

Instructions for Reproducing “Productivity Differences Between and Within Countries”

Daron Acemoglu Melissa Dell

July, 2009.

1 Geospatial Data

This section describes how to construct the municipality geospatial (GIS) measures used in the paper. For more information on using the geospatial software required to run these calculations, you may wish to consult the GIS course lecture notes available at the following link: <http://econ-www.mit.edu/grad/mdell/teaching>.

1.1 Population Data

To reproduce the municipality-level population measures, follow these steps:

1. Obtain the municipal GIS boundaries. The sources are as follows:
 - (a) *All Latin American countries besides Mexico and Peru:* These boundaries were obtained from the International Center for Tropical Agriculture (CIAT), *Sub National Boundaries of South and Central America GIS Dataset*. CIAT can be contacted as follows:

International Center for Tropical Agriculture
A.A. 6713; Cali, Colombia
Phone: +1 (650) 8336625
Fax: +1 (650) 8336626
Email: ciat@cgiar.org.

The boundaries database is not for sale, but CIAT may choose, at their discretion, to distribute it to researchers after conversing about the intended use of the data.

- (b) *Peru:* The CIAT data contains only provincial - not municipal - boundaries for Peru. Hence we obtained Peruvian municipal boundaries from the Peruvian Ministry of

Transportation. At the time of writing this document, the Ministry of Transportation did not have a functioning Web site, and the only way we know to obtain these proprietary data is to go in person or to call the Ministry of Transportation in Lima (address: Jiron Zorritos 1203, Lima, Peru; phone: 51-1-6157800).

- (c) *Mexico*: The CIAT data were missing some municipalities in Mexico, so instead we used the Earth Science Research Institute's (ESRI) "Municipalities of Mexico, 2006" GIS database. This can be purchased from ESRI at <http://www.esri.com/>. (Note that many universities also have a subscription to the ESRI database.)
- (d) *United States*: We obtained GIS maps of U.S. "public use micro areas - 5 percent" from the U.S. Census Bureau. These data can be downloaded at the following link: http://www.census.gov/geo/www/cob/bdy_files.html.
- (e) *Canada*: We used a GIS map of Canadian provinces (called CANcd.shp) produced by the Canadian Census Bureau. These data may be purchased from the Canadian Census Bureau, by emailing infostats@statcan.gc.ca.

Once the GIS boundaries have been obtained, run the file `aej_boundaries.py`. This file extracts and (when necessary) combines the municipal boundaries so that they correspond to the units geo-referenced in the income surveys and censuses. (Note that in a number of surveys, municipalities with less than 10,000 or 20,000 people are aggregated for confidentiality reasons, and in others, data are geo-referenced to larger regions; `aej_boundaries.py` produces the boundaries for these units.) You must have Python 2.4, ArcGIS 9.2, and a license for the ArcGIS spatial analysis extension on your machine to run this file. Before running the file, you must modify the directory path at the top of the file, where it states "INSERT DIRECTORY PATH HERE". Note that Python requires directory paths to be specified with double slashes. Then put the following .dbf files - posted on the AEA Web site - in the chosen directory: `BR3codes.dbf`, `CA1codes.dbf`, `CI1codes.dbf`, `CIcensus_codes.dbf`, `COcensus_codes.dbf`, `COcodes.dbf`, `codes4AR.dbf`, `CRed_codes.dbf`, `EC1codes.dbf`, `EScodes.dbf`, `MXcodes.dbf`, `PMcodes.dbf`, `UYcodes.dbf`, and `VE3codes.dbf` (these contain the information matching the municipality codes in the GIS boundary files to the municipality codes in the income data, as described in the `aej_boundaries.py` file). You may also need to modify the location of the Arc toolboxes on your machine (under the heading "# Load required toolboxes..." at the top of the python file).

2. The second step is to obtain the 2000 GIS population data: *Global Rural-Urban Mapping Project, Alpha Version: Population Grids*. These may be downloaded, after filling out a registration form, at <http://sedac.ciesin.columbia.edu/gpw>. In order to use the code files provided here, please download the data in .ascii format.

