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

Xiaohan Ma & Zhoudan Xie

## Overview

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).

## Data Availability and Provenance Statements

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 news articles analyzed. However, this repository provides the ProQuest IDs and some metadata for all 600,953 articles included in the baseline analysis (available in /measure\_uncertainty/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 repository includes demo data for five articles in XML format (located in /measure\_uncertainty/data/nlp\_demo/xml\_examples) along with the Python scripts for performing the textual analysis and estimating the uncertainty measures. The estimated uncertainty indexes using all articles are provided in the /measure\_uncertainty/data subdirectory.

Other economic data were obtained from publicly available data sources. See details in the Data Source section.

### 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.

### License for Data

The data are licensed under a Creative Commons/CC-BY-NC license.

### Summary of Availability

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

## Details on Datasets and Data Sources

There are two sets of data used in the paper. The first data set, stored in the /measure\_uncertainty/data subdirectory, contains data files used in the creation, analysis and visualization of the uncertainty measures. The second data set, stored in the /empirical\_analysis subdirectory, contains data used in the empirical analysis. Details on each data set are as follows, and a summary of all data and sources are available in Table D1.

### /measure\_uncertainty/data

This subdirectory contains the following data files:

* oil\_regulatory\_uncertainty\_index\_baseline.csv: estimated baseline oil regulatory uncertainty index (“RegUncertaintyIndex”).
* oil\_regulatory\_uncertainty\_index\_robust.csv: estimated alternative oil regulatory uncertainty indexes for robustness checks, including:
  + “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 in the paper.

**Table D1: Datasets and Data Sources**

| ***Data Name*** | ***Data File*** | ***Location*** | ***Provided*** | ***Citation*** |
| --- | --- | --- | --- | --- |
| Estimated oil regulatory uncertainty index (baseline) | oil\_regulatory\_uncertainty\_index\_baseline.csv | /measure\_uncertainty/data | TRUE | Created by authors |
| Estimated oil regulatory uncertainty indexes (for robustness) | oil\_regulatory\_uncertainty\_index\_robust.csv | /measure\_uncertainty/data | TRUE | Created by authors |
| Estimated oil supply uncertainty index | oil\_supply\_uncertainty\_index.csv | /measure\_uncertainty/data | TRUE | Created by authors |
| Calculated uncertainty scores for news articles | all\_uncertainty\_scores.csv | /measure\_uncertainty/data | TRUE | Created by authors |
| Noun chunk occurrences in regulatory sections | noun\_chunks\_by\_month\_reg.csv | /measure\_uncertainty/data | TRUE | Created by authors |
| Noun chunk occurrences in full articles | noun\_chunks\_by\_month\_general.csv | /measure\_uncertainty/data | TRUE | Created by authors |
| Demo news article text data | 282091173.xml  331720143.xml  391942977.xml  902573990.xml  2581635619.xml | /measure\_uncertainty/data/nlp\_demo/xml\_examples | TRUE | ProQuest |
| All news article text data | Not available | Not available | FALSE | ProQuest |
| Categorical EPU indexes | categorical\_epu\_index.xlsx | /measure\_uncertainty/data/supplementary\_data | TRUE | Baker et al. (2016) |
| CPU index | cpu\_index.csv | /measure\_uncertainty/data/supplementary\_data | TRUE | Gavriilidis (2021) |
| GPR index | gpr\_index.xls | /measure\_uncertainty/data/supplementary\_data | TRUE | Caldara and Iacoviello (2022) |
| EIA glossary | eia\_energy\_glossary.xlsx | /measure\_uncertainty/data/supplementary\_data | TRUE | EIA (2022) |
| Loughran and McDonald dictionary | lm\_sentiment.csv | /measure\_uncertainty/data/supplementary\_data | TRUE | Loughran and McDonald (2011) |
| Crosswalk of publication titles and newspaper names | pub\_title\_newspaper.csv | /measure\_uncertainty/data/supplementary\_data | TRUE | Created by authors |
| Fonts | coolvetica\_rg.otf  palatinolinotype\_roman.ttf | /measure\_uncertainty/data/supplementary\_data | TRUE | Dafont.com |
| Word cloud mask | wordcloud\_mask.png | /measure\_uncertainty/data/supplementary\_data | TRUE | Created by authors |
| Log U.S. crude oil prices deflated by CPI | data\_2024.dta | /empirical\_analysis | TRUE | EIA |
| Log S&P 500 index | data\_2024.dta | /empirical\_analysis | TRUE | S&P Dow Jones Indices LLC |
| Federal funds effective rate | data\_2024.dta | /empirical\_analysis | TRUE | Board of Governors of the Federal Reserve System |
| Log CPI | data\_2024.dta | /empirical\_analysis | TRUE | Bureau of Labor Statistics |
| Log industrial production | data\_2024.dta | /empirical\_analysis | TRUE | Board of Governors of the Federal Reserve System |
| Log U.S. oil drilling | data\_2024.dta | /empirical\_analysis | TRUE | EIA |
| Log U.S. oil production | data\_2024.dta | /empirical\_analysis | TRUE | EIA |
| Log world oil production | data\_2024.dta | /empirical\_analysis | TRUE | EIA |
| Growth rate of world economic activity | data\_2024.dta | /empirical\_analysis | TRUE | Kilian (2009) |
| California unemployment rate | data\_2024.dta | /empirical\_analysis | TRUE | U.S. Bureau of Labor Statistics |
| Texas unemployment rate | data\_2024.dta | /empirical\_analysis | TRUE | U.S. Bureau of Labor Statistics |
| New York unemployment rate | data\_2024.dta | /empirical\_analysis | TRUE | U.S. Bureau of Labor Statistics |
| New Mexico unemployment rate | data\_2024.dta | /empirical\_analysis | TRUE | U.S. Bureau of Labor Statistics |
| Political party dummy variable | data\_2024.dta | /empirical\_analysis | TRUE | U.S. House of Representatives: History, Arts & Archives |
| Macroeconomic uncertainty | data\_2024.dta | /empirical\_analysis | TRUE | Jurado et al. (2015) |
| EPU index | data\_2024.dta | /empirical\_analysis | TRUE | Baker et al. (2016) |
|  |  |  |  |  |

