. These are files that are generated via Stata do files for seasonally adjust some time series data by calling the Census X13 program. We include these files, so that the replicators can use these datasets without running these programs, because the execution of these programs requires additional steps as discussed in Section 5.3.3.

The final results are saved in Excel format and those files are included in the package. See Table 1 for details.

4 Computational Requirements

4.1 Environment

All codes are run on a MacBook with M1 Pro processor, with 32GB of memory and 1TB of a hard drive storage capacity. An exception to this is when we seasonally adjust the series using X-13ARIMA-SEATS Seasonal Adjustment Program (Lunix/Unix version). The seasonal adjustment procedures were run on a MacBook Pro with Intel Core i7 Processor and MacOS Mojave. This Seasonal Adjustment Program does not work on a newer version of Mac computers. One can run the codes on a windows machine by changing the path format to the windows style format.

4.2 Software Requirements

- Stata (code was last run with version 17)

- Matlab (code was last run with Matlab Release R2022_a)

- X-13ARIMA-SEATS Seasonal Adjustment Program

  - Available at: `https://www.census.gov/data/software/x13as.html`. Within this page, click on the tab "X-13ARIMA-SEATS." On that page, click on "Download" and then download either the Windows version or the Linux/Unix version. The latter version runs on Intel Macs but does not run on Apple silicon Macs. In order to run the seasonal adjustment programs, one needs to accommodate this restriction. More details can be found in Section 5.3.

Table 1: Datasets Included in the Package

	File location in the package	Source	Purpose/Notes
(1)	CPSData/CPSMonthly-Raw.zip	Census Bureau (2022a)	Unzipping it creates a folder for all monthly files
(2)	CPSData/ee_micro_final.dta	Output from Stata	Final micro data after imputation is completed
(3)	ExcelFiles/AgeGrpEEm.xlsx	Output from Stata	Fig. OA.6
(4)	ExcelFiles/BTwoDummiesJJSS1CycTr.xlsx	Output from Stata	Fig. A.3
(5)	ExcelFiles/BTwoDummiesJJSS1CycTrCont.xlsx	Output from Stata	Fig. OA.4
(6)	ExcelFiles/Composition.xlsx	Output from Stata	Fig. 11
(7)	ExcelFiles/DIRetired.xlsx	Output from Stata	Fig. OA.2, OA.3
(8)	ExcelFiles/EEm1.xlsx	Output from Stata	Fig. 2, 3, 5, 6, 7, 8, 9
(9)	ExcelFiles/EEmrCohort.xlsx	Output from Stata	Fig. 10
(10)	ExcelFiles/EEr_M_SA.xlsx	Output from Stata	Fig. 18
(11)	ExcelFiles/EEr_QfromM_SA.xlsx	Output from Stata	Fig. 1(b), 15(a)
(12)	ExcelFiles/ENQ_Q_SA.xlsx	Output from Stata	Fig. OA.11
(13)	ExcelFiles/Imp_TwoDummiesJJSS1CycTr.xlsx	Output from Stata	Fig. 1(a), 13, 14, 15(b), A.2, OA.9
(14)	ExcelFiles/Imp_TwoDummiesJJSS1CycTrCont.xlsx	Output from Stata	Fig. OA.4
(15)	ExcelFiles/Imp_ObsOnlyJJSS1CycTr.xlsx	Output from Stata	Fig. A.4
(16)	ExcelFiles/JJ_FF_MAR_Compare.xlsx	Manually edited	Fig.4. Created from (17) and (18)
(17)	ExcelFiles/JJ_FF_MAR_Compare_Temp.xlsx	Output from Stata	Copied into (16)
(18)	ExcelFiles/feds200434.xls	Fallick and Fleischman (2019)	Copied into (16)
(19)	ExcelFiles/matching_rate_edited.xlsx	Manually edited	Fig. A.1. Copied from files (18) and (19)
(20)	ExcelFiles/matching_rate_IDplus.xlsx	Output from Stata	Copied into matching_rate_edited.xlsx (17)
(21)	ExcelFiles/matching_rate.xlsx	Output from Stata	Copied into matching_rate_edited.xlsx (17)
(22)	ExcelFiles/OccMob.xlsx	Output from Stata	Fig. OA.1
(23)	ExcelFiles/RIP_Treatment.xlsx	Output from Stata	Fig. 12
(24)	ExcelFiles/SIPP_CPS_Comp.xlsx	Manually edited	Fig. 16. Copied from files (13) and (31)
(25)	ExcelFiles/tightness.xlsx	Bureau of Labor Statistics (2022a,b); Manually edited	Fig. OA.10. Averaged into quarterly series
(26)	ExcelFiles/j2jr-us.all.aggregations_s.xlsx	Census Bureau (2022b)	Copied into lehd_j2j.xlsx (27)
(27)	ExcelFiles/j2j_us.all_R2021Q2.csv	Census Bureau (2022b)	Input for EEbyGrp_LEHD.do. Fig. 19, OA.7, OA.8
(28)	ExcelFiles/joltsQ.xlsx	Bureau of Labor Statistics (2022a,b)	Fig. 17. Quarterly averages of monthly series
(29)	ExcelFiles/lehd_j2j.xlsx	Census Bureau (2022b); Manually edited	Fig. 17. Copied from (24)
(30)	ExcelFiles/NBER_recession_monthly.csv	NBER (2023)	Fig. 1(a), 13, 14, 15 (b), 16, 18, OA.9
(31)	SIPP/SIPP_1996-2013_data.zip	Census Bureau (2014)	Underlying micro data to produce (31)
(32)	SIPP/EE_SIPP.dta	Output from Stata	Copied into SIPP_CPS_Comp.xlsx (24)
(33)	SIPP/SIPP_2014_data.zip	Census Bureau (2014)	Underlying micro data to produce in-line numbers on p. 39
(34)	CPSData/DTA/EEr-grp_SA.dta	Output from Stata	Used to produce Figures 19, OA.7, OA.8
(35)	CPSData/EEr_M_SA.dta	Output from Stata	Stata dta version of (10)
(36)	CPSData/ENQ_Q_SA.dta	Output from Stata	Stata dta version of (12)

