Readme for Frias, Kaplan, Verhoogen, and Alfaro-Serrano, "Exports and Wage Premia: Evidence from Mexican Employer-Employee Data," Review of Economics and Statistics
Feb. 6, 2022

I. Introduction

This readme explains how to replicate the results in the above-referenced paper. Please refer to the paper for details on the names of datasets and how they were processed.

The datasets used in the paper are confidential but our access is not exclusive. The data to replicate the main results can be accessed at the Microdata Laboratory of INEGI, the Mexican statistical agency, in Mexico City.

The person to contact at the laboratory is:

Dra. Natalia Volkow
Director of Microdata Access
Instituto Nacional de Estadistica y Geografia (INEGI)
Avenida Héroe de Nacozari Sur 2301, Fraccionamiento Jardines del Parque, Aguascalientes, Aguascalientes, 20276, Mexico
Email: NATALIA.VOLKOW@inegi.org.mx
Tel: +52 55 5278 1000

The analysis using only IMSS data was carried out at the IMSS offices, while Judith Frias was on staff. Given the turnover of personnel at IMSS, we suggest that those interested in replicating our results first contact us. We will then put you in touch with the right person.

This readme is divided into three parts. Section II lists the datasets that are available outside of both IMSS and INEGI, and the programs to analyze them. Section III lists the datasets and programs needed to conduct the analysis that only uses IMSS data, which was conducted in the IMSS offices. Section IV lists the datasets and programs needed to conduct the analysis combining the IMSS and INEGI data, which was conducted in the INEGI Microdata Laboratory.

The programs are included in the replication package. Once the data have been accessed in the respective locations, the programs should be run in the order in which they are listed below.

II. Analysis using publicly available data

A. Datasets

1. Filename: banxicopriceindices19732006.dta. Contains Consumer Price Index and Producer Price Index, downloaded from the Banco de Mexico (https://www.banxico.org.mx/SieInternet/).

2. Filename: exchangerates.dta. Nominal peso/dollar exchange rate and consumer price indices for US and Mexico.

3. Filename: taggstats.dta. Contains aggregate labor force statistics from the 2005 Anuario Estadistico

de los Estados Unidos Mexicanos.

4. Filename: sat_salarios_minimos_1982_2008.dta. Contains minimum wages over time for the three minimum-wage zones in Mexico, originally downloaded from the Servicio de Administracion Tributaria (SAT), the Mexican tax authority, from the website:

http://www.sat.gob.mx/sitio_internet/asistencia_contribuyente/informacion_frecuente/salarios_minimos/45_7369.html

The link is no longer active. But the same information can be downloaded from the Comision Nacional de los Salarios Minimos (CONSAMI):

https://www.gob.mx/conasami/documentos/tabla-de-salarios-minimos-generales-y-profesionales-por-areas-geograficas

5. Filename: munwzones_updated2.dta. Contains a concordance between IMSS and INEGI geographic codes, made by hand from lists of geographic codes from the two institutes.

B. Code to analyze publicly available data

1. housekeeping-public.do. Sets paths and program parameters, to be called at the beginning of each subsequent program. Will need to be modified to fit local file structure.

2. auxiliary.do. Prepares the above datasets to be integrated in later analysis.

3. auxtablesfigs.do. Makes Table B1 and Figure 1, which only require publicly accessible data.

III. Analysis at IMSS

A. Datasets

1. Filenames: Ssci_ciz1_1m.txt, Ssci_ciz1_2m.txt, Ssci_ciz1_3m.txt, Ssci_ciz2_1m.txt, Ssci_ciz2_2m.txt, Ssci_ciz2_3m.txt, Ssci_ciz2_4m.txt, Ssci_ciz3_1m.txt, Ssci_ciz3_2m.txt, Ssci_ciz3_3m.txt, Sshc_ciz1_1m.txt, Sshc_ciz1_2m.txt, Sshc_ciz1_3m.txt, Sshc_ciz1_4m.txt, Sshc_ciz2_1m.txt, Sshc_ciz2_2m.txt, Sshc_ciz2_3m.txt, Sshc_ciz2_4m.txt, Sshc_ciz3_1m.txt, Sshc_ciz3_2m.txtSShc_ciz3_3m.txt, ASEGURADOS_20060905.TXT. Contain raw individual-level wage spells from IMSS administrative records.

2. imssdir_zona.dta. Contains a mapping from IMSS establishment ID to geographic areas (municipalities and minimum wage zones), derived from an IMSS directory file of establishment locations.

B. Programs

1. housekeeping-imss.do, housekeeping-imss.sas, housekeeping-imss.m. Set paths and program parameters. Called at the beginning of each program. Will need to be modified to fit local file structure.

2. readimssdata.sas. Reads raw text files with individual-level wage spells.

3. cleanimss.do. Does initial simple cleaning of individual-level wage-spell data.

4. makerawworkhistories.do. Constructs work histories based on data after simple cleaning.

5. morecleanimss.do. Does a second round of processing of individual-level wage information.

6. makeworkhistories.do. Constructs work histories based on data after additional cleaning.