3. Run the python file `aej_pop00.py`. This will produce a series of `.dbf` files. The output from steps (1) through (3) is provided in the files posted on the AEA Web site (`BLp.dbf`, `BRp.dbf`, `CAP.dbf`, `CIp.dbf`, `COp.dbf`, `ECp.dbf`, `ESp.dbf`, `GTP.dbf`, `HOp.dbf`, `MXp.dbf`, `PAP.dbf`, `PMp.dbf`, `USp.dbf`, `UYp.dbf`, `ARp.dbf`, `NUp.dbf`, `CI3p.dbf`, `CO3p.dbf`, `CR3p.dbf`, `EC3p.dbf`, `VEp.dbf`, `BRcensus3.csv`, and `MXcensus3.csv`). These can be placed in the desired directory to complete step (4) without doing steps (1) through (3).
4. Run the Stata file `aej_pop00.do`, in order to combine the output from ArcGIS into Stata files containing the municipality level population measures, by country. Note that you must modify the directory path at the top of the file, where it states “INSERT DIRECTORY PATH HERE”. The data produced in this step are used to weight the income and roads data for the paper’s inequality comparisons.

1.2 Climate Data

To reproduce the municipality-level climate measures, follow these steps:

1. Create the GIS boundary files, as described in the directions above for calculating the population averages.
2. Download the 30 arc second ESRI grids for max. temperature, min. temperature, and precipitation from the following link: <http://www.worldclim.org/current.htm>. There are a total of 36 files (12 monthly files for each of the climate measures). Note that these files are quite large.
3. Run the file `aej_climate.py`, changing the directory path at the top of the file to the desired location. Again, you must have Python 2.4, ArcGIS 9.2, and a license for the ArcGIS spatial analysis extension on your machine to run this file. This will produce a series of `.dbf` files. The output from steps (1) through (3) is provided in the files posted on the AEA Web site (`pre‘Y’‘X’.dbf`, `max‘Y’‘X’.dbf`, `min‘Y’‘X’.dbf`, where $X=1/12$ (data are by months), and $Y=BR, MX, PM, US, VE$ (FIPS country identifiers), `BRcensus3.csv`, and `MXcensus3.csv`). These can be placed in the desired directory to complete step (4) without doing steps (1) through (3).
4. Run the Stata file `aej_climate.do`, in order to combine the output from ArcGIS into Stata files containing the municipality level climate measures, by country. Note that you must modify the directory path at the top of the file as desired. The data produced in this step are used to calculate the partial correlations between incomes and proximity to roads that are reported in Table 3.

1.3 Terrain Data

To reproduce the municipality-level terrain measures, follow these steps:

1. Create the GIS boundary files, as described in the directions above for calculating the population averages.
2. Download the 30 arc second elevation data for the tiles covering the Americas from the following link: <http://www.dgadv.com/srtm30/>. Do not change the file names, as the provided python code files will merge these tiles together into a seamless raster.
3. Run the file `aej_elevslope.py`, changing the directory path at the top of the file to the desired location. This will produce a series of .dbf files. The output from steps (1) through (3) is provided in the files posted on the AEA Web site (`e3BR.dbf`, `e3MX.dbf`, `e3PM.dbf`, `e3US.dbf`, `e3VE.dbf`, `s3BR.dbf`, `s3MX.dbf`, `s3PM.dbf`, `s3US.dbf`, `s3VE.dbf`, `BRcensus3.csv`, `UScensus3.csv`, and `VEcensus3.csv`). These can be placed in the desired directory to complete step (4) without doing steps (1) through (3).
4. Run the Stata file `aej_elevslope.do`, in order to combine the output from ArcGIS into Stata files containing the municipality level terrain measures, by country. Note that you must modify the directory path at the top of the file as desired. The data produced in this step are used to calculate the partial correlations between incomes and proximity to roads that are reported in Table 3.

2 Income Data

The first step necessary to reproduce the income results in Tables 1, 2, and 3 is to obtain the proprietary income data used in the paper. Here are instructions, by country, for obtaining these data:

Bolivia: The income data are from *Encuesta de Hogares*, 2002. To purchase these data, please contact:

Instituto Nacional de Estadística

Calle Jos Carrasco No. 1391

La Paz, Bolivia 6129

Phone: (591-2) 2222333

Fax: (591-2) 2222885

Brazil: The income data are from the 2000 Brazilian Population Census 6% sample. To obtain these data, register with the IPUMS International Web site at <https://international.ipums.org/international/>.

