**Data archive index for “A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration”**

**Ran Abramitzky, Leah Boustan, Katherine Eriksson**

Step 1: Use the .do file “build\_panel\_forjpe.do” in the *Build\_data* folder to compile matched data on US natives and immigrants from 16 sending countries from the 1900-1910-1920 Census. The raw datasets that are the input to this .do file are proprietary and belong to Ancestry.com and Familysearch.org; therefore, we do not provide the raw files here. However, this .do file will be instructive to interested readers to understand the variable definitions in the main analysis file. The output from this file (which we do provide in the *Data\_analysis* folder, and which is the main data for the analysis in the paper) is **panel\_all\_forjpe.dta**.

Step 2: Use the .do file “ipums\_forjpe.do” in the *Build\_data* folder to read in and clean public-use Census data from 1900-1910-1920. We do not provide the raw data that is the input to this .do file because the dataset is large and easily downloadable from [www.ipums.org](http://www.ipums.org). The output from this file is **ipums\_forjpe.dta**, which is in the *Data\_analysis* folder.

Step 3: Use the .do file “analysis\_forjpe.do” in the *Data\_analysis* folder to replicate the tables and figures in the published paper. This .do file uses three datasets as inputs:

1. panel\_all\_forjpe.dta (see step 1)
2. ipums\_forjpe.dta (see step 2)
3. **urb\_rural\_occscores.dta** which contains mean incomes by occupation for urban workers for use in replicating Table 4. This data is provided in the *Data\_analysis* folder.

Note that the .do file “analysis\_forjpe.do” replicates all tables and figures in the paper except Table 7 and Figures 1 and 6.

Step 4: To replicate Table 7 and Figure 1 which use the HISCLASS occupation classification scheme, see the .do file “figure1\_table7\_jpe.do” in the *HISCLASS* folder. This .do file uses as inputs:

1. panel\_all\_forjpe.dta (see step 1)
2. ipums\_forjpe.dta (see step 2)
3. **hisclass50.dta** which matches the 1950-based occupation codes to HISCLASS classifications. This data is contained in the *HISCLASS* folder.

Step 5: To replicate Figure 6, see the .do file “figure6\_jpe.do” in the *Figure6* folder. This .do file uses as inputs:

1. panel\_all\_forjpe.dta (see step 1) for first generation immigrants and comparison natives
2. **f1st2ndGen\_onePtPerCtry.dta** for second generation immigrants and comparison natives. This data is contained in the *Figure6* folder.

In addition, you will need the .ado file “addFakeObs.ado,” which is also in the *Figure6* folder.