**April, 2021**

**Workforce of the Future Initiative**

**The CPS month-to-month transitions dataset**

#### **Summary**

We construct a dataset of job-to-job transitions using the Integrated Public Use Microdata Series (IPUMS) of the Current Population Survey (CPS) between 2003 and 2019.  We choose the CPS over other sources for its fidelity in representing the population and its high resolution in monthly observations.

The CPS basic survey is administered monthly to nearly 70,000 households in the US, recording relevant labor force and demographic information from its respondents across eight months, taking an eight-month break between the fourth and fifth samples. We use CPS person-level identifiers (CPSIDP)[[1]](#footnote-1) to match individuals’ answers across consecutive months, which allow us to identify month-to-month changes in workers’ occupation, industry, employment status, educational attainment, and other measures. Each basic monthly sample in the CPS is made of eight equally sized cohorts called rotation groups. Each month, the most recent rotation group answers for its first time, another for its second time, and so on through, which means that approximately 75 percent of the sample is common between two consecutive months.

Even though the CPS shows the unemployed person’s last occupation, we restrict this matching process to employed individuals. We also exclude matches that present differences in race, gender, or age information for the same individual across samples. The omission of unemployed workers helps exclude job changes starting from relative job insecurity. These presumably are not as representative of workers’ preferences or skills as decisions made from a stable situation. We also exclude the observations with imputed occupation or industry codes due to non-responses or refusals to answer.

The probability of being dropped from the sample correlates with demographic information. To correct for this, we reweight the new sample using inverse probability weighting to reobtain a representative sample[[2]](#footnote-2).

The final match of individuals across every CPS basic monthly sample within the period 2003–2019, results in a dataset with 8.1 million observations and approximately 228 thousand occupational transitions (equivalent to a transition rate of 2.8 percent). We define occupational transitions as month-to-month changes in occupational codes, excluding switches made by self-employed workers, transitions with unverifiable information about changes in last month’s employer[[3]](#footnote-3), active members of the armed forces, and individuals obtaining a higher educational degree over that period. These later filters diminish the transition rates by a small margin. **Table 1** shows the overall characteristics of the resulting dataset.

**Table 1. Broad characteristics of the CPS transitions’ sample**

Table

Description automatically generated

*Notes: The share of upward transitions is the aggregation of a relative mobility metric. By construction 50 percent of transitions are upward and 50 are not. We will present more details on such metrics in our upcoming report.*

#### **Access raw data**

Anyone interested in accessing the raw data should go to the IPUMS CPS website (Flood, et al. 2020), register, and then pick up the CPS monthly samples and variables of interest.

Tentative variables of interest:

* CPSID
* OCC2010
* OCC
* IND
* MHIS
* AGE
* SEX
* HISPAN
* RACE
* EMPSTAT
* WKSTAT
* EMPSAME
* ACTSAME
* CLASSWKR
* You can continue adding variables as you wish, IPUMS has a search function to find variables per topic that can be helpful.

#### **Making the CPS transitions dataset compatible with other datasets**

We aggregate transitions and other occupation level information custom occupational scheme called SOCXX[[4]](#footnote-4). This final occupational classification is a hybrid between the detailed, broad, major, and minor versions of the Standard Occupational Codes 2010, with 437 final occupations instead of the 840 available in its original version. This aggregation process makes it compatible with both information available by Census Occupational Codes (those extracted from the CPS and the ACS) and by Standard Occupational Codes (as those extracted from the BLS, O\*NET and BGT).

In total, 428 of these SOCXX codes were successfully integrated into the CPS transitions data, representing the set of occupations we use to create the network model of the labor market. This is the final set of occupations presented across the report, and accounts for 96 percent of total employment if according to estimates from the OES 2019.

We developed a similar scheme for industries that crosswalks Industry Census Codes into hybrid NAICS codes, referred to as NAICSXX.

#### **References**

Bollinger, C. R., & Hirsch, B. T. (2006). Match bias from earnings imputation in the current population survey: The case of imperfect matching. *Journal of Labor Economics*, *24*(3), 483–519. <https://doi.org/10.1086/504276>

Fujita, S. (2019). *Measuring Employer-to-Employer Reallocation*.

Julia Rivera Drew, Sarah Flood, and John Robert Warren, “Making Full Use of the Longitudinal Design of the Current Population Survey: Methods for Linking Records Across 16 Months,” Journal of Economic and Social Measurement 39, no. 3 (2014): 121–44.

Sarah Flood, Miriam King, Renae Rodgers, Steven Ruggles and J. Robert Warren. Integrated Public Use Microdata Series, Current Population Survey: Version 8.0 [dataset]. Minneapolis, MN: IPUMS, 2020. <https://doi.org/10.18128/D030.V8.0>

1. See Rivera et al. (2014) for more a throughout discussion of how to prepare the CPS for longitudinal analysis.  [↑](#footnote-ref-1)
2. See Bollinger and Hirsch (2006) for more details on why invalid answers are not randomly distributed and how to reweight CPS samples accordingly. [↑](#footnote-ref-2)
3. Fujita (2020) makes a detailed explanation of why the inclusion/exclusion of these records can affect aggregated mobility outcomes and designs a method to estimate missing answers. We have decided to include those observations but silenced the transitions associated with them. [↑](#footnote-ref-3)
4. This occupational classification constitutes a revised version of the SOC codes available in the [Census Bureau Occupation Code List Crosswalk](https://www.census.gov/topics/employment/industry-occupation/guidance/code-lists.html) and is made up of 446 occupations in total, while the 2010 Standard Occupation Classification System used in the OES 2018 has 808. [↑](#footnote-ref-4)