Once access has been approved, follow the instructions on the IPUMS extract page to create an extract containing the following variables: serial, pernum, statebr, munibr, wtper, age, sex, yrschl, and inearn.

Upon creating and downloading the extract, IPUMS will provide code files for converting the ASCII extract into formats used by a number of statistical packages. In order to analyze the data with the code provided here, please use the IPUMS-provided Stata do file for reading the extract into Stata format and save the file as “BRcensus.dta.”

Canada: The income data are from the 2001 Canadian Population Census 2.5% sample. To obtain these data, register with the IPUMS International Web site at <https://international.ipums.org/international/>.

Once access has been approved, follow the instructions on the IPUMS extract page to create an extract containing the following variables: serial, pernum, provca, wtper, age, sex, edattand, and inearn.

Upon creating and downloading the extract, IPUMS will provide code files for converting the ASCII extract into formats used by a number of statistical packages. In order to analyze the data with the code provided here, please use the IPUMS-provided Stata do file for reading the extract into Stata format and save the file as “canada_pums01.dta.’

Chile: The income data are from *VI Encuesta de Presupuestos Familiares*, 2006-2007. To purchase these data, please contact:

Instituto Nacional de Estadísticas

Paseo Bulnes 418

Santiago de Chile, 8330532

Phone: (56-2) 366 7777

Fax: (56 2) 671 2169

Email: ine@ine.cl

Colombia: The income data are from *Encuesta de Calidad de Vida*, 2003. To purchase these data, please contact:

Departamento Administrativo Nacional de Estadística

Carrera 59 No.26-70 Interior I - CAN

Bogotá D.C., Colombia 611

Phone: (571) 5978300

Fax (571) 5978399

Email: dane@dane.gov.co

Costa Rica: The income data are from *Encuesta Nacional de Ingresos y Gastos de los Hogares*, 2004. To purchase these data, please contact:

Instituto Nacional de Estadística y Censos

Calle Los Negritos, Edificio Ana Lorena

Mercedes de Montes de Oca

San Jos, Costa Rica 10 163 - 1000

Phone: (506) 2253-7579

Fax: (506) 2224-2221

Email: informacion@inec.go.cr

Ecuador: The income data are from *Encuesta de Condiciones de Vida*, 2005-2006. To purchase these data, please contact:

Instituto Nacional de Estadística y Censos

Juan Larrea N15-36 y José Riofro

Quito, 17-15 130C

Phone: 2529858

Fax: 2509836

Email: planta_central@inec.gov.ec

El Salvador: The income data are from *Encuesta de Propositos Multiples*, 2006. To purchase these data, please contact:

Dirección General de Estadísticas y Censos Alameda Juan Pablo II y Calle Guadalupe Edificio C1 - C2, Centro de Gobierno

San Salvador, El Salvador

Phone: (503) 2231-5600

Guatemala: The income data are from *Encuesta Nacional de Condiciones de Vida*, 2000. To download these data (free of charge), please visit <http://iresearch.worldbank.org/lsmss/lsmssurveyFinder.htm>, agree to the World Bank's terms and conditions, and follow the instructions on the screen for downloading the data and documentation.

Honduras: The income data are from *Encuesta de Condiciones de Vida*, 2004. To purchase these data, please contact:

Instituto Nacional de Estadística

Lomas del Guijarro, edificio Plaza Guijarro, 5to.piso

Tegucigalpa, M.D.C., Honduras

Mexico: The income data are from the 2000 Population Census 10.6% sample. To obtain these data, register with the IPUMS International Web site at <https://international.ipums.org/international/>.

Once access has been approved, follow the instructions on the IPUMS extract page to create an extract containing the following variables: serial, pernum, statemx, munimx, wtper, age, sex, yrschl, and inearn.

Upon creating and downloading the extract, IPUMS will provide code files for converting the ASCII extract into formats used by a number of statistical packages. In order to analyze the data with the code provided here, please use the IPUMS-provided Stata do file for reading the extract into Stata format and save the file as “MXcensus.dta.”

