READ ME

This directory contains the data and code to replicate the results in

"Trade and Inequality: From Theory to Estimation" (Elhanan Helpman, Oleg Itskhoki, Marc Muendler and Stephen Redding), May 2016

The paper uses confidential employer-employee data from RAIS, Brazil. We were given access to these data under the agreement that they could not be made publicly available. However, we interpret this agreement to permit third-party researchers access to the data on a secure server at UC San Diego for replication and extensions, in addition to the general raw data access at IPEA Brasilia and IBGE Rio de Janeiro. Several visiting scholars at UC San Diego have made use of RAIS in recent years. We will make available our code and hold the final estimation data ready for replication purposes.

Note that some of the estimation is computationally demanding and some of the files will take a number of hours to run even on a fast computer.

The files should be run in the order specified in Sections 2)-7) of this document.

To run the Stata and Matlab code, change the local directory assignment specified at the top of the files listed below (e.g. user=1).

1) SUMMARY

This table summarizes the files that replicate the figures and tables in the paper:

Table 1	dofiles/sector-occupation/soanalysis.do Table 1, Column (1), Rows (1)-(5) Table 1, Column (2), Rows (1)-(5)
Table 1	dofiles/sector-occupation/soanalysis_cnae.do Table 1, Column (1), Row (6) Table 1, Column (2), Row (6)
Table 1	dofiles/sector-occupation/csoanalysis.do Table 1, Column (1), Rows (7)-(8) Table 1, Column (2), Rows (7)-(8)
Table 2	dofiles/uncond/anova/analysis.do Table 2, Left Panel (Unconditional)
Table 2	dofiles/condwtvfe/anova/tanalysis.do Table 2, Right Panel (Conditional)
Table 3	dofiles/condwtvfe/wagesize/expwagesize.do
Table 4	matlab/brazil/main_MLE.m
Table 5	matlab/brazil/main_MLE.m
Table 6	matlab/brazil/main_MLE.m
Table 7	matlab/brazil/main_MLE.m
Table 8	dofiles/condwtvfe/mincer/semi-parametric-iv.do
Figure 1	matlab/brazil/COUNTERFACTUAL.m

Figure 2	matlab/brazil/GMMBounds2.m
Figure 3	matlab/brazil/COUNTERFACTUALmult.m
Table A1	dofiles/prepdata/descrip.do
Table A2	dofiles/prepdata/descrip.do dofiles/uncond/mincer/descripexporter.do
Table A3	matlab/brazil/COUNTERFACTUALmult.m
Figure A1	matlab/brazil/main_MLE.m
Figure A2	matlab/brazil/main_MLE.m
Figure A3	matlab/brazil/GMMBounds2.m
Figure A4	matlab/brazil/COUNTERFACTUALovertime.m matlab/brazil/COUNTERFACTUALmultovertime.m

This table summarizes the files that replicate miscellaneous results in the paper:

Footnote 10	dofiles/sector-occupation/soanalysis_cnae.do
Footnote 13	dofiles/condwtvfe/tanova/tanalysis_firm.do
Section 3.3	Observable and residual wage inequality within sector-occupations without including time-varying firm fixed effect dofiles/condwtvfe/tanova/tanalysis_robust.do
Footnote 15	dofiles/uncond/anova/analysis.do dofiles/conwtvfe/tanova/tanalysis.do

This table summarizes the files that replicate the figures and tables in the online supplement:

Table H1	Agriculture and mining robustness test dofiles/sector-occupation/soanalysis_agmining.do dofiles/sector-occupation/soanalysis_mining.do
Table H2	Regional robustness test for Sao Paulo state *Overall wage inequality for Sao Paulo* dofiles/sector-occupation/soanalysis_sao.do *Residual wage inequality for Sao Paulo* dofiles/sector-occupation/csoanalysis_sao.do *Overall wage inequality for sector-occupation-regions* dofiles/sector-occupation/soanalysis_sor.do *Residual wage inequality for sector-occupation-regions* dofiles/sector-occupation/csoanalysis_sor.do
Table H3	Coefficients on worker observables in Mincer regression dofiles/sector-occupation/csoanalysis.do data/sector-occupation/cso-mincer-coeff-results-year.csv data/sector-occupation/cso-mincer-se-results-year.csv
Table H4	Coefficients from Mincer regression using more disaggregated education categories dofiles/sector-occupation/csoanalysis_educ.do cso-mincer-educ-rob-coeff-results-year.csv cso-mincer-educ-rob-se-results-year.csv
Table H5	Within-between decomposition by occupation dofiles/uncond/anova/analysis_occup.do

