Overview

This README file describes the replication archive for this paper. There are two parts. Part 1 (ML) consists of the Python codes/data used for training the voice emotion model and the text sentiment model. Part 2 (Main) consists of Stata codes and data required to reproduce tables and figures in the paper. Some external data cannot be publicly posted and instructions on how to acquire access are included below.

Data Availability and Provenance Statements

Statement about Rights

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

- I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permission are documented in the LICENSE.txt file.

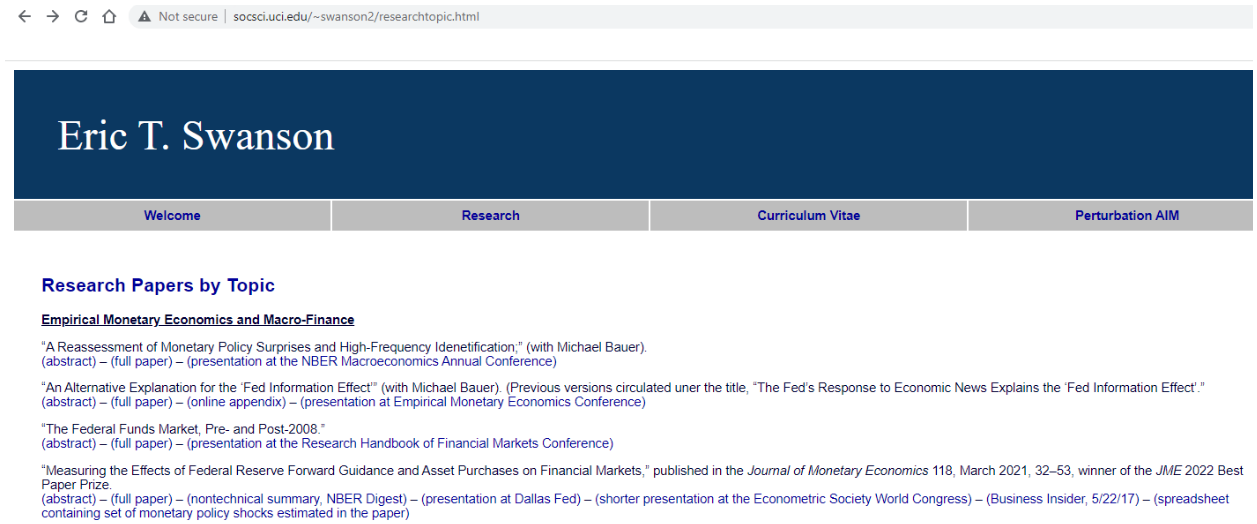
Summary of Availability

- Some data **cannot be made** publicly available.

Details on each Data Source

The following data are publicly available and are included in the repository:

- Series of policy shocks constructed by Swanson (2021): http://www.socsci.uci.edu/~swanson2/researchpublished.html. As shown in the author's personal webpage, this is a public dataset (see screenshot below).



Not secure | socsci.uci.edu/~swanson2/researchtopic.html

Eric T. Swanson

Welcome | Research | Curriculum Vitae | Perturbation AIM

Research Papers by Topic

Empirical Monetary Economics and Macro-Finance

“A Reassessment of Monetary Policy Surprises and High-Frequency Idenetification;” (with Michael Bauer).
(abstract) – (full paper) – (presentation at the NBER Macroeconomics Annual Conference)

“An Alternative Explanation for the ‘Fed Information Effect’” (with Michael Bauer). (Previous versions circulated uner the title, “The Fed’s Response to Economic News Explains the ‘Fed Information Effect’.”
(abstract) – (full paper) – (online appendix) – (presentation at Empirical Monetary Economics Conference)

“The Federal Funds Market, Pre- and Post-2008.”
(abstract) – (full paper) – (presentation at the Research Handbook of Financial Markets Conference)

“Measuring the Effects of Federal Reserve Forward Guidance and Asset Purchases on Financial Markets,” published in the *Journal of Monetary Economics* 118, March 2021, 32–53, winner of the *JME* 2022 Best Paper Prize.
(abstract) – (full paper) – (nontechnical summary, NBER Digest) – (presentation at Dallas Fed) – (shorter presentation at the Econometric Society World Congress) – (Business Insider, 5/22/17) – (spreadsheet containing set of monetary policy shocks estimated in the paper)