* 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 in the paper.
* /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 (EIA) (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.

### /empirical\_analysis

This subdirectory contains one data file, data\_2024.dta, that combines the estimated oil regulatory uncertainty indexes and economic data from publicly available data sources. This dataset is used in the baseline empirical analysis and robustness checks in the paper. Table D2 provides details on the data in data\_2024.dta.

**Table D2: Data in data\_2024.dta**

| ***Column Name*** | ***Description*** | ***Data Source*** |
| --- | --- | --- |
| “oilregunc2024” | Baseline oil regulatory uncertainty index | Created by authors (oil\_regulatory\_uncertainty\_index\_baseline.csv) |
| “oilregunc2024econ” | Economic-adjusted oil regulatory uncertainty index | Created by authors (oil\_regulatory\_uncertainty\_index\_robust.csv) |
| “oilregunc\_journal” | Journal-based oil regulatory uncertainty index | Created by authors (oil\_regulatory\_uncertainty\_index\_robust.csv) |
| “logrwti” | Log U.S. crude oil prices deflated by CPI | EIA |
| “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 | EIA |
| “logprodoil” | Log U.S. oil production | EIA |
| logworldprodoil | Log world oil production | EIA |
| “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) |

## Description of Programs/Code

The /measure\_uncertainty/python\_code subdirectory contains Python code to create, analyze, and visualize the uncertainty indexes. The /empirical\_analysis subdirectory contains Stata code used in the empirical analysis.

### /measure\_uncertainty/python\_code

* 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.
* 8\_word\_clouds.py: producing Figure 2 and Figure 4 in the paper.