	dofiles/condwtvfe/tanova/tanalysis_occup.do
Table H6	Within-between decomposition by sector dofiles/uncond/anova/analysis_sector.do dofiles/condwtvfe/tanova/tanalysis_sector.do
Table H7	Within-between decomposition by sector tanova/tanalysis_sector.do
Table H8	Column (1) (Unconditional) dofiles/uncond/anova/analysis.do Column (2) (Conditional) dofiles/condwtvfe/tanova/tanalysis.do
Table H9	Constant composition wage inequality dofiles/cond/mincer/cmincer_compos_anal.do
Table H10	Firm and worker fixed effects variance decomposition dofiles/condwtvfe/tanova/tanalysis-ack.do tanalysis-ack-num.csv tanalysis-ack.csv
Table H11	Matlab Colombian robustness test matlab/colombia/matlab/mainMLE.m (estimation) matlab/colombia/matlab/VML_c.m (standard errors)
Table H12	Columns (1)-(2) Table A1 Columns (3)-(4) dofiles/pnad/pnad-descrip-comcart.do Columns (5)-(6) dofiles/pnad/pnad-descrip.do
Table H13	Columns (1)-(2) Table A2 Columns (3)-(4) dofiles/pnad/pnad-descrip-comcart.do Columns (5)-(6) dofiles/pnad/pnad-descrip.do
Table H14	Column (1): Table 1, column (1), rows (1)-(6) Column (2), Rows (1)-(5) soanalysis-comcart.do Column (2), Row (6) soanalysis-comcart_ramo.do Column (3), Rows (1)-(5) soanalysis.do Column (3), Row (6) soanalysis_ramo.do
Table H15	Column (1): Table 1, column (1), rows (7)-(8) Column (2), Rows (7)-(8) csoanalysis-comcart.do Column (3), Rows (7)-(8) csoanalysis.do
Table I1	Detailed occupation breakdown
Table I2	Detailed industry breakdown
Figure D1	Monte Carlo results matlab/montecarlo/montecarlo_final.m
Figure H1	umincer/descripexporter.do
Figure H2	umincer/descripexporter.do
Figure H3	dofiles/sector-occupation/soanalysis.do
Figure H4	dofiles/sector-occupation/csoanalysis.do
Figure H5	dofiles/sector-occupation/ csoanalysis.do
Figure H6	dofiles/sector-occupation/csoanalysis.do
Figure H7	Sector-year fixed effects from Mincer regression dofiles/sector-occupation/csoanalysis_regioncontrol.do
Figure H8	dofiles/condwtvfe/tanova/tanalysis.do
Figure H9	Constant composition wage inequality

	dofiles/cond/mincer/cmincer_compos_anal.do
Figure H10	Matlab counterfactuals robustness test
Figure H11	Columbia firm distribution fit matlab/colombia/matlab/main_moments.m
Figure H12	Colombia worker distribution fit matlab/colombia/matlab/main_moments.m
Figure H13	Colombia counterfactuals matlab/colombia/matlab/main_counterfactual_fixed.m matlab/colombia/matlab/main_counterfactual_var.m

This table summarizes the files that replicate miscellaneous results in the online supplement

Section G	Summary statistics in text dofiles/apdx/apdx-stats.do dofiles/apdx/apdx-stats-samplecond.do
Section H7	Results discussed in text dofiles/sector-occupation/csoanalysis_educ.do
Section H13	Results discussed in text dofiles/condwtvfe/tanova/tanalysis_firm.do
Section I	Tabulations dofiles/apdx/sec-occ-samplecond.do

2) TRADE AND INEQUALITY DIRECTORY STRUCTURE

The directory structure on the COMPADV server is as follows:

/u/data/brazil/himr/code	Stata and Matlab code to replicate the results in the paper starting with the raw input datasets in /u/data/brazil/himr/input
/u/data/brazil/himr/input	Directory containing the raw input datasets
/u/data/brazil/himr/origins	Directory documenting the origin of the raw input datasets (the Stata files in this directory are slow to run and we do **not** recommend re-running them)
/u/data/brazil/himr/restud	Directory containing intermediate datasets that are generated by the code in /u/data/brazil/himr/code