4.3 Memory and Runtime Requirements

- The computation of our EE series requires to run `0_master_imputation.do`, which runs 9 Stata programs. This whole process takes about an hour on a standard desktop computer available in 2023.

- The master code for the additional analysis `MasterAdditionalAnalysis.do` takes about an hour to go through all the Stata do files listed within the master code.

- The code for visualizing the results all take a negligible amount of time.

- The code `SIPP_1996-2013_aggregate_EE.do` takes about 3 hours to run, using the extracted Stata data files that we provide. The extraction itself to generate the same Stata files using `extract_NBER.do` would take additional time. To produce the in-line numbers on page 39, which uses the SIPP 2014 panel, takes negligible amount of time.

- The programs were last run on a MacBook Pro with Apple M1 Pro chip, 32 GB of RAM, and 1 TB of local storage.

    - As noted above, the X13 seasonal adjustment procedures require a different environment. They were last run on a MacBook Pro with MacOS Mojave, Intel Core i7 Processor, 16 GB of RAM, and 500GB of local storage.

5 Description of Programs/Code

We first describe a set of Stata codes that produce our final EE rate series in Section 5.1. The programs for additional analysis are described in Section 5.3.2 as part of the production of figures and tables.

5.1 Imputation Procedure

The imputation procedure is executed in steps. The master code `0_master_imputation.do` executes all the steps as described below. It is necessary to edit the folder structure within the master code. No other changes are necessary. Note that during this process, a few datasets are generated and those derived datasets are used for the subsequent analysis.

- `0_master_imputation.do`: This is a master code that executes all codes below.

- `1_extract.do`: This code extracts variables of interests from raw CPS monthly files. It produces Stata data files for all months, starting in January 1994. Each monthly file is named `cpsmyyyymm.raw`, where `yyyy` refers to year and `mm` to month.

- `2_1_adjust_ind.do`: This code creates industry categories that are consistent over time, and overwrites monthly CPS Stata files.