- Series of shadow rate for the US constructed by Wu. and Xia (2016): https://sites.google.com/view/jingcynthiawu/shadow-rates. This is a public dataset, as stated in Footnote 1 of the related publication (https://onlinelibrary.wiley.com/doi/full/10.1111/jmcb.12300): “Our shadow rate data with monthly update are available at the Atlanta Fed (https://www.frbatlanta.org/cqer/research/shadow_rate.aspx) or our webpage (http://faculty.chicagobooth.edu/jing.wu/research/data/WX.html).” (Note that the author’s personal webpage has changed since the publication.)
- Speech data used to train the neutral network: the Ryerson Audio-Visual Database of Emotional Speech and Song (https://doi.org/10.5281/zenodo.1188975) and the Toronto emotional speech set (https://hdl.handle.net/1807/24487)
- Data for the text analysis: FOMC press releases and transcripts of FOMC press conferences (https://www.federalreserve.gov/monetarypolicy/fomc_historical.htm). FOMC_transcripts.zip contains all transcripts (in pdf format) of FOMC meetings over the 2011 – June 2019 period. The transcripts were converted into txt format and manually split into opening remarks, questions, and answers. The split data are also included in this depository.

The following data are publicly available but are not included in the repository due to redistribution rights:

- Series of daily bond and stock prices (except the data for ticker LQHD) are Tiingo (https://api.tiingo.com/products/end-of-day-stock-price-data). Free registration is required to use Tiingo's End-of-Day API. Data in csv format can be obtained by providing the ticker symbol in the following link: https://api.tiingo.com/tiingo/daily/{ticker}/prices?startDate=2011-1-1&format=csv. For example, the link to download data for &P 500 ETF (ticker symbol: SPY) is: https://api.tiingo.com/tiingo/daily/SPY/prices?startDate=2011-1-1&format=csv
- Series of daily prices for iShares Interest Rate Hedged Corporate Bond ETF (LQDH) are obtained from Yahoo Finance (https://finance.yahoo.com/quote/LQDH/history?p=LQDH). To obtain the data, follow this link, change the start date to 18-06-2014 (data prior to 17 June 2014 are not available) and the end date to 31-07-2019, apply, then download.
- Series of daily number of corporate earnings announcements in the US are obtained from Yahoo Finance (https://finance.yahoo.com/calendar/earnings/). Note that by default, Yahoo Finance only allows to view data for 1 month. Thus, one needs to obtain data on the rolling window basis. For example, the link to obtain data for 30 days prior to and including 01 March 2011 is https://finance.yahoo.com/calendar/earnings/?day=2011-03-01.

The following data are not publicly available:

- The tick data for SPY used to construct the intra-day returns in Figure 15 are not publicly available. The data can be obtained from the TAQ intraday dataset accessible via the Wharton Research Data Services (WRDS). Steps to obtain the data are as follows:
    - In WRDS interface, (https://wrds-www.wharton.upenn.edu/), click on All Data and choose TAQ

    - Choose Millisecond Trade and Quote – “Daily Product”. 2003 – present, updated daily > Consolidated Trades
    - Choose date range (1) (from 01-01-2011 to 31-07-2019); enter “SPY” in Company Codes (3); select Date of trade (DATE), Time of Trade or Quote in milliseconds (HHMMSSXXX), and price of trade (PRICE) in query variables (4); then download in preferred output format
- Citigroup US Economic Surprise Index data (used in Panel A – Figure 14) cannot be shared by the authors. The data can be purchased from Haver Analytics (https://www.haver.com/) with the series code: V111CSI@INTDAILY. Alternatively, the data can also be obtained from https://macrovar.com/united-states/us- citigroup-economic-surprise-index/ using a premium subscription.
- The raw text of news obtained from Nexis Uni of which the predicted sentiment is used in Panel C - Figure 14 and Panel B - Figure 16 cannot be shared by the authors. The data can be obtained from the Nexis Uni database by searching for news that contain the keyword “FOMC”. Within the returned search results, the following additional search criteria are applied:
    - Timeline: 01/01/2011 – 15/07/2019
    - Publication types: Choose “Newspapers”, “Newswires & Press Releases”, and “Webnews”
    - Search within results (i.e., additional keywords to be included): “interest rate” or “monetary” or “federal funds rate” or “fed funds rate”
    - Geography: North America > United States
    - Language: English
    - Sources: Untick “Targeted News Service”, "CQ Federal Department and Agency Documents“, “CQ Congressional Press Releases”, “US Fed News”, “US Official News"

  Note that the sample used in the paper was obtained on 18 June 2021. As Nexis Uni is constantly updated (i.e., some new sources are added, some others are removed), the new search with these above criteria might not return the exact same results as the one we obtained.