The directory /u/data/brazil/himr/code contains the following subdirectories:

himr/code/ado	Contains Stata ado files used in some of the Stata estimation
himr/code/stata	Contains the Stata code to replicate the results in the paper
himr/code/matlab	Contains the Matlab code to replicate the results in the paper

3) HIMR/ORIGINS

Here we describe the origin of the research data from the raw RAIS, SECEX and PNAD data as well as from auxiliary data (such as on doing business, or country or sector concordances). To save on replication time, we do not recommend to regenerate the research data from their raw sources. However, we document here the code that generates the Stata input files in ../himr/input. For each himr/input data set in the first column in the table below, the second column shows the Stata programs in reverse chronological order, where the last listed file processes the data from the original database (which is in turn explained in the respective original data folder whenever a database file is not in the raw format).

exp-cnpjradic-destbins.dta	secex-funcex-destbins.do, secex-funcex.do, firmagg-secex.do, secex-exp-prod.do
exp-dest10pls-cnpjradic.dta	secex-funcex.do, firmagg-secex.do, secex-exp-prod.do
expimp-cnpjradic.dta	secex-expimp.do
exp-usa-cnpjradic.dta	secex-funcex.do, firmagg-secex.do, secex-exp-prod.do
firms-`year'-hs6-dobusi.dta (`year' in 1986/1998)	f-shifter-hs6-dobusi.do, doing-business-2.do, wtf-brazil-iso-subsibge.do, wtf-brazil.do
municip2meso.dta	municip2meso.do
municnpj.csv	muni-to-plants-cnpj.sas, muni-to-plants-cnpj.pl
rais-secex-all-workers-`year'.dta (`year' in 1986/1998)	prep-workers-annual.do
workers-template-`year'-educ.dta (`year' in 1986/1998)	prep-workers-annual-educ.do, extract-edu.sas, extract-edu.pl
workers-template-`year'-nacl.dta (`year' in 1986/1998)	prep-workers-annual-natlty.do, extract-natlty.sas, extract-natlty.pl
pnad`year'pes_comp81_likerais.dta (`year' in 1986(1)1990 1992(1)1993 1995(1)1998)	pnad-likerais.do
meso-mass-layoff-cnae.dta	mass-layoff-foreigners-cnae.do

For additional information, the file ./himr/origins/sequence-of-programs .txt lists all required programs in their sequence of use as well as the input and output files for each program step. All the input and output files continue to be held on the server himr.rc.fas.harvard.edu in the stated original locations. The subdirectories /perl/, /sas/, /stata/ contain copies of all relevant Perl scripts, SAS executables and Stata do files for convenience (while the original program files remain in the according data preparation folders on himr.rc.fas.harvard.edu).

4) HIMR/CODE/STATA

The Stata code to replicate the results in the paper is organized around the following folders.

4.1) himr/code/stata/dofiles/routines

This directory contains subroutines that are run from other Stata files discussed in later subsections of this document.

samplecond.do	Constructs our baseline sample in the manufacturing sector
samplecond_mining.do	Constructs a robustness sample including manufacturing and mining
samplecond_agmining.do	Constructs a robustness sample including manufacturing, mining and agriculture

4.2) himr/code/stata/dofiles/prepdata

First run the following data preparation do files:

prepstatecodes.do	Prepare the state codes for regional analysis using the following input datasets ./data/himr/input/municip2meso.dta ./data/himr/input/municnpj.csv
cleannationality.do	Cleans up data on nationalities of workers using the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/rais-secex-all-workers-`yr'.dta ./data/himr/input /workers-template-`yr'-nacl.dta
cleanfuncex.do	Cleans up trade transactions data using the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/expimp-cnpjradic.dta ./data/himr/input/exp-usa-cnpjradic.dta ./data/himr/input/exp-dest10pls-cnpjradic.dta ./data/himr/input/exp-cnpjradic-destbins.dta
cleantradeinst.do	Clean trade instrumental variables using the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/firms-`yr'-hs6-dobusi.dta
dispemp.do	Creates an occupation-industry-year employment dataset and measures of the dispersion of employment using the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/rais-secex-all-workers-`yr'.dta
empweights.do	Creates employment share weights for summarizing employment-weighted averages of results
descrip.do	Creates descriptive tables of occupation and industry characteristics