Panama: The income data are from the 2000 Population Census 10% sample. To obtain these data, register with the IPUMS International Web site at <https://international.ipums.org/international/>.

Once access has been approved, follow the instructions on the IPUMS extract page to create an extract containing the following variables: serial, pernum, provpa, distpa, wtper, age, sex, yrschl, and inearn.

Upon creating and downloading the extract, IPUMS will provide code files for converting the ASCII extract into formats used by a number of statistical packages. In order to analyze the data with the code provided here, please use the IPUMS-provided Stata do file for reading the extract into Stata format and save the file as “PMcensus.dta.”

Paraguay: The income data are from *Encuesta Integrada de Hogares*, 2001. To purchase these data, please contact:

Dirección General de Estadística, Encuestas y Censos

Naciones Unidas y Saavedra - Fernando de la Mora

Asunción, Paraguay

Phone: (595 21) 511-016

Fax: (595 21) 508-493

Peru: The income data are from *Encuesta Nacional de Hogares*, 2001. To download these data (free of charge), please go to <http://www.inei.gob.pe/>, click on the link “microdatos” on the righthand side of the page, and follow the instructions provided (in Spanish only).

United States: The income data are from the 2000 Population Census 5% sample. To obtain these data, register with the IPUMS USA Web site at <http://usa.ipums.org/usa/>.

Once access has been approved, follow the instructions on the IPUMS USA extract page to create an extract containing the following variables: `pernum`, `perwt`, `statefip`, `puma`, `age`, `sex`, `educrec`, `inccarn`.

Upon creating and downloading the extract, IPUMS USA will provide code files for converting the ASCII extract into formats used by a number of statistical packages. In order to analyze the data with the code provided here, please use the IPUMS USA-provided Stata do file for reading the extract into Stata format and save the file as “`us_pums00.dta`.”

Uruguay: The income data are from *Encuesta de gastos y ingresos de hogares*, 2005-2006. To download these data (free of charge), please go to <http://www.ine.gub.uy/>, click on the link “microdatos” at the top of the page, and follow the instructions provided (in Spanish only).

Venezuela: The income data are from the 2001 Population Census 10% sample. To obtain these data, register with the IPUMS International Web site at <https://international.ipums.org/international/>.

Once access has been approved, follow the instructions on the IPUMS extract page to create an extract containing the following variables: `serial`, `pernum`, `stateve`, `munive`, `wtpcr`, `age`, `sex`, `yrschl`, and `inccarn`.

Upon creating and downloading the extract, IPUMS will provide code files for converting the ASCII extract into formats used by a number of statistical packages. In order to analyze the data with the code provided here, please use the IPUMS-provided Stata do file for reading the extract into Stata format and save the file as “`VEcensus.dta`.”

Save these income data in the desired directories (using the file names assigned by the national statistical agencies). Make sure the GIS outputs produced by following the steps in Section 1 are also stored in the appropriate directories. (The code for extracting the income data is designed so that data from each country can be stored in a separate directory, as some of the files from national statistical agencies from different countries have the same names.) The paper’s results can then be reproduced by running a series of Stata and Matlab programs in the order listed below. Note that you must insert the appropriate directory path in each `.do` file where it states “INSERT DIRECTORY PATH HERE”. The top of each of the `.do` or `.m` files also contains a list of all input files required to run the program, a description of all output files produced, details of how to obtain any Stata add-on packages needed to run the program, and a detailed description of what the code file does. The code is commented throughout. Because some of the datasets are extremely large, computation requires a substantial amount of memory and if possible will be greatly aided by the use of Stata MP (multi-processor).