- The raw text of tweets obtained from Twitter of which the predicted sentiment is used in Panel A - Figure 16 cannot be shared by the authors. To obtain the historical tweets, please follow one of the two approaches below. For both approaches, you need to have an approved Twitter developer account.
  Approach 1:
  - Apply for Twitter's the Academic Research API by providing the required information (https://developer.twitter.com/en/products/twitter-api/academic-research). Note that each application is subject to Twitter's review and approval.
  - Upon approval, use the Search Tweets API to obtain all historical tweets posted by the Fed accounts with the following search criteria:
    - query: (from:stlouisfed OR from:sffed OR from:RichmondFed OR from:philadelphiafed OR from:NewYorkFed OR from:MinneapolisFed OR from:KansasCityFed OR from:federalreserve OR from:DallasFed OR from:ClevelandFed OR from:ChicagoFed OR from:BostonFed OR from:AtlantaFed)
    - start_time=2011-01-01T00:00:00.000Z
    - end_time=2019-07-16T00:00:00.000Z
  - Detailed documentation/sample code of the Search Tweets API can be found at https://developer.twitter.com/en/docs/twitter-api/tweets/search/introduction

  Approach 2: One could also use the tweet ids provided in file "media_all.xlsx" in this repository (variable "tweet_id", sheet "tweets_sentiment", file "media_all.xlsx") to re-hydrate the tweets. To do so, you need to use Twitter's Tweets lookup API of which detailed documentation can be found at https://developer.twitter.com/en/docs/twitter-api/tweets/lookup/api-reference/get-tweets-id.

- The audio files of the FOMC press conferences cannot be posted publicly due to Youtube's Terms of Service (https://www.youtube.com/static?gl=GB&template=terms). The files covering the sample period (April 2011 – June 2019) should be downloaded from the Fed's Youtube channel (https://www.youtube.com/watch?v=H-dV2VUNh4E&list=PL159CD41EB36CFE86&ab_channel=FederalReserve). Manual pre-processing is required to cut audio files into segments: opening remarks, questions, and answers.

Dataset list

Data file	Source	Notes	Provided
`ML/text/data/dictionary.csv`			Yes
`ML/text/data/FED_classification.csv`			Yes
`ML/text/data/FED_text.xlsx`	FOMC	Questions-answers split based on the original FOMC transcripts, serves as input for text_prediction.py.	Yes
`FOMC_transcripts.zip`	FOMC	FOMC transcripts (in pdf format)	Yes
`ML/voice/data/RAVDESS.zip`	Livingstone and Russo (2018)		Yes
`ML/voice/data/TESS.zip`	Pichora-Fuller and Dupuis (2010)		Yes
`Main/data/raw/shadowrate.xls`	Wu and Xia (2016)		Yes
`Main/data/raw/policyshocks_swanson.xlsx`	Swanson (2021)		Yes
`Main/data/raw/media_all.xlsx`		Predicted sentiment for Fed tweets and FOMC-related news	Yes
`Main/data/raw/fomc_all.xlsx`		Predicted sentiment for FOMC texts	Yes

Main/data/raw/FED_QA_timing.dta			Yes
Main/data/raw/CESI.xlsx	MacroVar		No
Main/data/raw/EUR.csv	Tiingo		No
Main/data/raw/GBP.csv	Tiingo		No
Main/data/raw/GLD.csv	Tiingo		No
Main/data/raw/GOVT.csv	Tiingo		No
Main/data/raw/IEF.csv	Tiingo		No
Main/data/raw/IEI.csv	Tiingo		No
Main/data/raw/IVR.csv	Tiingo		No
Main/data/raw/JPY.csv	Tiingo		No
Main/data/raw/LQD.csv	Tiingo		No
Main/data/raw/LQDH.xlsx	Yahoo Finance		No
Main/data/raw/SHV.csv	Tiingo		No
Main/data/raw/SHY.csv	Tiingo		No
Main/data/raw/SPY.csv	Tiingo		No
Main/data/raw/TIP.csv	Tiingo		No
Main/data/raw/TLH.csv	Tiingo		No
Main/data/raw/TLT.csv	Tiingo		No
Main/data/raw/VIX.csv	Tiingo		No
Main/data/raw/VIXM.csv	Tiingo		No
Main/data/raw/VIXY.csv	Tiingo		No
Main/data/yahoo_earnings.dta	Yahoo Finance		No
Main/data/est_*.dta		All estimation samples in dta format, needed to reproduce tables and figures	Yes