4.3) himr/code/stata/dofiles/sector-occupation

Contains the following do files for decomposing overall wage inequality into its within and between sector-occupation components:

soanova.do	Constructs datasets for decomposing overall wage inequality into within and between sector-occupation components using the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/rais-secex-all-workers-`yr'.dta
soanalysis.do	Generates the results of the decomposition of overall wage inequality into within and between sector-occupation components using the datasets created by soanova.do
soanova_agmining.do	Constructs datasets for robustness test decomposing overall wage inequality into within and between sector-occupation components including agriculture and mining industries. This file uses the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/rais-secex-all-workers-`yr'.dta
soanalysis_agmining.do	Generates the results of the decomposition of overall wage inequality into within and between sector-occupation components using the datasets created by soanova_mining.do
soanova_mining.do	Constructs datasets for robustness test decomposing overall wage inequality into within and between sector-occupation components including mining industries. This file uses the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/rais-secex-all-workers-`yr'.dta
soanalysis_mining.do	Generates the results of the decomposition of overall wage inequality into within and between sector-occupation components using the datasets created by soanova_mining.do
soanova_cnae.do	Constructs datasets for the robustness test decomposing overall wage inequality into within and between sector-occupation components for more disaggregated (CNAE) industries using the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/rais-secex-all-workers-`yr'.dta
soanalysis_cnae.do	Generates the results of the robustness test decomposing residual wage inequality into within and between sector-occupation components for more disaggregated (CNAE) industries using the datasets created by soanova_cnae.do (Footnote 10 in the paper)
soanova_sao.do	Constructs datasets for the robustness test decomposing overall wage inequality into within and between sector-occupation components for Sao Paulo state using the following input datasets, where `yr' is a local with the numeric value of the year

	./data/himr/input/rais-secex-all-workers-`yr'.dta
soanalysis_sao.do	Generates the results of the robustness test decomposing overall wage inequality into within and between sector-occupation components using the datasets created by soanova_sao.do
soanova_sor.do	Constructs datasets for the robustness test decomposing overall wage inequality into within and between sector-occupation-region components using the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/rais-secex-all-workers-`yr'.dta
soanalysis_sor.do	Generates the results of the robustness test decomposing overall wage inequality into within and between sector-occupation-region components using the datasets created by soanova_sor.do (Table H2 in the online supplement)
csoanova.do	Constructs datasets for decomposing residual wage inequality into within and between sector-occupation components using the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/rais-secex-all-workers-`yr'.dta
csoanalysis.do	Generates the results of the decomposition of residual wage inequality into within and between sector-occupation components using the datasets created by csoanova.do (Table 1 in the paper)
csoanova_regioncontrol.do	Constructs datasets for the robustness test decomposing overall wage inequality in within and between sector-occupation components after controlling for worker observables, sector and region
csoanalysis_regioncontrol.do	Reports the results of the robustness test decomposing overall wage inequality in within and between sector-occupation components after controlling for worker observables, sector and region
csoanova_educ.do	Constructs datasets for the robustness test decomposing residual wage inequality into within and between sector-occupation components for more disaggregated education categories using the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/rais-secex-all-workers-`yr'.dta
csoanalysis_educ.do	Generates the results of the robustness test decomposing residual wage inequality into within and between sector-occupation components for more disaggregated education categories using the datasets created by csoanova_educ.do (Table H4 in the online supplement)
csoanova_sao.do	Constructs datasets for the robustness test decomposing residual wage inequality into within and between sector-occupation components for Sao Paulo state using the following

	input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/rais-secex-all-workers-`yr'.dta
csoanalysis_sao.do	Generates the results of the robustness test decomposing residual wage inequality into within and between sector-occupation components using the datasets created by csoanova_sao.do
csoanova_sor.do	Constructs datasets for the robustness test decomposing residual wage inequality into within and between sector-occupation-region components using the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/rais-secex-all-workers-`yr'.dta
csoanalysis_sor.do	Generates the results of the robustness test decomposing residual wage inequality into within and between sector-occupation-region components using the datasets created by csoanova_sor.do

