README file for Replication Package of " 'Workhorses of Opportunity': Regional Universities Increase Local Social Mobility"

Greg Howard* Russell Weinstein†

August 13, 2025

This README file contains the programs and data to replicate the figures and tables in the paper. The IPUMS USA Usage License does not allow redistribution of any data from the IPUMS USA Full Count Data (IPUMS, 2025). We include the programs used to input the data into Stata, as well as the analysis programs, but we do not include the data for this portion of the analysis. As the full count data are publicly available on the IPUMS USA website (`https://usa.ipums.org/usa/cite.shtml`), individuals may download the data from the website and then run the programs in this replication package.

Our analysis of The Freshman Survey (TFS) data utilizes a restricted dataset at the individual-level that includes the survey year and the following characteristics of the individual's university: TFS university ID (acerecode), IPEDS ID, name, state, city, county, and zipcode. We obtained one observation per TFS university ID and year in the restricted data, and merge to the public-access data using TFS university ID and year. Thus, the restricted dataset could also be at the university-survey year level.

For a fee, researchers can request these restricted data through an application process (`https://heri.ucla.edu/data-access-for-researchers/`). The public access raw data files are available for download through the HERI Data Archive (`https://heri.ucla.edu/`

*University of Illinois, Urbana-Champaign. glhoward@illinois.edu
†University of Illinois, Urbana-Champaign. weinst@illinois.edu

heri-data-archive/). Individuals need to register for the archive. After registering and receiving a username and password, the public access data we use for this project can be found by navigating to the The Freshman Survey (TFS) Trends Datasets. The data are available as SPSS files, which we then converted to Stata files. Because HERI requires users to register for the public-access data, we have not shared any of the TFS data.

For all analysis, we include the programs for replicating our results. For all analysis except the restricted data described above we include the raw data, programs, and final files for analysis.

To run the replication file, users should run the file $Maindofile_allinputs.do$. This will run all the figures and tables in the paper except those using restricted data as described above. The do files contain comments describing which files create which tables and figures. If a user obtains the restricted-access data described above and puts them in the replication folder, the do file will run all of the figures and tables in the paper. The do files for inputting the raw full count census data (starting with $usa_$) also explain the filters used when requesting the full count census data from IPUMS USA.

References

IPUMS. 2025. "IPUMS New USA Registration." https://uma.pop.umn.edu/usa/user/new?return_url=https%3A%2F%2Fusa.ipums.org%2Fusa-action%2Fmenu, Accessed 5/21/25.