Computational requirements

Software Requirements

- Stata (code was last run with version 17): `outreg2, rego, reghdfe, ftools` (as of 01-12-2022)

- Python 3.6+. The specific Python and dependencies requirements can be found in the requirements.txt files in "ML/text/codes" and "ML/voice/codes" folders.

Memory and Runtime Requirements

Summary

Approximate time needed to reproduce the main analyses (using Stata codes) on a standard (2022) desktop machine:

- 8-24 hours

Approximate time needed to reproduce the voice and text analyses (using Python codes) on a standard (2022) desktop machine:

- Not feasible to run on a desktop machine, as described below.

Details

- The Python codes to train the voice emotion and text sentiment neural network models were run on the HPE server with 384 GB memory, 2 CPUs (Intel® Xeon(R) CPU E5-2650 @ 2.6GHz x 8 cores) using Ubuntu 20.04 LTS operating system. The approximate time needed is 72 hours.

- All Stata codes including those to produce tables and figures in the paper can be run on a standard laptop or desktop machine such as a HP Z420 Workstation with 16GB memory, Intel® Xeon(R) CPU E5-1650 @ 3.5GHz x 6 cores using Windows 10 operating system. The approximate time needed to run Readme.do is 17 hours.

Description of programs/code

- Programs in `ML/text/codes` will analyze the FOMC texts, FOMC news, and the Fed tweets to produce the predicted text sentiment and sentiment intensity used in the manuscript.

- Programs in `ML/voice/codes` will process and analyze FOMC audios to produce the predicted speech emotions used to construct the *VoiceTone* measure.

- Programs in `Main/codes` generate all tables and figures in the main body of the article and online appendix (except Table C1 which is produced by running voice_model.py; Tables D1-D2 which are not empirical, and Figures B1-B3 which are examples). The program `Main/codes/Readme.do` will run them all in order. Each program called from `Readme.do` identifies the table or figure it creates (e.g., `Table1.do`). Output files are called appropriate names (e.g., `Table1.txt`, `Fig1.png`) and should be easy to correlate with the manuscript.

Instructions to Replicators

Details

Machine learning model for voice emotion detection

1. Folder `ML/voice/codes` contains the following files:
    - `requirements.txt`: Python packages to be installed before running the codes
    - `voice_preprocessing.py`: pre-process audio files and extract vocal features
    - `voice_model.py`: train models using features extracted previously. This should be run after voice_preprocessing.py
    - `voice_prediction.py`: use the trained model to predict emotions in the prediction data. This should be run after voice_model.py
2. Folder `ML/voice/data/raw` contains the following files:
    - `RAVDESS.zip`: The zipped folder contains the Ryerson Audio-Visual Database of Emotional Speech and Song. Note that we only used the speech data. The

filename consists of a 7-part identifier (e.g., 03-01-06-01-02-01-12.mp4). Filename identifiers are as follows:

  - Modality (03 = audio-only).
  - Vocal channel (01 = speech).
  - Emotion (01 = neutral, 02 = calm, 03 = happy, 04 = sad, 05 = angry, 06 = fearful, 07 = disgust, 08 = surprised).
  - Emotional intensity (01 = normal, 02 = strong).
  - Statement (01 = "Kids are talking by the door", 02 = "Dogs are sitting by the door").
  - Repetition (01 = 1st repetition, 02 = 2nd repetition).
  - Actor (01 to 24. Odd numbered actors are male, even numbered actors are female).

