

```
*****
/* readme file for:
                               */
/* "Corporate Taxation Under Weak Enforcement" */
/* Authors: Pierre Bachas and Mauricio Soto */
/* First Edited: January 2014 */
/* Last Edited: November 2020 */
*****
```

** This Dofile summarizes the code and data used for the manuscript "Corporate Taxation Under Weak Enforcement" accepted at the AEJ Policy

** The manuscript uses three types of datasets explained below:

** (1) The main data used in the analysis is Costa Rican administrative firm level tax returns, accessed through the Ministry of Finance (MoF). The data are physically stored on computers at the Ministerio de Hacienda, in San Jose, Costa Rica, and due to security reasons the data may not be transferred to computers outside the Finance Ministry. Researchers interested in obtaining access to the data employed in this paper can contact the Ministerio de Hacienda, Departamento de Estadisticas Fiscales, Division de Politica Fiscal. These data are not meant to become public and will remain restricted-access. However, aggregated/tabulated data which was provided to us by the MoF and which groups firms by bins of revenue is provided, which enables the replication of most figures and tables in the paper. These tabulated data are directly provided to us, upon description of the bins of revenue and thus do not have a source file. The Ministry of Finance does not keep archival snapshots of its database.

** (2) The other information used in the appendix includes administrative register data from the Central bank of Costa Rica. The dataset is titled "Registro de variables economicas del Banco Central de Costa Rica (Revec)" and contains firm-level tax return data, combined with information on employment from social security, and data on economic affiliates. The data are physically stored on computers at the Banco Central, in San Jose, Costa Rica, and due to security reasons the data may not be transferred to computers outside the Central Bank. Researchers interested in obtaining access to the data employed in this paper can contact the Banco Central de Costa Rica, Division Economica, Departamento de Investigacion Economica. These data are not meant to become public and will remain restricted-access. Aggregated/tabulated data which was provided to us by the MoF and which groups firms by bins of revenue is provided, which enables the replication of most figures and tables in the paper. These tabulated data are directly provided to us, upon description of the bins of revenue and needs and thus do not have a source file. The Central Bank does not keep archival snapshots of its database.

** (3) Figure 1 contains public access data and the sources are detailed below.

```
*****  
* PRELIMINARIES
```

```
*****  
** (1) Run: config_stata.do --> wrapper with ssc instal packages  
** Installs packages: grc1leg, expand, estout, outreg, outtable,  
frmttable
```

```
** (2) Create in the working directory three new folders:  
** proc (store all processed data), graphs (stores figures), tables  
(stores tables)
```

```
*****  
* AEJ:EP Paper Code
```

```
*****  
*****  
* Figures
```

```
*****  
** Figure 1: Corporate Income Taxes across Countries  
** Corporate_tax_cross_country.do  
** Inputs: "ICTDGRD_Dec2015_CentralGeneralMergedFull.xlsx" ,  
"Statutory_data.dta" , "GDP_data_CITCostaRica.xlsx"  
** Description: country level data on corporate tax revenues and on  
top statutory tax rates, merged with countries GDP and population.  
Corporate tax revenue data obtained from the ICTD/UNU-WIDER  
"Government Revenue Dataset" which can be accessed at https://  
www.wider.unu.edu/project/government-revenue-dataset. Top statutory  
tax rates were collected by the authors, by online search looking at  
the websites of the ministries of finance of each country. Data on  
country's per capita GDP and population were obtained from the World  
Bank development indicators on November 11th 2020: exact data query  
can be accessed at https://databank.worldbank.org/  
GDP\_data\_CITCostaRica/id/f7611cd8 . As can be seen from the query, the  
dataset collects data on per capita GDP in PPP, current international  
USD, and on population from all countries available over all years.  
The code generating figure 1, then keeps the latest year available for  
each country, and restricts the sample to countries with more than 1  
million inhabitants.  
** Output: Figure 1 in manuscript  
  
** Figure 2: Costa Rica corporate tax schedule  
** Not applicable  
  
** Figure 3: Theoretical Bunching
```

** Not applicable

** Figure 4: Firm Density and Average Profit Margin

** revenue_profitmargin.do

** Inputs: Margin.xlsx, sheet 1 : Tabulated data on number of firms and average profit by 1/2 Mil CRC bins of revenue (relative to first and to second threshold)

** Output: Figure 4 in manuscript, Figure B1 in Online Appendix, Figure B3 in Online Appendix

** Figure 5: Revenue Bunching Estimation

** revenue_bunching.do

** Inputs: Margin.xlsx, sheet 1 : Tabulated data on number of firms and average profit by 1/2 Mil CRC bins of revenue

** Output: Figure 5 in manuscript

** Figure 6: Donut-Hole Discontinuity in Cost by Revenue

** cost_discontinuity.do

** Inputs: Margin.xlsx, sheet 1 : Tabulated data on number of firms and average profit by 1/2 Mil CRC bins of revenue

** Output: Figure 6 in manuscript

** Figure 7: Profit Margin Change Across Years for Growing Firms

** dynamics.do

** Inputs: Margin.xlsx, sheets 3,4,5 : Tabulated data of transition matrices at year t and year t+1 by 2M bins of revenue.

