

README and Guidance for replication

by Sandra García-Urbe, garciauribe.sandra@gmail.com.

Research Economist, Banco de España

For reference to the content provided in this package, please cite: García-Urbe, S. (2022), "The Effects of Tax Changes on Economic Activity: A Narrative Approach to Frequent Anticipations", *The Economic Journal*.

Overview

This document is a guide to replicate the results of García-Urbe, S. "The Effects of Tax Changes on Economic Activity: A Narrative Approach to Frequent Anticipations", forthcoming at The Economic Journal. It constructs the results from several data sources that we document below and uses Stata, Python, R and Matlab for the purpose.

Data Availability and Provenance Statements

Statement about Rights

- [X] I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Summary of Availability

- [X] All data **are** publicly available.

Details on each Data Source

Economic Activity Data

The file "usdata.xlsx" contains all the series of economic activity indicators used in the analysis. Each indicator is the first difference of the natural logarithm of each series in levels. The exact reference for each series in levels is given below.

- Quarterly GDP data is "usqgdp.xlsx". It corresponds to the series "Gross Domestic product, Billions of Dollars, Quarterly, Seasonally Adjusted Annual Rate". U.S. Bureau of Economic Analysis, Gross Domestic Product [GDP], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/GDP>, July 21, 2016.
- GDP deflator is "deflatorGDP_2009_sa.xlsx". It corresponds to the series "Gross Domestic Product: Implicit Price Deflator, Index 2009=100, Quarterly, Seasonally Adjusted". U.S. Bureau of Economic Analysis, Gross Domestic Product: Implicit Price Deflator [GDPDEF], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/GDPDEF>, July 21, 2016.
- IPI is "ipi_2012_sa.xlsx". It corresponds to the series "Industrial Production Index, Index 2012 =100, Monthly, Seasonally Adjusted". Board of Governors of the Federal Reserve System (US), Industrial Production: Total Index [INDPRO], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/INDPRO>, July, 21, 2016.
- Manufactures is "manuf_2009_sa.xlsx". It corresponds to the series "Real Manufacturing and Trade Industries Sales, Millions of Chained 2009 Dollars, Monthly, Seasonally Adjusted". U.S. Bureau of Economic Analysis, Real Manufacturing and Trade Industries Sales [CMRMT], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/CMRMT>, July 21, 2016.
- Income is "realpi_2009_sa.xlsx". It corresponds to the series "Real personal income excluding current transfer receipts, Billions of Chained 2009 Dollars, Monthly, Seasonally Adjusted Annual Rate". U.S. Bureau of Economic Analysis, Real personal income excluding current transfer receipts [W875RX1], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/W875RX1>, July 21, 2016.
- Employment is "employ_sa.xlsx". It corresponds to "All Employees: Total Nonfarm Payrolls, Thousands of Persons, Seasonally Adjusted". U.S. Bureau of Labor Statistics, All Employees, Total Nonfarm [PAYEMS], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/PAYEMS>, July 21, 2016.
- PCE is "RealPce.xlsx". It corresponds to "Personal Consumption Expenditures, Index numbers, 2012=100, seasonally adjusted". U.S. Bureau of Economic Analysis, Personal Consumption Expenditures [PCE], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/PCE>, January 31, 2020.
- Durables is also in "RealPCE.xlsx". It corresponds to "Personal Consumption Expenditures: Durable Goods, Index numbers, 2012=100, seasonally adjusted". U.S. Bureau of Economic Analysis, Personal Consumption Expenditures [PCE], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/PCE>, January 31, 2020.
- Nondurables is also in "RealPCE.xlsx". It corresponds to "Personal Consumption Expenditures: Nondurable Goods, Index numbers, 2012=100, seasonally adjusted". U.S. Bureau of Economic Analysis, Personal Consumption Expenditures [PCE], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/PCE>, January 31, 2020.
- Farm self-employment is "Employment level - Agriculture and Related Industries, Self-Employed Workers, Unincorporated, Thousands of Persons, Monthly, Seasonally Adjusted". U.S. Bureau of Labor Statistics, Employment Level - Agriculture and Related Industries, Self-Employed Workers, Unincorporated [LNS12032185], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/LNS12032185>, August 25, 2020.
- Non-farm self-employment is "Employment level - Nonagriculture, Self-Employed Workers, Unincorporated, Thousands of Persons, Monthly, Seasonally Adjusted". U.S. Bureau of Labor Statistics, Employment Level - Nonagriculture, Self-Employed Workers, Unincorporated [LNS12032192], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/LNS12032192>, March 25, 2021.
- Compensation income is "compensation_workers.xlsx". It corresponds to "Compensation of employees, received, Billions of Dollars, Monthly, Seasonally Adjusted Annual Rate". U.S. Bureau of Economic Analysis, Compensation of employees, received [W209RC1], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/W209RC1>, February 15, 2021. It is deflated by "Personal Consumption Expenditures: Chain-type Price Index, Index 2012 = 100, Seasonally Adjusted". U.S. Bureau of Economic Analysis, Personal