- `TESS.zip`: The zipped folder contains the Toronto emotional speech set. The filename consists of a 5-part identifier (e.g., 03-01-06-back-26.wav). Filename identifiers are as follows:
  - Modality (03 = audio-only).
  - Vocal channel (01 = speech).
  - Emotion (01 = neutral, 03 = happy, 04 = sad, 05 = angry, 06 = fearful, 07 = disgust, 08 = pleasant surprised).
  - Target word (in total 200 target words).
  - Actor (26 and 28).
- *For editorial purposes, we provide the `FED_audios.zip` file which contains the audio files of FOMC press conferences. This file cannot be shared with replicators other than data editors.*

Machine learning model for text sentiment analysis

1. Folder `ML/text/codes` contains the following files:
   - `requirements.txt`: Python packages to be installed before running the codes
   - `text_model.py`: train a model using embeddings obtained from BERT/RoBERTa

- `text_prediction.py`: apply the trained model on the FOMC data to predict text sentiment. This should be run after text_model.py
- `finbert.py`: generate text sentiment using FinBert
- `search_and_count.py`: generate number of hawkish/dovish phrases following Neuhierl and Weber (2019)
- `sentiment_intensity.py`: generate a measure of sentiment intensity following Kozlowski et al. (2019) and Jha et al. (2021)

2. Folder `ML/text/data` contains the following files:
   - `FED_classification.csv`: labelled text for training
   - `FED_text.xlsx`: text data of the FOMC meetings/press conferences. Pre-processing was performed to split text of press release and opening remarks into paragraphs and text of the Q&A session into questions and answers.
   - `dictionary.csv`: the hawkish/dovish pairs used to calculate the baseline dovish dimension (input of `sentiment_intensity.py`)

Stata codes and data to produce tables and figures are included in the `Main` folder

1. The `Main/data/raw` folder contain the following files:
   - `fomc_all.xlsx`: predicted voice emotions for answers and predicted text sentiment for answers, opening remarks, and press releases
   - `media_all.xlsx`: predicted text sentiment for FOMC-related media news and tweets created by the Fed' Twitter accounts.
   - `policyshocks_swanson.xlsx`: data on policy shocks in Swanson, E.T., 2021. "Measuring the effects of Federal Reserve forward guidance and asset purchases on financial markets." Journal of Monetary Economics 118:32-53.
   - `shadowrate.xlsx`: data from Wu, J.C. and Xia, F.D., 2016. "Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound." Journal of Money, Credit and Banking 48(2-3):253-291.
   - `FOMC_QA_timing.dta`: data on the timing of each answer in a press conference.
   - *For editorial purposes, we provide the following files which cannot be shared with replicators other than data editors:*

- `CESI.xlsx` file which is the Citigroup Economic Surprise Index data
- csv/excel files containing stock prices
- `SPY_tick.dta` file which is the raw tick data for SPY

2. The `Main/data` folder contains the following files:
   - `est_*.dta`: estimation samples
   - *For editorial purposes, we provide the following file which cannot be shared with replicators other than data editors:*
     - `yahoo_earnings.dta`
3. The `Main/codes` folder contains the Stata codes needed to process, merge data, and run estimations. `Readme.do` will run all do files (from pre-processing to estimation steps) in order.
   - The codes in Block 1 are to pre-process financial and FOMC data. If daily prices data are not available, skip lines 97-195 in `pre_process0.do` and the entire `merge_all.do`. If tick data for SPY (`SPY_tick.dta`) are not available, skip `pre_process1.do`.
   - All programs in Blocks 3 and 4 except `Figure15.do` in `Readme.do` can be run without additional confidential data or data that cannot be redistributed, but Panels A and B of Figure 14 will not be replicated.

List of tables and programs

The provided code reproduces:

- Selected tables and figures in the paper, as explained and justified below.

Figure/Table #	Program	Line Number	Output
Tables and Figures in the paper			
Table 1	`Main/codes/Table1.do`	Entire program	Table1.csv
Figure 1	`Main/codes/Figures1-2.do`	79	Fig1.png