7. AKMestimation.do, Pcg12.m, FindConnectedSet5.m. Estimate two-way fixed models for periods 1988-1990, 1992-1994, 1996-1998, 2000-2002, following Abowd, Kramarz, and Margolis (Econometrica, 1999; hereafter AKM) and Card, Heining, and Kline (QJE, 2013; hereafter CHK). AKMestimation.do calls two Matlab programs: FindConnectedSet5.m to find largest connected set of establishments, and Pcg12.m, which implements a pre-conditioned conjugate gradient (PCG) algorithm, following CHK.

8. JobChangesCoworkers.do. Generates Figures 3, 4, B2 and B3.

9. JobChangesCoworkers-nominal.do. Generates Figures B4 and B5 (similar to Figures 3 and 4 but using nominal wages).

10. AKMdiagnostics.do. Generates Table 3.

11. preparedata-kss.do. Prepares data to be passed to Matlab programs for "leave-one-out" estimation.

12. dkss19921994.m, dkss19961998.m, dkss20002002.m, exportkss.m, leave_out_COMPLETE_vDAS.m: Together implement "leave-one-out" estimators from Kline, Saggio and Solvsten (Econometrica, 2020; hereafter KSS). dkss19921994.m, dkss19961998.m, dkss20002002.m implement the KSS estimator for the periods 1992-1994, 1996-1998, and 2000-2002. These programs require leave_out_COMPLETE_vDAS.m, which should be placed in the main working directory. They also require the replication programs from KSS, which should be placed in a subdirectory named KlineSaggioSolvstenEMA2020-code/ in the main working directory (or the subdirectory should be using "kss" in housekeeping.m).

13. KSStable.do. Generates Table B5.

14. IndivDescripStats.do. Generates Tables 1, B2.

15. wagepercentiles.do. Generates Figure B1.

16. makelistsofconnectedsets.do. Makes lists of establishments in largest connected sets in each period.

14. stayerswages.do. Calculates average wage changes at the establishment level for workers who stay in same establishment across periods.

15. makeimssonlypanel.do. Constructs balanced establishment-level panel, including AKM decomposition, using only IMSS data (for eventual use in Table 7).

16. packforinegi.do. Makes two establishment-level datasets to be transferred to INEGI: (1) a dataset

combining the AKM estimates, the panel of stayers' wages, and the IMSS-only panel (filename: imss-inegi.dta), and (2) a dataset containing lists of establishments in the largest connected sets in each period (filename: imss-inegi-sets.dta).

IV. Analysis at INEGI Microdata Laboratory

A. Datasets

The following datasets are accessible at the INEGI microdata laboratory.

1. Filename: eia_enestyc_eric_tesis.dta. This file is the panel constructed for: Eric Verhoogen, "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector," (QJE, 2008). See the appendix of that paper for details. The file combines information the EIA sample first drawn in 1993 (covering 1993-2001 in this file) and information from the 1992, 1995 and 1999 waves of the ENESTyC. (The files also contains information from the EIA panel first drawn in 1984, covering 1984-1994, although that information is not used in this paper.) INEGI refers to this dataset as Eric Verhoogen's "tesis" dataset, since it was part of his doctoral thesis.

2. Filename: lad_eia_1994_2003.dta. This is INEGI's standard file for the 1993-2003 EIA panel.

3. Filename: eim_var_1994_2008.dta. This is INEGI's standard file for the 1993-2003 EIM panel.

4. Filename: links_eia_imss.dta. Contains a crosswalk between INEGI establishment IDs and IMSS establishment IDs. Originally created at INEGI under the aegis of a data-sharing agreement between IMSS and INEGI.

5. Filenames: imss-inegi.dta, imss-inegi-sets.dta. Datasets derived from IMSS data, created by packforinegi.do (see above).

B. Programs

1. housekeeping-inegi.do. Sets paths and program parameters. Called at the beginning of each program. May need to be modified to fit local file structure.

2. unpackimssdataatinegi.do. Reads the datasets that were transferred from IMSS to INEGI, imss-inegi.dta and imss-inegi-sets.dta, and prepares them for further analysis.

3. add20022003.do. Extends the "tesis" dataset until 2003, using INEGI's EIA files for 2002 and 2003.

4. clean-eia19932003.do. Cleans the 1993-2003 EIA panel, following steps very similar to those in Verhoogen (QJE, 2008).

5. addregistro.do. Uses IMSS-INEGI crosswalk of plant IDs (links_eia_imss.dta) to add IMSS plant ID to cleaned EIA 1993-2003 panel.

6. prepareeiaimssdata-plantyear.do. Merges AKM estimates and other IMSS information into EIA panel. Also constructs predicted exporter, predicted export share, and TFP variables at plant-year level.

7. prepareeiaimssdata-plantperiod.do. Collapse EIA-IMSS panel to plant-period-level.

8. extendeiaimsspanel.do. Extends EIA-IMSS panel using IMSS-only information from previous years (1988-1990).

9. eiaimssfigs.do. Makes Figure 2.

10. eiaimsstables.do. Makes tables that use EIA or EIA-IMSS data: Tables 2, 4, 5, 6, 7, B3, B4, B6, B7, B8, B9, B10, B11.