- `2_2_adjust_occ.do`: This code creates occupation categories that are consistent over time, and overwrites monthly CPS Stata files.

- `3_match.do`: This code matches individuals across two adjacent months using the existing files and saves each matched file as `matchedyyyymm.raw`.

- `4_genvars_Probit.do`: This code generates additional variables, taking each matched dataset. In particular, it runs Probit regressions, computes predicted probabilities of having an missing answer to the EMPSAME question, and overwrites the data.

- `5_append_matched.do`: This code appends all monthly matched files into one dataset and saves it as `cps_all_matched.dta`.

- `6_genvars_matched.do`: This code takes `cps_all_matched.dta`, generates additional variables, and saves it as `cps_matched_all_additional_vars.dta`.

- `7_UEr_JJrSS_append.do`: This code uses `cps_matched_all_additional_vars.dta`, generates aggregate time series of the UE transition rate and EE transition rates for selected groups (the overall SS group and the SS group between the first and second interviews), merges these aggregate series back to the original dataset, and overwrites it. The EE rate for the SS group between the first and second interviews is used in the imputation as the trend and cyclical variables, as described in the paper.

- `8_1_Imputation_base.do`: This code implements the imputation of the missing observations and generates the aggregate time series including the final EE rate for the U.S. economy. The final results are stored in `BTwoDummiesJJSS1CycTr.xlsx` and `Imp_2DummiesJJSS1CycTr.xlsx`. The first one includes average bias series for the five respondent groups and the later one includes EE rates for various groups. As explained below, Matlab codes take these series and draw graphs in the paper. This code also saves the final micro data file as `ee_micro_final.dta`.

5.2 Additional analysis and Data Visualization

This section describes the codes for the data visualization and additional analysis we conduct throughout the paper on top of the production of our final EE rate series. The description is organized by figures and tables. Notes for the replicators:

- All figures except for Figures 19, OA5, OA7, and OA8 are produced in Matlab by simply running `MasterMatlabFigures.m`. This master program lists all Matlab programs, which are explained below one by one. To run individual programs, one can comment out unnecessary programs on the master program. Figures 19, OA5, OA7, and OA8 are produced within Stata, as described below.

- The Stata do files that are used for the additional analysis can be executed through the master code `MasterAdditionalAnalysis.do`. Modify the folder structure at the beginning of the code.

Note that some Stata programs are written to seasonally-adjust time-series data. Executing them requires additional steps as explained below. Note also that one of the programs (`EEbyGrp_merge.do`) executed through `MasterAdditionalAnalysis.do` uses a Stata dataset `EEr_grp_SA.dta` that stores seasonally-adjusted data. This dta file is included in this package and should be placed in the folder specified by a global macro "dta" in the master code. In addition, one also needs to place `j2j_us_all_R2021Q2.csv`, which can be found in `ExcelFiles`, in the same folder ("dta").

The instructions below describe the production process of figures and tables in detail. See also the NOTES section in the master code.

Figures 1(a), 13, 14, 15(b), A.2, OA.9: Run `EE_Figures.m` in Matlab. The code takes the final Excel file (`Imp_2DummiesJJSS1CycTr.xlsx`) produced in the last stage of the imputation as described above. It also uses `NBER_recession_monthly.csv`, which includes the NBER's recession dates.

Figures 1(b), 15(a) : Run `E2E_Save.do` in Stata. The code takes `ee_micro_final.dta` and collapses the data to produce the time series of EE and UE rates (saved in `EE_M.dta`). Next run `E2E_MSA.do`, which takes `EE_M.dta`, seasonally adjusts the series, produces quarterly averages of monthly SA data, and outputs the Excel file `EEr_QfromM_SA.xlsx`. Finally, run `EE_Quarterly.m` in Matlab, which produces the graphs.

Figures 2, 3, 5, 6, 7, 8, 9: Run `EEm_by_date.do`. The code computes various shares of missing observations (using `cps_matched_all_additional_vars.dta`) and outputs an Excel file `EEm1.xlsx`. Using this Excel file, `EEm_Figures.m` produces these figures in Matlab.