4.3) himr/code/stata/dofiles/uncond

Contains the following do files for unconditional wage inequality analysis:

anova/anova.do	Constructs dataset for decomposition of wage inequality within sector-occupations into within and between firm components using the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/rais-secex-all-workers-`yr'.dta
anova/analysis.do	Reports results of decomposition of wage inequality within sector-occupations into within and between firm components using the datasets created by anova/anova.do
anova/analysis_sector.do	Robustness test reporting results of decomposition of wage inequality within sector-occupations into within and between firm components by sector using the datasets created by anova/anova.do
anova/analysis_occup.do	Robustness test reporting results of decomposition of wage inequality within sector-occupations into within and between firm components by occupation using the datasets created by anova/anova.do
mincer/umincer.do	Estimates mean wages by sector-occupation-year using the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/rais-secex-all-workers-`yr'.dta
mincer/descripexporter.do	Descriptive figures of export participation and real exports over time

4.4) himr/code/stata/dofiles/condwtvfe

Contains the following do files for wage inequality analysis conditional on time-varying firm fixed effects:

tanova/tanova.do	Constructs datasets for decomposition wage inequality within sector-occupations into within and between firm components after controlling for worker observables using the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/rais-secex-all-workers-`yr'.dta
tanova/tanalysis.do	Reports results for decomposition wage inequality within sector-occupations into within and between firm components after controlling for worker observables using the datasets created by tanova.do
tanova/tanalysis_sector.do	Reports results by sector for decomposition wage inequality within sector-occupations into within and between firm components after controlling for worker observables using the datasets created by tanova.do
tanova/tanalysis_occup.do	Reports results by occupation for decomposition wage inequality within sector-occupations into within and between firm components after controlling for worker observables using the datasets created by tanova.do
tanova/tanova_educ.do	Generates datasets for robustness test decomposing wage inequality within sector-occupations into within and between firm components after controlling for worker observables (using more disaggregated education measures) and using the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/rais-secex-all-workers-`yr'.dta ./data/himr/input/workers-template-`yr'-educ.dta
tanova/tanalysis_educ.do	Reports results for robustness test decomposing wage inequality within sector-occupations into within and between firm components after controlling for worker observables (using more disaggregated education measures) using the datasets generated by tanova/tanova_educ.do
tanova/tanova_firm.do	Generates datasets for robustness test decomposing wage inequality within sectors into within and between firm components after controlling for worker observables (using time-varying firm fixed effects instead of time-varying firm-occupation fixed effects) using the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/rais-secex-all-workers-`yr'.dta
tanova/tanalysis_firm.do	Results of robustness test decomposing wage inequality within sectors into within and between firm components after controlling for worker observables (using time-varying firm fixed effects instead of time-varying firm-occupation fixed

	effects)
tanova/tanova_raw.do tanova/tanova_ack.do	Robustness test decomposing wage inequality within sectors into within and between firm components using the Abowd, Creecy and Kramarz methodology (including firm and worker fixed effects) using the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/rais-secex-all-workers-`yr'.dta
tanova/tanova_robust.do	Generates datasets for robustness test decomposing wage inequality within sector-occupation-years, excluding firm fixed effects, and using the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/rais-secex-all-workers-`yr'.dta
tanova/tanalysis_robust.do	Reports results of robustness test decomposing wage inequality within sector-occupation-years, excluding firm fixed effects, and using the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/rais-secex-all-workers-`yr'.dta
mincer/tmincer.do	Estimates firm-occupation-year wage components in a Mincer wage regression using the following input datasets, where `yr' is a local with the numeric value of the year ./data/himr/input/rais-secex-all-workers-`yr'.dta
mincer/aggtmincer.do	Aggregates firm-occupation-year wage components generated by mincer/tmincer.do to the firm-year level for Matlab estimation of the model
mincer/semi-parametric-iv.do	Semi-parametric selection model estimation
wagesize/expwagesize.do	Estimates wage premia by employment size and export status for Table 3 in the paper

4.5) himr/code/stata/dofiles/cond

Contains the following do files for wage inequality analysis conditional on worker observables:

cond/mincer/cmincer_compos.do	Generates datasets for robustness test for constant composition residual wage inequality analysis following Lemieux (2006)
cond/mincer/cmincer_compos_anal.do	Reports results for robustness test for constant composition residual wage inequality analysis following Lemieux (2006)

