**Replication Package for “The Economic Impact of Uncertainty about U.S. Regulations of the Energy Sector”**

Xiaohan Ma & Zhoudan Xie

This repository provides the data and code to replicate the results from the paper titled “**The Economic Impact of Uncertainty about U.S. Regulations of the Energy Sector**.” The package includes two main components:

1. **/measure\_uncertainty**: Data and Python code for measuring regulatory uncertainty, as described in Section 2 of the paper.
2. **/empirical\_analysis**: Data and Stata code for generating empirical results, as described in Section 3 of the paper.

Questions and comments can be directed to the authors at [xiaohan.ma@ttu.edu](mailto:xiaohan.ma@ttu.edu) or [zxie@gwu.edu](mailto:zxie@gwu.edu).

# **Directory Structure and Files**

## **/measure\_uncertainty**

This paper analyzes the full text and metadata of newspaper articles from the U.S. Newsstream database, accessed through ProQuest’s [TDM Studio](https://tdmstudio.proquest.com/home). Due to copyright restrictions, the authors cannot distribute the full text of the articles analyzed. However, this directory provides the ProQuest IDs and some metadata for all 600,953 articles included in the analysis (available in /data/all\_uncertainty\_scores.csv). Researchers with access to ProQuest’s content can use this information to retrieve the full text and additional metadata of the articles.

To demonstrate the textual analysis process, this directory includes demo data for five articles in XML format (located in /data/nlp\_demo/xml\_examples) along with the Python scripts for performing the textual analysis and estimating the uncertainty indexes.

The monthly uncertainty indexes, estimated using all articles, are provided in the /data subdirectory. A detailed description of the specific files in this directory is outlined below.

### **/data**

* oil\_regulatory\_uncertainty\_index.csv: estimated oil regulatory uncertainty indexes, including:
  + “RegUncertaintyIndex”: baseline oil regulatory uncertainty index;
  + “RegUncertaintyIndex\_Econ”: oil regulatory uncertainty index estimated while controlling for uncertainty scores of economic sections (i.e., economic-adjusted index);
  + “RegUncertaintyIndex\_Journal”: oil regulatory uncertainty index estimated using trade journals and magazines;
  + “RegUncertaintyIndex\_Broad”: oil regulatory uncertainty index estimated using a broader set of energy terms.
* oil\_supply\_uncertainty\_index.csv: estimated general oil supply uncertainty index (“UncertaintyIndex”).
* all\_uncertainty\_scores.csv: ProQuest ID, publication title, publication date, and uncertainty scores calculated based on different sections of an article, including:
  + “RegUncertaintyScore”: uncertainty score from the regulatory section;
  + “EconUncertaintyScore”: uncertainty score from the economic section;
  + “UncertaintyScore”: uncertainty score from the full text of the article.
* noun\_chunks\_by\_month\_reg.csv: all noun chunks and their occurrences from the regulatory sections with positive regulatory uncertainty scores that were published during a given month, used to generate Figure 2.
* noun\_chunks\_by\_month\_general.csv: all noun chunks and their occurrences from the full news articles with positive uncertainty scores that were published during a given month, used to generate Figure 4.
* /nlp\_demo/xml\_examples: demo data containing XML files of five randomly selected news articles.
* /supplementary\_data:
  + categorical\_epu\_index.xlsx: U.S. categorical economic policy uncertainty (EPU) index originated from Baker et al. (2016) (available at <https://www.policyuncertainty.com/categorical_epu.html>).
  + cpu\_index.csv: climate policy uncertainty (CPU) index from Gavriilidis (2021) (available at <https://www.policyuncertainty.com/climate_uncertainty.html>).
  + gpr\_index.xls: geopolitical risk (GPR) index from Caldara and Iacoviello (2022) (available at <https://www.matteoiacoviello.com/gpr.htm>).
  + eia\_energy\_glossary.xlsx: glossary terms in the natural gas and petroleum categories from the U.S. Energy Information Administration (available at <https://www.eia.gov/tools/glossary/>).
  + lm\_sentiment.csv: 2018 version of the Loughran and McDonald dictionary (available at <https://sraf.nd.edu/loughranmcdonald-master-dictionary/>).
  + pub\_title\_newspaper.csv: a crosswalk of publication titles and newspaper names, created by the authors; publication titles are from ProQuest, and a newspaper name can correspond to multiple publication titles.
  + coolvetica\_rg.otf: a font file for generating figures.
  + palatinolinotype\_roman.ttf: a font file for generating figures.
  + wordcloud\_mask.png: a shape image for generating the word cloud figures.