Figure 4: Run `match_FF.do` and then `append_matched_FF.do`, which produces `cps_all_matched_FF.dta`. `EE_FF_replication.do` computes EE rates based on the methodology of Fallick and Fleischman (2004) and under the assumption of missing at random. The results are saved in `JJ_FF_MAR_Compare_Temp.xlsx`. For the original Fallick and Fleischman (2004) series, download the Excel file `feds200434.xls` from the website Fallick and Fleischman (2019), copy the series `EEhaz` into the previous Excel file, and save it as `JJ_FF_MAR_Compare.xlsx`. `FF_MAR_compare.m` produces the two graphs using this Excel file.

Figure 10: Run `EEm_by_cohort.do` in Stata. The code computes the share of missing observations by cohort (using `cps_matched_all_additional_vars.dta`) and outputs an Excel file `EEmrCohort.xlsx`. Using this Excel file, `EEm_Cohort.m` produces the figure.

Figure 11: Run `Composition_1.do`, which estimates Probit regressions for the SS group on each monthly dataset and computes predicted probabilities for the remaining four groups. `Composition_2.do` appends all monthly datasets and `Composition_3.do` collapses the data and saves the actual and predicted series in `Composition.xlsx`. Using this Excel file, `Composition.m` produces the graphs.

Figure 12: Run `Control_Treatment.do` in Stata. The code runs the regressions using the data in `cps_matched_all_additional_vars.dta`. The output is `RIP_Treatment.xlsx`. Using this Excel file, `Treatment.m` produces the graphs.

Figure 16: This figure uses the SIPP EE rate series following the methodology in Moscarini and Postel-Vinay (2016). The Stata code `SIPP_1996-2013_aggregate_EE.do` produces the final SIPP EE series (`EE_SIPP.dta`). The replication of the other two series in the figure are already discussed above. The Excel file `SIPP_CPS_Comp.xlsx` includes the three series (assembled manually) plotted in this figure. `SIPP_CPS_Compare.m` reads the Excel file together with `NBER_recession_monthly.csv` and draws the graph.

Figure 17: Run `lehd_compare.m` in Matlab. The code reads three Excel files that includes LEHD J2J separation rate (`lehd_j2j.xlsx`), JOLTS quit rate `joltsQ.xlsx`, and CPS EE rates (`EEr_QfromM_SA.xlsx`), and plots them together. Note that the JOLTS quit rate is time aggregated into quarterly averages.

Figure 18: Run `EE_UE_Recent.m` in Matlab. The code reads `EEr_M_SA.xlsx`, which is generated within `E2E_MSA.do` discussed above, together with `NBER_recession_monthly.csv`, and draws the graph.

Figures 19, OA.7, OA.8: Run `EEbyGrp.do`, which reads `EE_micro_final.dta` and generates `EE_grp.dta`. `EEGrpSA.do` reads this data and seasonally adjusts and saves it as `EEr_grp_SA.dta`, which is included in the package for convenience, as noted at the beginning of this section. For LEHD data, `EEbyGrp_LEHD.do` puts together seasonally-adjusted LEHD-based EE rate series by demographic groups by reading the original data `j2j_us_all_R2021Q2.csv`, and saves it as `EEr_LEHD.dta`. `EEbyGrp_merge.do` merges the CPS-based and LEHD-based EE rate series. The merged dataset is saved as `CPS_LEHD_EErates_Q_10Groups.dta`. `BlaiseRIPPropGr.do` reads this dta file and `RIPPropensity.dta`, and produces the graphs. The production of `RIPPropensity.dta` is discussed below for Figure OA.6.

Figure A.1: Run `match_FF.do` and `append_matched_FF.do` that produces `cps_all_matched_FF.dta`. Run `match_IDs.do` and `append_matched_IDs` that produces `cps_all_matched_IDs.dta`. Then run `match_rates_observables.do` and `match_rates_IDs.do` that output the two Excel files `matching_rate_IDplus.xlsx` and `matching_rate.xlsx`, respectively. `matching_rate_edited.xlsx` puts together the first series from the former and all four series from the latter (together with the dates). This last file is included in this package. `matching_rates_gra.m` draws the two graphs in Matlab using this file.