Tax data

The file "romer_tax_data.xlsx" contains the series of exogenous tax shocks together with dates of approvals, implementations and the nominal magnitude. - Tax shocks retrieved from the paper: "Romer, Christina D., and David H. Romer. 2010."The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks." American Economic Review, 100 (3): 763-801. - Dates of approval of each associated tax bill were manually collected from: Romer, C.D. and Romer, D.H. (2009). "A narrative analysis of postwar tax changes", Unpublished paper, University of California, Berkeley.

Television News Data

Data on Television News were collected from the Vanderbilt Television News Archive (VTDA). I use all the news in the archive from August 1968 to December 2007. The author has no proprietary rights on this data, thus it has no rights to redistribute the data but the data is publicly accessible. Data can be accessed from <https://tvnews.vanderbilt.edu/> by using the advanced search, "Specials" were excluded from analysis. Currently, one possible strategy to search for the data is collecting each piece of information by iterating from 1 to over one million approximately on # over this url <https://tvnews.vanderbilt.edu/broadcasts/#>. Another strategy is to search by using the Broadcast Index <https://tvnews.vanderbilt.edu/siteindex>. However, these strategies might be invalid for future changes in website.

Data was structured in an spreadsheet containing variables corresponding to each broadcast piece in the following order: year, month, day, title, starting time, ending time, duration (seconds), day of the week, tv channel, text abstract. This data files were named "Vanderbilt_TV_*.csv", saved in the folder named "Vanderbilt Dataset" and processed using "02_newsuniverse.py" as detailed below.

Congressional Activity Data

Data on congressional activity was collected from <https://www.govtrack.us/start> and it provided in file "cong_activity.xlsx". Total votes collected from <https://www.govtrack.us/congress/votes> searching by session. Total laws enacted and total legislation activity from <https://www.govtrack.us/congress/bills/statistics>. All taxation bills and taxation bills enacted from <https://www.govtrack.us/congress/bills/subjects/taxation/6342> searching by session. Total tax related bills and enacted from <https://www.govtrack.us/congress/bills/browse> searching by session and text item "tax". Gross Domestic Product data used in this analysis corresponds to the series from U.S. Bureau of Economic Analysis, Real Gross Domestic Product [GDPCA], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/GDPCA>, December 20, 2019. It is provided under "GDPCA.csv".

Dataset list

Data file	Source	Notes	Provided
data/Vanderbilt Dataset/Vanderbilt_TV_*.csv	Vanderbilt Television Data Archive	Public and proprietary	No
data/Tax changes/romer_tax_data.xlsx	All listed	As per terms of use	Yes
data/Economic Activity/usdata.xlsx	All listed	As per terms of use	Yes
data/Congressional Activity/cong_activity.xlsx	All listed	As per terms of use	Yes
data_mining/cleaned_data/episodedata_*.xls	romer_tax_data.xlsx structured data with 01_taxshocks.do used for input data to the FF	As per terms of use	Yes
data_mining/cleaned_data/rr_episodes.xls	romer_tax_data.xlsx structured data with 01_taxshocks.do used for input data to the FF	As per terms of use	Yes
analysis/tax_effects/taxshocks_*.xlsx	romer_tax_data.xlsx structured data with 01_taxshocks.do used for input data to the FF	As per terms of use	Yes
analysis/tax_effects/FF_*.xlsx	media anticipations measures computed with codes in 04_news_anticipation.R used for 05_taxeffects.m	As per terms of use	Yes

Computational requirements

Software Requirements

- Stata MP 15
- Python 2.7.13 with Anaconda2 and Spyder.
 - [nltk 3.2.4](#)
 - [pandas 0.20.1](#)