4.6) himr/code/stata/dofiles/pnad

Contains the following do files for decomposing overall wage inequality into its within and between sector-occupation components in the PNAD household data:

pnad/pnad-descrip.do	Creates descriptive tables of occupation and industry characteristics for any PNAD household member with work (formal or informal employment). Tables H12

	and H13.
pnad/pnad-descrip-comcart.do	Creates descriptive tables of occupation and industry characteristics only for PNAD household members with work and "*Carteira*" (formal employment). Tables H12 and H13.
pnad/pnad-soanova.do	Constructs datasets for decomposing overall wage inequality into within and between sector-occupation components using the following input datasets, where `year' is a local with the numeric value of the year ./data/himr/input/pnad`year'pes_comp81_likerais.dta for any PNAD household member with work (formal or informal employment)
pnad/pnad-soanova-comcart.do	Constructs datasets for decomposing overall wage inequality into within and between sector-occupation components using the following input datasets, where `year' is a local with the numeric value of the year ./data/himr/input/pnad`year'pes_comp81_likerais.dta only for PNAD household members with work and "*Carteira*" (formal employment)
pnad/pnad-soanalysis.do	Generates the results of the decomposition of overall wage inequality into within and between sector-occupation components using the datasets created by pnad-soanova.do for any PNAD household member with work (formal or informal employment). Table H14.
pnad/pnad-soanalysis-comcart.do	Generates the results of the decomposition of overall wage inequality into within and between sector-occupation components using the datasets created by pnad-soanova-comcart.do only for PNAD household members with work and "*Carteira*" (formal employment). Table H14.
pnad/pnad-soanova_ramo.do	Constructs datasets for the robustness test decomposing overall wage inequality into within and between sector-occupation components for more disaggregated (RAMO) activities using the following input datasets, where `year' is a local with the numeric value of the year ./data/himr/input/pnad`year'pes_comp81_likerais.dta for any PNAD household member with work (formal or informal employment)
pnad/pnad-soanova-comcart_ramo.do	Constructs datasets for the robustness test decomposing overall wage inequality into within and between sector-occupation components for more disaggregated (RAMO) activities using the following input datasets, where `year' is a local with the numeric value of the year ./data/himr/input/pnad`year'pes_comp81_likerais.dta

	only for PNAD household members with work and "*Carteira*" (formal employment)
pnad/pnad-soanalysis_ramo.do	Generates the results of the robustness test decomposing residual wage inequality into within and between sector-occupation components for more disaggregated (RAMO) activities using the datasets created by pnad-soanova_ramo.do for any PNAD household member with work (formal or informal employment). Table H14.
pnad/pnad-soanalysis-comcart_ramo.do	Generates the results of the robustness test decomposing residual wage inequality into within and between sector-occupation components for more disaggregated (RAMO) activities using the datasets created by pnad-soanova-comcart_ramo.do only for PNAD household members with work and "*Carteira*" (formal employment). Table H14.
pnad/pnad-csoanova.do	Constructs datasets for decomposing residual wage inequality into within and between sector-occupation components using the following input datasets, where `year' is a local with the numeric value of the year ./data/himr/input/pnad`year'pes_comp81_likerais.dta for any PNAD household member with work (formal or informal employment)
pnad/pnad-csoanova-comcart.do	Constructs datasets for decomposing residual wage inequality into within and between sector-occupation components using the following input datasets, where `year' is a local with the numeric value of the year ./data/himr/input/pnad`year'pes_comp81_likerais.dta only for PNAD household members with work and "*Carteira*" (formal employment)
pnad/pnad-csoanalysis.do	Generates the results of the decomposition of residual wage inequality into within and between sector-occupation components using the datasets created by pnad-soanova.do for any PNAD household member with work (formal or informal employment). Table H15.
pnad/pnad-csoanalysis-comcart.do	Generates the results of the decomposition of residual wage inequality into within and between sector-occupation components using the datasets created by pnad-soanova-comcart.do only for PNAD household members with work and "*Carteira*" (formal employment). Table H15.