Figure A.3: Run `BhatGraphs.m` in Matlab. The code produces the graphs using the Excel file `BTwoDummiesJJSS1CycTr.xlsx`.

Figure A.4: Run `Imputation_obs_only.do` in Stata. The code conducts the imputation using the observables only and saves the results in `Imp_ObsOnlyJJSS1CycTr.xlsx`. `EErObsOnly.m` produces the two graphs using the Excel file.

Figure OA.1: Run `DI_Analysis_Occ.do`. The code produces the series plotted in the figure using `EE_micro_final.dta` and saves the results in `OccMob.xlsx`. `OccMobility.m` draws the graph using the Excel file in Matlab.

Figures OA.2, OA.3: Run `DI_Analysis_Retired.do` in Stata. The code produces the series in the figure using `EE_micro_final.dta` and saves the results in `DIRetired.xlsx`. Using the Excel file, `DI_Retired.m` draws the graph.

Figure OA.4: Run `Contributions.m`. The code produces the graphs by using the two Excel files `Imp_2DummiesJJSS1CycTrCont.xlsx` and `BTwoDummiesJJSS1CycTrCont.xlsx`.

Figure OA.5: Run `EECI.do`. The code calculates the standard errors, using variance-covariance matrices and Jacobian, which are previously calculated and saved as the data files (in `4_genvars_Probit.do` and `8_Imputation_base.do`). The results are saved in `SE_EEm.dta`. `EECI_Graphs.do` merges this file into `E2ETemp.dta`, which is created in `8_Imputation_base.do`, and then draws the graphs in Stata.

Figure OA.6: Run `BlaiseRIPPropensity.do` in Stata. The code saves the results in `AgeGroupEEm.xlsx`. Using this Excel file, `AgeGrpEEmP.m` produces the graphs in Matlab. The code also saves `RIPPropensity.dta` (which is used for Figure OA.7 and Figure 19).

Figure OA.10: Run `timing_around_GR.m`. The code reads the Excel files `tightness.xlsx` and `EEr_QfromM_SA.xlsx`, and draws the graph. The first file includes quarterly averages of the number of job openings (from the JOLTS) and unemployed (from the CPS) that are used to compute the tightness measure. It also includes the LEHD J2J rate, which is an identical series as the one used for Figure 17. The original data on the number of job openings and the number of unemployed are both downloaded as monthly series, as noted in Section , but converted into quarterly series before they were saved in the Excel file.

Figure OA.11: Run `EN_EE.do` in Stata. The code generates the monthly CPS-based transition rates, using `ee_micro_final.dta`. `EN_EE_SA.do` seasonally adjusts the monthly series, computes quarterly averages, and saves the results in `ENQ_Q_SA.xlsx`. The JOLTS quit rate (quarterly averages of monthly series) is manually added to this Excel file. `EE_EN.m` draws the graphs using the Excel file.

Tables OA.1-OA.3: In the final step of the imputation, a log file (`ImputationResults.smcl`) is generated and placed into the folder as the data folder. It is included in the folder `ExcelFile` as part of this package. The tables are manually produced from this log file. Note that RIPFLAG = 1 in the Stata output indicates BLAISE dummy = 1 and RIPFLAG = 2 indicates RIP dummy = 1. The top portion of the log file corresponds to OA.1 and all variables are listed in the same order. The results for the two interaction terms are stacked in the log file, while they are split into two tables, AO.2 and AO.3, respectively.

Section II.C. CPS In-line number: On page 16, we mention "over 70%." One can reproduce this figure by running `PSSS_EEm.do`. The result (71.3%) is shown on the screen at the end.

Section V.A. IPUMS-ABACUS in-line numbers: On page 39, we refer to the calculations based on IPUMS (n.d.). To obtain these results, go to

`https://usa.ipums.org/usa/abacus.shtml`

and launch Abacus. Add MIGRATE1 available under Migration. Restrict the sample to those employed by selecting the subcategories, Work, EMPSTAT, and Employed. Select time 2014 ACS through 2018 ACS.

