

README improved version

Data Documentation : README intial version

This document describes the meaning of files that are necessary to replicate results in "International Students and Labor Market Outcomes of U.S. Natives" at the *Canadian Journal of Economics* by Murat Demirci.

"Demirci_CJE_2020.dta" is the data file in STATA format. The data set includes data from the American Community Survey for the years between 2009 and 2012. The sample is restricted to recent graduate bachelor's-degree holders (i.e., age below 24) and doctoral degree holders (aged below 31) and all master's degree holders in the data.

"Demirci_CJE_2020_figure1.dta" is the data file that is necessary to replicate Figure 1. It includes the IPEDS data that show the number of degrees granted to international students between 1977 and 2015.

"Demirci_CJE_2020_figure2.dta" is the data file that is necessary to replicate Figure 2. It includes the SEVIS data that show the potential and actual labor supply of international students on F-visas between 2004 and 2012.

"Demirci_CJE_2020.do" is the code in STATA format. It replicates all tables and figures of the paper. Table 1 and Table 2 are summary statistics. The code generates them while running, and this output can be accessed via the log file. Table 3 to Table 6 are regression outputs. The code runs them quietly and save the results in Excel files, named as Table 3 and so on. Both Figure 1 and 2 have six sub-panels. The code generates and saves them in .PNG format.

Source for the American Community Surveys

Ruggles, S. J., T. Alexander, K. Genadek, R. Goeken, M. B. Schroeder and M. Sobek (2010): Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database], Minneapolis: University of Minnesota. <https://usa.ipums.org/usa/> (accessed March 4, 2015).

Source for the IPEDS Data

Integrated Postsecondary Education Data System Completion Surveys. 1977-2015. National Center for Education Statistics. <https://nces.ed.gov/ipeds/use-the-data> (accessed April 21, 2018).

Source for the SEVIS Data

The data about the supply of international students are obtained from the United States Immigration and Customs Enforcement via the Freedom of Information Act Request. (See <https://www.ice.gov/foia/overview>.) The data set includes visa and educational records of international students who completed their studies between 2004 and 2014 from the Student and Exchange Visitor Information System (SEVIS).

Source for the National Survey of Recent College Graduates

The descriptive features of labor market outcomes of recent college graduates that are reported in the bottom panel of Table 1 are tabulated based on the 2008 and 2010 waves of the National Survey of Recent College Graduates. The restricted-use version of these data where all variables can be observed was used at the University of Virginia. The publicly available versions are available at <https://sestat.nsf.gov/datadownload/>.

Additional informations

Raw data

- ☒ figure1_raw.dta : IPEDS raw data for Figure 1
- ☒ SEVIS_aggregates.dta : aggregated raw data on international students labor supply
- ☒ Demirci_CJE_2020.dta : ACS raw data for econometric estimation
- ☒ 2004-2013SEVIS.dta : SEVIS raw data for Figure 2

Adding missing data cleaning code

Here are codes that allow data cleaning or processing to obtain the analysis data :

- ☒ figure1_clean.do : code that prepares IPEDS raw data for Figure 1
- ☒ figure2_clean.do : code that prepares the raw SEVIS data for Figure 2
- ☒ SEVIS_aggregates_clean.do : code that generates the SEVIS_aggregates.dta

Diagramm for onnecting display items with inputs

From this diagram you can clearly see the raw data transformation into analysis data, and cleaning and analysis codes that have been used. Also you can distinguish the data files merge that produced the final analysis data used for econometric estimations.

FIGURE 1

```

└─ Demirci_CJE_2020.do
    └─ Demirci_CJE_2020_figure1.dta
        └─ figure1_clean.do
            └─ figure1_raw.dta

```

FIGURE 2

```

└─ Demirci_CJE_2020.do
    └─ Demirci_CJE_2020_figure2.dta
        └─ figure2_clean.do
            └─ 2004-2013SEVIS.dta

```

TABLE X

```

└─ Demirci_CJE_2020.do
    └─ Demirci_CJE_2020.dta
        └─ MERGE
            └─ Demirci_CJE_2020_1.dta
                └─ SEVIS_aggregates_clean.do
                    └─ Demirci_CJE_2020_0.dta
            └─ SEVIS_aggregates.dta
                └─ SEVIS_aggregates_clean.do
                    └─ 2004-2013SEVIS.dta

```

Unused data sources:

Not available

Unused analytic data:

None - all analytic data files were used

NB : Demirci_CJE_2020.dta_0 and Demirci_CJE_2020.dta_1 files are only the initial and cleaned version of Demirci_CJE_2020.dta analysis data file.

More information on access to raw data

- **Access to American Community Surveys data**

This data is available on IPUMS, a website for downloading socio-economic data on the US population. The only requirement is that you have to create an account. Also, there is a time lag (usually less than 10 minutes) between submitting the selected data and receiving them.

- **Access to American Community Surveys data**

It is not allowed to share the raw SEVIS data. Anyone interested can access them by submitting a request to United States Immigration and Customs Enforcement.

Here is some useful informations :

- ☒ **Institution:** United States Immigration and Customs Enforcement
- ☒ **URL:** <https://www.ice.gov/>

Contact information

- ☒ *Generic email:* ICE-FOIA@dhs.gov
- ☒ *Specific contact*
 - *Department:* Freedom of Information Act (FOIA) Office
 - *Email:* ICE-FOIA@dhs.gov
 - *Phone number:* (866) 633-1182
 - *Fax:* (202) 732-4265
 - *Address:* 500 12th Street, S.W., Stop 5009, Washington, D.C. 20536-5009

Conditions of access

- ☒ **Restrictions:** On site only
- ☒ **Requirements:** Submit a Freedom of Information Act (FOIA) request
- ☒ **Fees:** No fees
- ☒ **Type of data:**
 - ☒ Raw data only
 - ☐ Analysis data only
 - ☐ Raw and analysis data