Migration Patterns Data Dictionary

Data Files:

- *od_pooled.csv* – Dataset contains count and fractions of people who move between each origin and destination (OD) commuting zone pair. These data are unique at the OD level. The variable **pool** takes the value "pooled".
- *od_race.csv* – Dataset contains count and fractions of people who move between each OD commuting zone pair separately by race/ethnicity. These data are unique at the OD x race level. The variable **pool** takes the values "Asian", "Black", "Hispanic", "White", and "Other". Note that "Other" includes not only other races but also people who are not linked to the Decennial Census or ACS.
- *od_inc.csv* – Dataset contains count and fractions of people who move between each origin and destination commuting zone pair separately by parental income quintile. These data are unique at the OD x income quintile level. The variable **pool** takes the values "Q1" (poorest parental income quintile), …, "Q5" (richest parental income quintile).
- *od.csv* – Dataset contains count and fractions of people who move between each origin and destination commuting zone pair separately by race/ethnicity and parental income quintile. These data are unique at the OD x race/ethnicity x parental income level. The variable **pool** takes the values of the pairwise combinations of race ("Asian", "Black", "Hispanic", "White", and "Other") and parental income quintile ("Q1" - "Q5"). Note that the "Other" race/ethnicity category includes not only other races but also people who are not linked to the Decennial Census or ACS.

Sample:

The sample includes all children in the Census Numerical Identification Database (Numident) of Social Security Number holders who are born in the U.S. between 1984-92. To ensure our estimates preserve anonymity and satisfy differential privacy requirements we infuse a small amount of noise to each cell. In practice, this adds or subtracts a couple of people from each demographic-by-origin-by-destination migration count cell. As a result, some cells contain negative values.

Location at age 16 and 26 is assigned using Census, tax, and HUD information. Child race is measured using information from the 2010 Decennial Census and American Community Survey (ACS). Parental income is measured as a 5-year average of family income when the children are aged 14-18 based on the tax form 1040 of the parent who claims them as a dependent.

Geographic variables:

- **o_cz**: The numeric identifier of the commuting zone (CZ) in which the individual resided in childhood/at age 16 (referred to as origin CZ hereafter).
- **o_cz_name**: The name of the origin CZ.
- **o_state_name**: The name of the state the origin CZ is in.
- **d_cz**: The numeric identifier of the CZ in which the individual resides in young adulthood/at age 26 (referred to as destination CZ hereafter).
- **d_cz_name**: The name of the destination CZ.
- **d_state_name**: The name of the state the destination CZ is in.

Migration variables:

- **n**: The number of individuals from origin *o* living in destination *d*.
- **n_tot_o**: Total number of individuals from origin *o*. Generated by summing *n* across d_cz cells for each o_cz.
- **n_tot_d**: Total number of individuals who live in destination *d*. Generated by summing *n* across o_cz for each d_cz.
- **pr_d_o**: Pr{D|O}. The probability an individual lives in destination *d* given they grew up in origin *o*. Generated as n/n_tot_o.
- **pr_o_d**: Pr{O|D}. The probability an individual is from origin *o* given they live in destination *d*. Generated as n/n_tot_d.