Section V.A. SIPP 2014 in-line numbers: On page 39, we refer to the calculations based on Census Bureau (2014). Run `SIPP_2014_panel_movers.do` after changing the root folder if necessary.

5.3 Instructions to Replicators

5.3.1 Imputation

To run `0_master_imputation.do`, follow the instructions below. All the programs for the imputation can be found in `Stata/Imputation` in this package. It requires the replicator to create new folders and modify the paths specified in the master file as follows.

1. Create a folder to store all the results. In the master file, this folder is specified as:

   ```
   global dta = "$bases/CPSData/DTA".
   ```

2. Create a folder to store the Stata do files. Move all do files from `Stata/Imputation` in the data repository to this newly created folder, so that the master file can refer to these do files. In the master file, this folder is specified as:

   ```
   global dos= "$bases/Stata/Imputation".
   ```

3. Unzip the file `CPSMonthly-Raw.zip` to create a folder of the same name that includes all monthly CPS files. The master file refers to this folder.

4. Run `0_master_imputation.do`, which will go through all the steps laid out in Subsection 5.1.

5.3.2 Additional Analysis and Data Visualization

After running the imputation process, which creates output files in the folder `$bases/CPSData/DTA` mentioned in subsection 5.3.1, you can proceed to additional analysis, using Stata programs in the folder `Stata/AdditionalAnalysis` in this package, that use those imputation output files. This part can be executed by running `MasterAdditionalAnalysis.do` in Stata, after modifying the folder structure at the beginning of the code, as was also discussed above for the imputation. Note that one also needs to place the two data files, `EEr_grp_SA.dta` and `j2j_us_all_R2021Q2.csv`, in the folder specified by the global macro "`dta`" in the master code. These two data files are, respectively, available in `CPSData` and `ExcelFiles` within this package.

Data visualization is performed mostly in Matlab and can be executed by running `MasterMatlabFigures.m`, which lists all Matlab programs. This will create and save figures in eps format, whose name correspond exactly to figure numbers used in the manuscript. These eps files are saved in the same folder as the Matlab programs. The data files used for the visualization are included in the folder `ExcelFiles` in the package and should be placed in the folder that is specified at the beginning of this master program. But note that most of these data files are generated as part of the analysis from Stata and thus saved in the `$bases/CPSData/DTA` folder (noted above), when one executes the Stata programs.

5.3.3 X13 Seasonal Adjustment

There are three Stata codes within `AdditionalAnalysis` that are written for the purpose of seasonally adjusting time series data. Those three do files are:

- `EN_EE_SA.do`

- `E2E_MSA.do`

- `EEGrpSA.do`

Running these files requires one to obtain X13 program, `x13as`, as documented above. Each Stata code calls this external program that is placed in a location, specified at the beginning of the code as:

- `global path_SA ...`

One also needs to place a "spec" file placed in the same folder. The first two do files above use the spec file `SA.spc` and the last one uses the spec file `SAq.spc`. The above three do files and the two spec files are included `Stata/AdditionalAnalysis` in this package. To run the do files, one needs to place `x13as` and the spc files in the folder specified as a global macro "`path_SA`" at the beginning of the three do files. Note also that `x13as` does not run on Apple silicon Macs, as discussed in Section 4.2. This is the reason why the seasonal adjustments cannot be integrated into `MasterAdditionalAnalysis.do`. We included in this package the three dta files, `EEr_grp_SA.dta`, `EEr_M_SA.dta`, and `ENQ_Q_SA.dta` that are output from the above three do files. They are available in the `CPSData` folder.