Code files to reproduce Tables 1 and 2

1. `aej_yearn.do` - Extracts income and demographic data and merges it with population data. Output is used by `aej_mincer.do`, `aej_dcy.do`, and `aej_dcycntry.do`.
2. `aej_mincer.do` - Decomposes income into predicted and residual components for males between the ages of 18 and 55. Output is used by `aej_dcysmall.do`, `aej_dcxb.do`, `aej_dcyres.do`, and `aej_dcrescntry.do`.
3. `aej_dcy.do` - Decomposes income inequality into cross-country and within-country components, all individuals. Output is used by `aej_inequality.m`.
4. `aej_dcycntry.do` - Decomposes income inequality into within and between municipality components, by country, all individuals. Output is used by `aej_inequality.m`.
5. `aej_dcysmall.do` - Decompose total income inequality for 18 to 55 year old males into between country and within country components. Output is used by `aej_inequality.m`.
6. `aej_dcxb.do` - Decompose predicted income inequality for 18 to 55 year old males into between country and within country components. Output is used by `aej_inequality.m`.
7. `aej_dcyres.do` - Decompose residual income inequality for 18 to 55 year old males into between country and within country components. Output is used by `aej_inequality.m`.
8. `aej_dcrescntry.do` - Decomposes total, predicted, and residual income inequality, for males between the ages of 18 and 55, into inequality across municipalities and inequality within municipalities, by country. This file produces the country-by-country results reported in Table 2, and the output is used as an input by `aej_inequality.m`.
9. `aej_inequality.m` - Decomposes total income inequality (all individuals), total income inequality (males 18-55), predicted income inequality (males 18-55), and residual income inequality (males 18-55) into across-country, across-municipality, and within-municipality components. Produces the results reported in Table 1 and the summary rows of Table 2.

3 Roads

To reproduce the municipality-level roads measures, follow these steps:

1. Create the GIS boundary files, as described in the directions above for calculating the population averages.
2. Obtain the GIS roads data.
 - (a) The primary source for these data is the International Center for Tropical Agriculture (CIAT), *Latin America and the Caribbean Roads Database*. CIAT can be contacted as follows:

International Center for Tropical Agriculture
A.A. 6713; Cali, Colombia
Phone: +1 (650) 8336625
Fax: +1 (650) 8336626
Email: ciat@cgiar.org.

The roads database is not for sale, but CIAT may choose, at their discretion, to distribute it to researchers after conversing about the intended use of the data.

- (b) The CIAT data for Peru were somewhat outdated and hence were updated using a 2006 GIS roads database produced by the Peruvian Ministry of Transportation. At the time of writing this document, the Ministry of Transportation did not have a functioning Web site, and the only way we know to obtain these proprietary data is to go in person or to call the Ministry of Transportation in Lima (address: Jiron Zorritos 1203, Lima, Peru; phone: 51-1-6157800). Note that this update to the Peru roads data makes only a very marginal difference for the results.
 - (c) Finally, the roads data for the United States and their corresponding documentation can be downloaded free of charge at <http://www-atlas.usgs.gov/metadata/roadtrl020.faq.html>.
3. Run the python file `aej_roads.py`. You must have Python 2.4, ArcGIS 9.2, and a license for the ArcGIS spatial analysis extension on your machine to run this file. Before running the file, you must modify the directory path at the top of the file, where it states “INSERT DIRECTORY PATH HERE”. Note that Python requires directory paths to be specified with double slashes. You may also need to modify the location of the Arc toolboxes on your machine (under the heading “# Load required toolboxes...” at the top of the file). This will produce a series of .dbf files. The output from steps (1) through (3) is provided in the files posted on the AEA Web site (`dpAR.dbf`, `dpBL.dbf`, `dpBR.dbf`, `dpCI.dbf`, `dpCO.dbf`, `dpEC.dbf`, `dpES.dbf`, `dpGT.dbf`, `dpHO.dbf`, `dpMX.dbf`, `dpNU.dbf`, `dpPA.dbf`, `dpPM.dbf`, `dpUS.dbf`, `dpUY.dbf`, `dpVE.dbf`, `BRcensus3.csv`, `UScensus3.csv`, and `VEcensus3.csv`). These can be placed in the desired directory to complete step (4) without doing steps (1) through (3) (note that step 4 requires the income data discussed in Section 2).
 4. Run the Stata file `aej_roads.do`, in order to reproduce the results in Table 3. Note that you must modify the directory path at the top of the file, where it states “INSERT DIRECTORY PATH HERE”. The GIS and income .dta files produced by following the directions in Steps 1 and 2 must be in this directory for `aej_roads.do` to run.