** Rows: bin relative to threshold at time t. Columns: bin relative to threshold at time t+1

** Output: Figure 7 in manuscript, Appendix Figures B6, B7

** Figure 8: Profit Margin Distributions Away From the Threshold

** profit_distributions.do

** Inputs:"Import_profit_margin_distributions.txt : tabulations of the percentiles of profit margin distribution within each 2 Mil CRC revenue interval

** Output: Figure 4 in manuscript.

** Figure 9: Notifications of Discrepancies with Third-Party Data

** corrections.do

** Inputs: "Sales_and_Corrections_2M.dta" : Tabulated data on number of firms and number of desk audits by 2M bins of revenue for 2012

** Output: Figure 9 in manuscript

* Tables

** Table 1: Elasticity Estimates Point of Convergence Method

** revenue_elasticity_estimates.do

** Inputs: Outcome of bunching estimation

** s.e. from bootstrap_residuals_T1 and bootstrap_residuals_T2
** Output: Table 1 in manuscript (Note: cost elasticity estimates
come from file below.)

** Table 2: Donut-Hole Cost Discontinuity by Revenue
** cost_discontinuity.do
** Inputs: Margin.xlsx : Tabulated data on number of firms and
average profit by 1/2 bins of revenue around 1st threshold
** Output: Table 2 in manuscript

** Table 3: Industry Level Results (First Threshold)
** sector_analysis.do
** Inputs: sector_data.dta : Tabulated data on number of firms and
average profit by 1/2 bins of revenue by economic sector
** Output: Table 3 in manuscript. Appendix Figure B4

** Table 4: Variation in Audit Risk at the Sector Level
** Audits_DDD.do
** Inputs: Directly run on the micro tax return data held on the
server, can not be disclosed.
** Output: Figure 4 in manuscript

* ADDITIONAL FIGURES AND TABLES IN ONLINE APPENDIX

* Figures

* Figure B1: See above with Figure 4

* Figure B2:
* Profit_elasticity_robustness.do
* profit_elasticity_T1.dta , profit_elasticity_T2.dta --> generated
in bootstrap_residuals_T1 & bootstrap_residuals_T2.

* Figure B3: See above with Figure 4

* Figure B4: see above with Table 3

* Figure B5:
* yearly_figures.do
* Inputs: Yearly_data.xlsx : tabulated data at the yearly level, for
the number of firms per revenue bin and the average profit margin,
covers 2008-2014

* Figure B6: see above with Figure 7

* Figure B7: see above with Figure 7

```

* Figure B8:
* Analysis_microdatos_161107.do, obtained from lines 141-152,
produces the data: three_period_data.dta
* then figure from three_period.do

* Figure B9:
* Cost_breakdown.do
* Inputs: Costos.xlsx --> tabulated data on breakdwon of costs in 6
categories, by revenue bin
* profit_elasticity_T1.dta --> generated in bootstrap_residuals_T1 ,
data allows for a quick merge with revenue and profits by revenue bins

* Figure B10:
* REVEC_data_analysis.do
* Inputs: "REVEC_January_data.xlsx"
* Contains averages of variables of employment, salaries, share of
subsidiary firms, by revenue bins provided to us by the Central Bank
from the REVEC dataset.

* Figure B11:
* Assets.do
* Inputs: Assets.xlsx
* Tabulated data on average value of assets reported in tax
declaration by bins or reported tax revenue (Tax declaration form
D101), provided to us by the MoF from the corporate tax return
dataset.

* Figure B12: See above with figure B10

*****
* Tables
*****

* Table B1:
* revenue_bunching_robustness.do
* Inputs: Margin.xlsx

* Table B2: See above with Figure B2

* Table B3: See above with Figure B2

* Table B4:
** Discontinuities_sectors_geo.do
** Inputs: Directly run on the micro tax return data held on the
server, can not be disclosed.

* Table B5
** Analysis_microdatos_161107.do, obtained from lines 192-196
** Inputs: Directly run on the micro tax return data held on the
server, can not be disclosed.

```

* Table B6
** Audits_DDD.do
** Inputs: Directly run on the micro tax return data held on the server, can not be disclosed.

* Appendix A

* Figure A1:
* Appendix_OptimalTax_DataCreation.do --> inputs "proc/profit_elasticity_T1.dta", outputs "proc/bunching_T1_adjusted_ey_24.dta", dataset of counterfactual revenue without the change in tax rate
* Appendix_OptimalTax_simulation_OneRate.do --> runs optimal tax simulations assuming on flat tax rate on firms, inputs "proc/bunching_T1_adjusted_ey_24.dta", outputs "proc/RevenueVsTau_onerate.dta"
* Appendix_OptimalTax_simulation_TwoRate.do --> runs optimal tax simulations assuming XX, inputs "proc/bunching_T1_adjusted_ey_24.dta", outputs "proc/RevenueVsTau_tworates.dta"
* Figure created in Appendix_OptimalTax_Figure.do --> using "proc/RevenueVsTau_onerate.dta" and "proc/RevenueVsTau_tworates.dta" produces Figure A1.

* In text numbers:

In Text Numbers

"per capita GDP of 15,000 USD in PPP", pg. 9, uses "Corporate_tax_crosscountry.do" line 45
"The government collects 24% of its GDP in revenue, of which 60% are tax revenues and the rest social security contributions" pg. 9 uses: <https://data.worldbank.org/indicator/GC.REV.XGRT.GD.ZS?locations=CR> , <https://data.worldbank.org/indicator/GC.TAX.TOTL.GD.ZS?locations=CR>

"The data contain 617,588 firm-year observations and 222,352 unique firms." pg. 9, Provided to us by the Ministerio de Hacienda
"the corporate income tax raises 18% of tax revenue, around 2.5%" pg. 9 Provided to us by the Ministerio de Hacienda
"They represent 81% of the firm population, declare 20% of total profits, and generate 15% of corporate tax revenue" Provided to us by the Ministerio de Hacienda
"The ILO estimates that among formal firms in Costa Rica, only 11% of

employees are informal" footnote 39, from "ILO: 2012 Statistical Update on Employment in the Informal Economy, Table 1"