5.3.4 SIPP

EE series. From the weblink listed, and following instructions, in Section 2.3, download the Core SIPP datasets (.zip) for years 1996 through 2008 (all waves), the associated extraction programs (.do), and the dictionary files (.dct) into folders called `sippYY` where `YY` is 96, 01, 04, and 08. Run the provided program `extract_NBER.do`. This will extract the raw data and merge the dictionary files to create Stata files, called `sippYYwi.dta` where `i` is the wave. These Stata files are already provided for convenience in the zipped folder, `SIPP SIPP_1996-2013_data.zip`, along with `FFseries.dta` from Fallick and Fleischman (2019). The code `SIPP/SIPP_1996-2013_aggregate_EE.do` reads these Stata files and generates the final aggregate EE time series, saved in `EE_SIPP.dta` and shown on Figure 16. See also the instruction for Figure 16 above.

In-line numbers on page 39. The `SIPP` folder contains a zipped folder `SIPP_2014_data.zip`, where we placed the raw data for the SIPP 2014 panel in Stata format to replicate in-line numbers on geographical mobility on p.39 of the manuscript. The replicator can unzip `SIPP_2014_data.zip`, open the Stata code `SIPP_2014_panel_movers.do`, change the root directory if necessary, and run it. The code will produce tabulations from which one can read, among others, the numbers that we quote on page 39.

6 List of Figures and Tables, and Programs

The programs and data provided in this package reproduce:

- all numbers provided in text in the paper; and

- all tables and figures in the paper.

Production of Figures and Tables is described above, but the following table provides the summary.

Table 2: Figures and Tables List

Figure/Table #	Program	Output file	Note
Figure 1 (a)	Matlab/EE_Figures.m	Fig1a.eps	
Figure 1 (b)	Matlab/EE_Quarterly.m	Fig1b.eps	
Figure 2	Matlab/EEm_Figures.m	Fig2.eps	
Figure 3	Matlab/EEm_Figures.m	Fig2.eps	
Figure 4 (a), (b)	Matlab/FF_MAR_compare.m	Fig4a.eps, Fig4b.eps	
Figure 5	Matlab/EEm_Figures.m	Fig5.eps	
Figure 6	Matlab/EEm_Figures.m	Fig6.eps	
Figure 7	Matlab/EEm_Figures.m	Fig7.eps	
Figure 8	Matlab/EEm_Figures.m	Fig8.eps	
Figure 9 (a)-(d)	Matlab/EEm_Figures.m	Fig9a.eps,···, Fig9d.eps	
Figure 10	Matlab/EEm_Cohort.m	Fig10.eps	
Figure 11 (a)-(e)	Matlab/Composition.m	Fig11a.eps,···, Fig11e.eps	
Figure 12 (a)-(d)	Matlab/Treatment.m	Fig12a.eps,···, Fig12d.eps	
Figure 13 (a), (b)	Matlab/EE_Figures.m	Fig13a.eps, Fig13b.eps	
Figure 14 (a),(b)	Matlab/EE_Figures.m	Fig14a.eps, Fig14b.eps	
Figure 15 (a)	Matlab/EE_Quarterly.m	Fig15a.eps	
Figure 15 (b)	Matlab/EE_Figures.m	Fig15b.eps	
Figure 16	Matlab/SIPP_CPS_Compare.m	Fig16.eps	
Figure 17	Matlab/lehd_compare.m	Fig17.eps	
Figure 18	Matlab/EE_UE_Recent.m	Fig17.eps	
Figure 19	Stata/AdditionalAnalysis/BraiseRIPPropGr.do	Fig18.eps	See Section 5.2 for more details
Figure A1 (a), (b)	Matlab/matching_rates_gra.m	FigA1a.eps, FigA1b.eps	
Figure A2	Matlab/EE_Figures.m	FigA2.eps	
Figure A3 (a)-(e)	Matlab/BhatGraphs.m	FigA3a.eps,···, FigA3e.eps	
Figure A4 (a), (b)	Matlab/EErObsOnly.m	FigA4a.eps, FigA4b.eps	
Figure OA1 (a), (b)	Matlab/OccMobility.m	FigOA1a.eps, FigOA1b.eps	
Figure OA2 (a)-(e)	Matlab/DI_Retired.m	FigOA2a.eps, ···, FigOA2e.eps	
Figure OA3 (a)-(e)	Matlab/DI_Retired.m	FigOA3a.eps, ···, FigOA3e.eps	
Figure OA4 (a)-(e)	Matlab/Contributions.m	FigOA4a.eps, ···, FigOA4e.eps	
Figure OA5 (a)-(e)	Stata/AdditionalAnalysis/EECI_Graphs.do	FigOA5a.eps, ···, FigOA5e.eps	See Section 5.2 for more details
Figure OA6 (a)-(j)	Matlab/AgeGrpEEmP.m	FigOA6a.eps, ···, FigOA6j.eps	
Figure OA7	Stata/AdditionalAnalysis/BraiseRIPPropGr.do	FigOA7.eps	See Section 5.2 for more details
Figure OA8 (a), (b)	Stata/AdditionalAnalysis/BraiseRIPPropGr.do	FigOA8a.eps, FigOA8b.eps	See Section 5.2 for more details
Figure OA9	Matlab/EE_Figures.m	FigOA9.eps	
Figure OA10	Matlab/timing_around_GR.m	FigOA10.eps	
Figure OA11 (a)-(c)	Matlab/EE_EN.m	FigOA11.eps,···, FigOA11c.eps	
Table 1	–	Created directly in the paper	
Table OA1	Stata/Imputation/0_master_imputation.do	ImputationResults.smcl	See Section 5.2 for more details
Table OA2	Stata/Imputation/0_master_imputation.do	ImputationResults.smcl	See Section 5.2 for more details
Table OA3	Stata/Imputation/0_master_imputation.do	ImputationResults.smcl	See Section 5.2 for more details
Inline # ("over 70%") on page 17	Stata/AdditionalAnalysis/PSSS_EEm.do	On screen	
Inline #s IPUMS-ABACUS	–	On screen	See Section 5.3.2 for more details
Inline #s SIPP	SIPP/SIPP_2014_panel_movers.do	On screen	See Section 5.3.4 for more details