- [unicodedata2 9.0.0.post4](#)
 - [regex 2017.7.28](#)
 - [Pattern 2.6](#)
 - [numpy 1.13.1](#)
- Matlab R2019a with Paralell Computing Toolbox and Econometrics Toolbox.
- R 3.6.2 with RStudio 1.2.5033
 - [readxl '1.3.1'](#)
 - [xts '0.11.2'](#)
 - [zoo '1.8.6'](#)
 - [NLP '0.2.0'](#)
 - [tm '0.7.7'](#)
 - [psych '1.9.12'](#)
 - [wordcloud '2.6'](#)
 - [randomForest '4.6.14'](#)
 - [fuzzyforest '1.0.6'](#)
 - [foreach '1.4.7'](#)
 - [doRNG '1.7.1'](#)
 - [WGCNA '1.68'](#)
 - [ggplot2 '3.2.1'](#)
 - [MLmetrics '1.1.1'](#)

Details

The code was last run on a **4-core Intel-based desktop with Windows 10 Enterprise version 1909, 64 GB of RAM and 1 TB of fast local storage**. Computation took 25 days, 24 days of which correspond to the computation of “five_taxeffects.m” using Matlab R2019a paralell processing tool `parpool` with four workers.

To obtain the Python version first [download the distribution Anaconda2 4.4.0 from Anaconda.org repository](#). Then open Anaconda Prompt and run `pip install package==version` for each package in the Software requirements list, except for `pandas` which comes with the exact version. Before running the replication package, open Python and run `import nltk` then run `nltk.download()` and download four packages: `punkt`, `averaged_perceptron_tagger`, `maxent_ne_chunker`, `words` and `stopwords`.

To setup R first [download the R 3.6.2 version](#) and install it. You would also need to [download and install Rtools 3.5](#) to install packages from CRAN source. Please, find a guide to setup Rtools [here](#). To install each package, open R and install package `devtools`, then `require(devtools)`, then run `install_version("packagename", version = "X", repos = "http://cran.us.r-project.org")` for each package required, where packagename is each of the package names listed and X is the corresponding version number. The exception is package `WGCNA` which requires running `install.packages("BiocManager")` and `BiocManager::install("WGCNA")`.

The results that the package generates depend on the specified requirements. A slight departure from the requirements, even the operating system version, may also produce a slight variation of the results. In particular, Figure 2 layout changes though the represented frequencies won't. Importantly, the replicator must bear in mind that if the requirements are not satisfied the package may not work.

Instructions to Replicators

- `data` should contain all the required raw datasets as detailed above.
- `data_mining/tax_changes` contains the codes that create all data files containing monthly tax shocks features that use subsequent codes.
- `data_mining/news_universe` contains the codes that create the files containing preprocessed text data on the TV news and episode dates that use subsequent codes.
- `analysis/media_anticipation` contains the code that estimates the anticipation measures that subsequent codes use to estimate anticipation effects and the plots and tables that document the performance of Fuzzy Forests and comparasion to Random Forest which are stored in `results`.
- `analysis/tax_effects` contains the output from Fuzzy Forest, the measures of tax changes and the codes (at `Semi_bootstrap_SE` and `VAR`) that execute the estimation of the effects of tax changes on economic activity for the different specifications of the paper and appendix. The results are stored in `results`.
- `analysis/cong_activity_endogeneity` contains the code that executes a robustness analysis of endogeneity of tax approvals to economic activity and the output table.

Details

To replicate the output presented in the paper the following files should be executed in order from the main folder `replication package`.

- `01_taxshocks.do` :
 - It executes several do-files in sequence to process tax shocks data into files for use in final datasets for anticipation measure estimation and tax effects estimation. It also creates tables of summary statistics.
 - These programs were last run in April, 2021.
 - It should take about 1 minute to run.
 - If running programs individually, note that ORDER IS IMPORTANT.
 - To adjust the directory to run this code type `cd` followed by the directory pointing to `replication package` in the command prompt.
- `02_newsuniverse.py` :
 - It executes several python scripts in sequence to process VTDA data into final datasets to use in the estimation of anticipations measure.
 - The programs were last run top to bottom on March, 2021.
 - It should take about 5 hours to run.