A summary of the inputs and outputs for each Python script is available in Table D3.

**Table D3: Inputs and Outputs of Python Scripts**

|  |  |  |
| --- | --- | --- |
| ***Python Script*** | ***Input File and Location*** | ***Output File and Location*** |
| 1\_parse\_xml.py | All files in /measure\_uncertainty/data/nlp\_demo/xml\_examples | parsed\_xml.pkl (/measure\_uncertainty/data/nlp\_demo) |
| 2\_clean\_data.py | parsed\_xml.pkl (/measure\_uncertainty/data/nlp\_demo) | parsed\_xml\_clean.pkl (/measure\_uncertainty/data/nlp\_demo) |
| 3\_match\_keywords.py | parsed\_xml\_clean.pkl (/measure\_uncertainty/data/nlp\_demo) | parsed\_xml\_clean.pkl (/measure\_uncertainty/data/nlp\_demo) |
| 4\_extract\_regulatory\_sections.py | parsed\_xml\_clean.pkl (/measure\_uncertainty/data/nlp\_demo) | parsed\_xml\_clean.pkl (/measure\_uncertainty/data/nlp\_demo) |
| 5\_quantify\_uncertainty.py | parsed\_xml\_clean.pkl (/measure\_uncertainty/data/nlp\_demo) | parsed\_xml\_clean.pkl (/measure\_uncertainty/data/nlp\_demo) |
| 6\_estimate\_uncertainty\_index.py | all\_uncertainty\_scores.csv  (/measure\_uncertainty/data) | oil\_regulatory\_uncertainty\_index\_baseline.csv; oil\_supply\_uncertainty\_index.csv (/measure\_uncertainty/data) |
| 7\_visualize\_indexes.py | oil\_regulatory\_uncertainty\_index.csv; oil\_supply\_uncertainty\_index.csv  (/measure\_uncertainty/data) | figure1.jpg; figure3.jpg; appendixC.jpg; appendixD.jpg (/measure\_uncertainty/output) |
| 8\_word\_clouds.py | noun\_chunks\_by\_month\_reg.csv; noun\_chunks\_by\_month\_general.csv  (/measure\_uncertainty/data) | figure2.jpg; figure4.jpg (/measure\_uncertainty/output) |

### /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.

### License for Code

The code is licensed under a MIT license.

## List of Figures and Programs

The provided code reproduces:

* ☒ All figures in the paper.