References

Bureau of Labor Statistics, "Employment Situation Summary: Current Population Survey: Civilian Unemployment (Seasonally Adjusted) - LNS13000000," 2022. U.S. Department of Labor, https://data.bls.gov/cgi-bin/surveymost?bls, (accessed January 31, 2023).

⎯ , "Job Openings and Labor Turnover Survey (JOLTS): Quits rate, Total Nonfarm (Seasonally adjusted) - JTS000000000000000QUR; Job Openings Level, Total nonfarm (Seasonally adjusted) - JTS000000000000000JOL," 2022. U.S. Department of Labor, https://data.bls.gov/cgi-bin/surveymost?jt, (accessed January 31, 2023).

Census Bureau, "Survey of Income and Program Participation 1996-2014," 2014. https://www.census.gov/programs-surveys/sipp/data/datasets.html. The Department of Commerce.

⎯ , "Basic Monthly CPS (1994-2022)," 2022. U.S. Department of Commerce, https://www.census.gov/data/datasets/time-series/demo/cps/cps-basic.html, (accessed on various dates).

⎯ , "Longitudinal Employer-Household Dynamics: Version R2022Q4," 2022. U.S. Department of Commerce, https://lehd.ces.census.gov/data/, (accessed January 31, 2023).

Fallick, Bruce and Charles Fleischman, "Employer-to-Employer Flows in the U.S. Labor Market: The Complete Picture of Gross Worker Flows," 2004. Federal Reserve Board Finance and Economics Discussion Series 2004-34.

⎯ **and** ⎯ , "Data for Employer-to-Employer Flows in the U.S. Labor Market: The Complete Picture of Gross Worker Flows," 2019. https://www.federalreserve.gov/econresdata/researchdata/feds200434.html. (Accessed January 31, 2023).

IPUMS, "Abacus." https://usa.ipums.org/usa/abacus.shtml. (Accessed January 31, 2023).

Moscarini, Giuseppe and Fabien Postel-Vinay, "Wage Posting and Business Cycles," *American Economic Review Papers and Proceedings*, 2016, *106* (5), 208–213.

NBER, "US Business Cycle Expansions and Contractions," 2023. https://www.nber.org/research/data/us-business-cycle-expansions-and-contractions.