### **/python\_code**

A virtual Python environment should be created using requirements.txt in the home directory. Please run “pip install -r requirements.txt” as the first step. See [https://pip.pypa.io/en/stable/user\_guide/#ensuring-repeatability](https://pip.pypa.io/en/stable/user_guide/%23ensuring-repeatability) for further instructions on creating and using the requirements.txt file.

* master.py: a master file to execute all Python scripts in this subdirectory.
* 1\_parse\_xml.py: parsing the full text and metadata of each news article from XML files (executed on demo data).
* 2\_clean\_data.py: cleaning the parsed data and dropping duplicates (executed on demo data).
* 3\_match\_keywords.py: searching energy keywords in the full text to determine relevance (executed on demo data).
* 4\_extract\_regulatory\_sections.py: extracting “regulatory sections” from news articles (executed on demo data).
* 5\_quantify\_uncertainty.py: calculate uncertainty scores for each regulatory section and the full text of each article (executed on demo data).
* 6\_estimate\_uncertainty\_index.py: estimating the baseline oil regulatory uncertainty. index, the economic-adjusted oil regulatory uncertainty index, and the general oil supply uncertainty index (using pre-saved uncertainty scores in all\_uncertainty\_scores.csv).
* 7\_visualize\_indexes.py: producing Figure 1, Figure 3, Appendix C, and Appendix D in the paper, saved in the /output subdirectory.
* 8\_word\_clouds.py: producing Figure 2 and Figure 4 in the paper, saved in the /output subdirectory.

## **/empirical\_analysis**

* var\_energy2024\_baseline.do: producing baseline VAR analysis and Figures 5, 6, 7, and 8 in the paper.
* var\_energy2024\_robust.do: producing robustness checks using alternative VAR specifications in Appendices E and G.
* LPoil2024.do: producing robustness checks using local projections in Appendix H1.
* LPoil2024drill.do: producing robustness checks using local projections in Appendix H2.
* data\_2024.dta: data used in the baseline and robustness exercises (see the following table for a summary).

| ***Column Name*** | ***Description*** | ***Reference & Source*** |
| --- | --- | --- |
| oilregunc2024 | Oil regulatory uncertainty index (baseline) | Estimated in this paper |
| oilregunc2024econ | Oil regulatory uncertainty index (economic-adjusted) | Estimated in this paper |
| logrwti | Log U.S. crude oil prices deflated by CPI | U.S. Energy Information Administration |
| logstock | Log S&P 500 index | S&P Dow Jones Indices LLC |
| ffr | Federal funds effective rate | Board of Governors of the Federal Reserve System |
| logcpi | Log CPI | Bureau of Labor Statistics |
| logipm | Log industrial production | Board of Governors of the Federal Reserve System |
| logdrill | Log U.S. oil drilling | U.S. Energy Information Administration |
| logprodoil | Log U.S. oil production | U.S. Energy Information Administration |
| logworldprodoil | Log world oil production | U.S. Energy Information Administration |
| kiliamindex\_100 | Growth rate of world economic activity | Kilian (2009) |
| caur | California unemployment rate | U.S. Bureau of Labor Statistics |
| txur | Texas unemployment rate | U.S. Bureau of Labor Statistics |
| nyur | New York unemployment rate | U.S. Bureau of Labor Statistics |
| nmur | New Mexico unemployment rate | U.S. Bureau of Labor Statistics |
| dum | Dummy variable = 1 when the U.S. president is Republican | U.S. House of Representatives: History, Arts & Archives |
| jurado | Macroeconomic uncertainty | Jurado et al. (2015) |
| epu | EPU index | Baker et al. (2016) |
| cpu | CPU index | Gavriilidis (2021) |
| gpr | GPR index | Caldara and Iacoviello (2022) |

# **References**

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Caldara, D. and Iacoviello, M. (2022). Measuring geopolitical risk. *American Economic Review*, 112(4):1194–1225.

Gavriilidis, K. (2021). Measuring climate policy uncertainty. Available at SSRN 3847388.

Jurado, K., Ludvigson, S. C., and Ng, S. (2015). Measuring uncertainty. *American Economic Review*, 105(3):1177–1216.

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99(3):1053–69.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65.