4.7) himr/code/stata/dofiles/apdx

Contains the following do files for descriptive statistics in the Data Appendices G and I:

apdx/apdx-stats.do	Generates descriptive statistics on the unrestricted RAIS data, presented in Data Appendix G.
apdx/sec-occ.do	Provides tabulations of CNAE industries and CBO occupations based on the unrestricted RAIS data. (Robustness, not reported.)
apdx/apdx-stats-samplecond.do	Generates descriptive statistics on the restricted RAIS sample (firms whose workers are reported with complete demographics), presented in Data Appendix G.
apdx/sec-occ-samplecond.do	Provides tabulations of CNAE industries and CBO occupations based on the restricted RAIS sample (firms whose workers are reported with complete demographics) presented in Data Appendix I.

5) HIMR/CODE/MATLAB/BRAZIL

This folder implements the structural estimation in Sections 5 and 6 of the paper

main_MLE.m	Main Matlab code that generates the maximum likelihood estimation results in Section 5.1 of the paper using the following files created by the Stata code above: agg-tmincer`yr'-fun.csv
VML_c.m	Computes standard errors for the baseline maximum likelihood estimation in Section 5.1 of the paper and generates Figure A1 of the parameter estimates and standard errors over time using the parameter estimates generated by main_MLE.m
main_moments.m	Computes the firm and worker moments in the model and data and generates Figure A2 with the kernel density estimates of the distributions of wages and employment in the model and data
COUNTERFACTUAL.m	Undertakes the baseline maximum likelihood counterfactuals in Section 5.2 of the paper using the parameter estimates generated by main_MLE.m
mainMLEmultdest.m	Undertakes the maximum likelihood estimation for the multi-destination specification in Section 6 of the paper using the following files created by the Stata code above: agg-tmincer`yr'-fun.csv
COUNTERFACTUALmult.m	Undertakes the counterfactuals for the multi-destination specification in Section 6 of the paper using the parameter estimates generated by mainMLEmultdest.m
COUNTERFACTUALovertime.m	Undertakes the counterfactuals over time for Figure A4 in the appendix to the paper
ROBUSTNESSovertime.m	Undertakes the counterfactuals for each year for Figure H10 of the online supplement

GMMBounds.m	Undertakes the GMM Bounds estimation and generates Figure A3 in the appendix to the paper
GMMBoundsCounterfactualAutarky.m	Undertakes the GMM Bounds autarky counterfactuals for Figure 2 in the paper
GMMBoundsCounterfactualLocal.m	Undertakes the GMM Bounds local changes in variable trade costs counterfactuals for Figure 2 in the paper
llf.m	Function file with the likelihood function
llf_c.m	Function file with the constrained likelihood function
strconstr.m	Function file with the structural constraint
sj_c.m	Function file used to compute the standard errors
ssj_c.m	Function file used to compute the standard errors
llf_md.m	Function file with the likelihood function for the multi-destination specification
strconstr_md.m	Function file with the structural constraint for the multi-destination specification

6) HIMR/CODE/MATLAB/MONTECARLO

This folder implements the Monte Carlo in Section D.5 of the Online Supplement

montecarlo/montecarlo_final.m	Matlab code to implement the Monte Carlo
montecarlo/llf.m	Function file with likelihood function called by montecarlo_final.m
montecarlo/strconstr.m	Function file with structural constraint called by montecarlo_final.m

7) HIMR/CODE/MATLAB/COLOMBIA

This folder implements the robustness test using Colombian data from Section H17 of the Online Supplement. The Colombia data were supplied by Jim Tybout from Penn State University. Please contact Jim Tybout directly to request access to these data.

This Stata file sets up the Colombian data for the Matlab estimation.

colombia/stata/colmerge.do	Stata file to merge and prepare the Colombian data for Matlab estimation

These Matlab files implement the structural estimation using the Colombian data.

colombia/matlab/mainMLE.m	Matlab file to estimate the model's parameters
colombia/matlab/llf.m	Function file with likelihood function called by mainMLE.m
colombia/matlab/strconst.m	Function file with structural constraint called by mainMLE.m
colombia/matlab/VML_c.m	Matlab file to compute sandwich standard errors for the model's parameters
colombia/matlab/llj_c.m	Function file called by VML_c.m

colombia/matlab/sj_c.m	Function file called by sj_c.m
colombia/matlab/ssj_c.m	Function file called by ssj_c.m
colombia/matlab/ main_counterfactual_fixed.m	Matlab file to undertake fixed cost counterfactuals
colombia/matlab/ main_counterfactual_variable.m	Matlab file to undertake variable cost counterfactuals
colombia/matlab/ main_moments.m	Matlab file to compare the model's fit to the data