**Table D4: All Figures and Corresponding Programs**

| ***Figure #*** | ***Program*** | ***Line #*** | ***Output File and Location*** |
| --- | --- | --- | --- |
| Figure 1 | 7\_visualize\_indexes.py | 46-133 | figure1.jpg (/measure\_uncertainty/output) |
| Figure 2 | 8\_word\_clouds.py | 114-127 | figure2.jpg (/measure\_uncertainty/output) |
| Figure 3 | 7\_visualize\_indexes.py | 149-222 | figure3.jpg (/measure\_uncertainty/output) |
| Figure 4 | 8\_word\_clouds.py | 169-182 | figure4.jpg (/measure\_uncertainty/output) |
| Figure 5 | var\_energy2024\_baseline.do | 14-31 | figure5.eps (/empirical\_analysis) |
| Figure 6 | var\_energy2024\_baseline.do | 14-31 | figure6.eps (/empirical\_analysis) |
| Figure 7 | var\_energy2024\_baseline.do | 34-44 | figure7a.eps & figure7b.eps (/empirical\_analysis/) |
| Figure 8 | var\_energy2024\_baseline.do | 47-59 | figure8.eps (/empirical\_analysis) |
| Appendix C | 7\_visualize\_indexes.py | 237-293 | appendixC.jpg (/measure\_uncertainty/output) |
| Appendix D | 7\_visualize\_indexes.py | 364-441 | appendixD.jpg (/measure\_uncertainty/output) |
| Appendix E.1.1 | var\_energy2024\_robust.do | 125-136 | figureE11.eps (/empirical\_analysis) |
| Appendix E.1.2 | var\_energy2024\_robust.do | 125-136 | figureE12.eps (/empirical\_analysis) |
| Appendix E.2.1 | var\_energy2024\_robust.do | 12-24 | figureE21.eps (/empirical\_analysis) |
| Appendix E.2.2 | var\_energy2024\_robust.do | 12-24 | figureE22.eps (/empirical\_analysis) |
| Appendix G.1.1 | var\_energy2024\_robust.do | 27-38 | figureG11.eps (/empirical\_analysis) |
| Appendix G.1.2 | var\_energy2024\_robust.do | 27-38 | figureG12.eps (/empirical\_analysis) |
| Appendix G.2.1 | var\_energy2024\_robust.do | 41-52 | figureG21.eps (/empirical\_analysis) |
| Appendix G.2.2 | var\_energy2024\_robust.do | 41-52 | figureG22.eps (/empirical\_analysis) |
| Appendix G.3.1 | var\_energy2024\_robust.do | 55-66 | figureG31.eps (/empirical\_analysis) |
| Appendix G.3.2 | var\_energy2024\_robust.do | 55-66 | figureG32.eps (/empirical\_analysis) |
| Appendix G.4.1 | var\_energy2024\_robust.do | 69-80 | figureG41.eps (/empirical\_analysis) |
| Appendix G.4.2 | var\_energy2024\_robust.do | 69-80 | figureG42.eps (/empirical\_analysis) |
| Appendix G.5.1 | var\_energy2024\_robust.do | 83-94 | figureG51.eps (/empirical\_analysis) |
| Appendix G.5.2 | var\_energy2024\_robust.do | 83-94 | figureG52.eps (/empirical\_analysis) |
| Appendix G.6.1 | var\_energy2024\_robust.do | 97-108 | figureG61.eps (/empirical\_analysis) |
| Appendix G.6.2 | var\_energy2024\_robust.do | 97-108 | figureG62.eps (/empirical\_analysis) |
| Appendix G.7.1 | var\_energy2024\_robust.do | 111-122 | figureG71.eps (/empirical\_analysis) |
| Appendix G.7.2 | var\_energy2024\_robust.do | 111-122 | figureG72.eps (/empirical\_analysis) |
| Appendix H.1 | LPoil2024.do | all lines | figureH1.eps (/empirical\_analysis) |
| Appendix H.2 | LPoil2024drill.do | all lines | figureH2.eps (/empirical\_analysis) |

## Computational Requirements

### Software Requirements

* ☒ The replication package contains one or more programs to install all dependencies and set up the necessary directory structure.
* Python 3.10+
  + 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.
* Stata (code was last run with version 18)
  + var (as of April 2023)

### Memory, Runtime, Storage Requirements

Approximate time needed to reproduce the analyses on a standard 2025 desktop machine:

* ☒ 1-3 days

Approximate storage space needed:

* ☒ 25 MB - 250 MB

The code was last run on a **11th Gen Intel(R) Core(TM) i7-11700 @ 2.50GHz with 16 GB RAM**.

Note: The original analysis of the full text data of all news articles was run on a **16 vCPU Amazon EC2 instance with 64 GB RAM and 70 GB storage**, consuming 2-8 hours.

## Instructions to Replicators

* Download the replication package.
* Set up the working environment (see instructions in “Computational Requirements”).
* Run /measure\_uncertainty/python\_code/master.py to execute all Python scripts in the subdirectory and reproduce the output files, as outlined in Table D4.
* Run Stata code (.do files) in the /empirical\_analysis subdirectory to reproduce results of the empirical analysis.

## 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.

## Acknowledgements

The README template recommended by the journal was used to generate this document:

Lars Vilhuber, Connolly, M., Koren, M., Llull, J., & Morrow, P. (2022). A template README

for social science replication packages. Social Science Data Editors. <https://social-science-data-editors.github.io/template_README/>.