Figure 2	`Main/codes/Figures1-2.do`	128	Fig2.png
Figure 3	`Main/codes/Figures3-13.do`	174	Fig3.png
Figure 4	`Main/codes/Figures3-13.do`	175	Fig4.png
Figure 5	`Main/codes/Figures3-13.do`	176	Fig5.png
Figure 6	`Main/codes/Figures3-13.do`	177	Fig6.png
Figure 7	`Main/codes/Figures3-13.do`	178	Fig7.png
Figure 8	`Main/codes/Figures3-13.do`	254	Fig8.png
Figure 9	`Main/codes/Figures3-13.do`	180	Fig9.png
Figure 10	`Main/codes/Figures3-13.do`	181	Fig10.png
Figure 11	`Main/codes/Figures3-13.do`	182	Fig11.png
Figure 12	`Main/codes/Figures3-13.do`	183	Fig12.png
Figure 13	`Main/codes/Figures3-13.do`	184	Fig13.png
Figure 14	`Main/codes/Figure14.do`	429	Fig14.png
Figure 15	`Main/codes/Figure15.do`	216	Fig15.png
Figure 16	`Main/codes/Figure16.do`	217	Fig16.png
Tables and Figures in the Online Appendix			
Table A1	`Main/codes/AppendixTablesA1-A2.do`	13	AppendixTableA1.xlsx
Table A2	`Main/codes/AppendixTablesA1-A2.do`	24	AppendixTableA2.xlsx
Figure A1	`Main/codes/AppendixFiguresA1-A8.do`	174	AppendixFigA1.png
Figure A2	`Main/codes/AppendixFiguresA1-A8.do`	175	AppendixFigA2.png
Figure A3	`Main/codes/AppendixFiguresA1-A8.do`	176	AppendixFigA3.png
Figure A4	`Main/codes/AppendixFiguresA1-A8.do`	177	AppendixFigA4.png
Figure A5	`Main/codes/AppendixFiguresA1-A8.do`	178	AppendixFigA5.png
Figure A6	`Main/codes/AppendixFiguresA1-A8.do`	179	AppendixFigA6.png
Figure A7	`Main/codes/AppendixFiguresA1-A8.do`	180	AppendixFigA7.png
Figure A8	`Main/codes/AppendixFiguresA1-A8.do`	181	AppendixFigA8.png

Figure A9	`Main/codes/AppendixFigureA9.do`	100	AppendixFigA9.png
Figure B1	Example (not created from code or data)		
Figure B2	Example (not created from code or data)		
Figure B3	Example (not created from code or data)		
Table C1	`ML/voice/codes/voice_model.py`	138	
Figure D1	`Main/codes/AppendixFigureD1.do`	94	AppendixFigD1.png
Figure D2	`Main/codes/AppendixFigureD2.do`	138	AppendixFigD2.png

References

Federal Open Market Committee. 2011-2019. "FOMC Press Conferences". https://www.youtube.com/watch?v=HdV2VUNh4E&list=PL159CD41EB36CFE86&ab_channel=FederalReserve (accessed 28 November 2022)

Federal Open Market Committee. 2011-2019. https://www.federalreserve.gov/monetarypolicy/fomc_historical.htm (accessed 28 November 2022)

Haver Analytics. 2022. https://www.haver.com/ (accessed 28 November 2022)

Livingstone, S.R. and Russo, F.A., 2018. "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English: Dataset." *PLoS ONE*. https://doi.org/10.5281/zenodo.1188975 (accessed 28 November 2022)

Livingstone, S.R. and Russo, F.A., 2018. "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English." *PLoS ONE* 13(5):e0196391.

Nexis Uni. 2011-2019. https://www.nexisuni.com (access 28 November 2022)

Pichora-Fuller, M.K and Dupuis, K., 2010. "Toronto emotional speech set." University of Toronto. https://hdl.handle.net/1807/24487 (accessed 28 November 2022)

Tiingo. 2011-2019. https://api.tiingo.com/documentation/end-of-day (accessed 28 November 2022)

Twitter Full-archive search API. 2022. https://developer.twitter.com/en/docs/twitter-api/tweets/search/introduction (accessed 28 November 2022)

WRDS. 2022. “TAQ”. https://wrds-www.wharton.upenn.edu/ (accessed 28 November 2022)

Yahoo Finance. 2011–2019. “Company Earnings Calendar.” https://finance.yahoo.com/calendar/earnings/ (accessed 28 November 2022)

Yahoo Finance. 2011-2019. “iShares Interest Rate Hedged Corporate Bond ETF (LQDH)” https://finance.yahoo.com/quote/LQDH/history?p=LQDH (accessed 28 November 2022)

Swanson, E.T., 2021. “Measuring the effects of Federal Reserve forward guidance and asset purchases on financial markets: Dataset.” *Journal of Monetary Economics* (accessed 28 November 2022)

- shadowrate.xlsx: data from Wu, J.C. and Xia, F.D., 2016. “Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound.” Journal of Money, Credit and Banking 48(2-3):253-291.