- If running programs individually, note that ORDER IS IMPORTANT.
 - To adjust the directory to run this code modify variable `ORPATH`.
- 03_summary_rawnews.do : Code that generates summary statistics of the data on tax news and tax news coverage plot across time. The programs were last run top to bottom on March, 2021. This program should take about 1 minute to run. To adjust the directory to run this code type `cd` followed by the directory pointing to `replication package` in the command prompt.
- 04_news_anticipation.R :
 - It executes descriptive plots on tax shocks and tax salience, tax wordcloud and top tax news keywords frequency. Then estimates the baseline measure of tax anticipation, the size measure of tax anticipations and compares results from Fuzzy Forests and Random Forests.
 - The programs were last run top to bottom on March, 2021.
 - It should take about 6 hours to run.
 - To adjust the directory to run this code type `setwd(x)` in the command prompt where `x` is the directory pointing to `replication package`.
- five_taxeffects.m :
 - It computes impulse responses with standard error bands of the effects of tax anticipations and tax changes on economic activity by means of Mixed Frequency Dynamic Factor model (MFDF) at `Semi_bootstrap_SE` and Vector Autoregressive model (VAR) at `VAR`.
 - It provides the resulting figures presented in the paper and the online appendix.
 - The programs were last run top to bottom on February, 2021.
 - It should take about 24 days to run under specified technical requirements above. If the user can dedicate more than four cores to run this code, she may increase the speed by modifying the parameter inside `parpool` of the files that are called in lines 4 and 11, at lines 194 and 190, respectively.
 - If running programs individually, note the ORDER IS IMPORTANT.
 - To adjust the directory to run this code type `cd` followed by the directory pointing to `replication package` in the command prompt.
- 06_endogeneity.do : Code that generates analysis on the relationship of economic activity and congressional activity. The programs were last run in March, 2021. It should run in 2 minutes. It generates Table 2 presented in the paper. To adjust the directory to run this code type `cd` followed by the directory pointing to `replication package` in the command prompt.

List of tables and programs

The provided code reproduces all tables and figures in the paper. Those tables and figures which use VTDA data are indicated.

Figure/Table #	Program	Line Number	Output file	Note
Table 1	03_summary_rawnews.do	30		VTDA data required for reproduction
Table 2	06_endogeneity.do		AppEnd.tex	
Figure 1	n.a. (no data)			Manually designed
Figure 2	TaxNews_ExplorationR.R	78	wordcloud_universe_R.png	VTDA data required for reproduction
Figure 3	fuzzyrf_relnews_sign.R	256-308	fitinc.png and fitcut.png	VTDA data required for reproduction
Figure 4	fuzzyrf_relnews_sign.R	179-182 & 207-209	ff_news_inc.png and ff_news_cut.png	VTDA data required for reproduction
Figure 5	monthly_taxshocks_RR_MR_RFsign_simu.m	270-308	fig5.png	
Figure 6	monthly_taxshocks_RR_MR_RFsign_simu.m	317-365	fig6a.png and fig6b.png	
Figure 7	monthly_taxshocks_comparison.m	32-55 & 85-110	fig7a.png and fig7b.png	
Figure 8	monthly_taxshocks_RR_nolinMRlessleads_RFsign_simu.m	267-289	fig8.png	
Figure 9	VARfinal_sign_linear.m	140-312	fig9.png	
Figure 10	monthly_taxshocks_RR_MR_RFsize_simu.m	274-340	fig10.png	
Table A1	descriptive_analysis.do	68		
Table A2	descriptive_analysis.do	30-31		
Table A3	descriptive_analysis.do	64		
Table A4	descriptive_analysis.do	28 and 65		
Table A5	n.a. (no data)			Manually generated
Table A6	monthly_taxshocks_RR_MR_RFsign_simu.m	77-80	tableA6.xlsx	
Table A7	TaxNews_ExplorationR.R	100-103		VTDA data required for reproduction

Figure/Table #	Program	Line Number	Output file	Note
Table A8	comparisonRF_FF.R	130-143 & 172-185		VTDA data required for reproduction
Figure A1	Description_taxes.R	14	implementation.png	
Figure A2	Description_taxes.R	26	approval.png	
Figure A3	Description_taxes.R	40	salience_tvnews.png	
Figure A4	03_summary_rawnews.do	82	annual_coverage.png	VTDA data required for reproduction
Figure A5	fuzzyrf_relnews_sign.R	240-252	oobinc.png and oobcut.png	VTDA data required for reproduction
Figure A6	comparisonRF_FF.R	210-217	oob_comparison.png	VTDA data required for reproduction
Figure A7	monthly_taxshocks_RR_MR_RFsign_simu_RR2robustness.m	266-304	figA7.png	
Figure A8	monthly_taxshocks_RR_MR_RFsign_simu_1mwindow.m	270-306	figA8.png	
Figure A9	VARfinal_sign_linear_A9.m	140-346	figA9.png	
Figure A10	VARfinal_sign_linear_A10.m	140-346	figA10.png	
Figure A11	VARfinal_sign_linear_A11.m	140-348	figA11.png	

References

- Romer, C.D. and Romer, D.H. (2010). “The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks”, *American Economic Review*, vol. 100(3), pp. 763-801.
- Romer, C.D. and Romer, D.H. (2009). “A narrative analysis of postwar tax changes”, Unpublished paper, University of California, Berkeley.