README – Replication Materials for *Warning words in a warming world: Central bank communication and climate change*

Emanuele Campiglio[a,b,c], Jérôme Deyris[d], Davide Romelli[e], and Ginevra Scalisi[f]

[a] *University of Bologna, Department of Economics, Italy*
[b] *CMCC Foundation - Euro-Mediterranean Center on Climate Change, Italy*
[c] *LSE Grantham Research Institute on Climate Change and the Environment, United Kingdom*
[d] *Sciences Po Paris, Centre d'Études Européennes, France*
[e] *Trinity College Dublin, Ireland*
[f] *Scuola Superiore Sant'Anna, Italy*

Introduction

This archive contains the replication files for the paper ***Warning words in a warming world: Central bank communication and climate change*** by Emanuele Campiglio, Jérôme Deyris, Davide Romelli, and Ginevra Scalisi, *European Economic Review*.

The replication material can be downloaded from: www.cbspeeches.com.

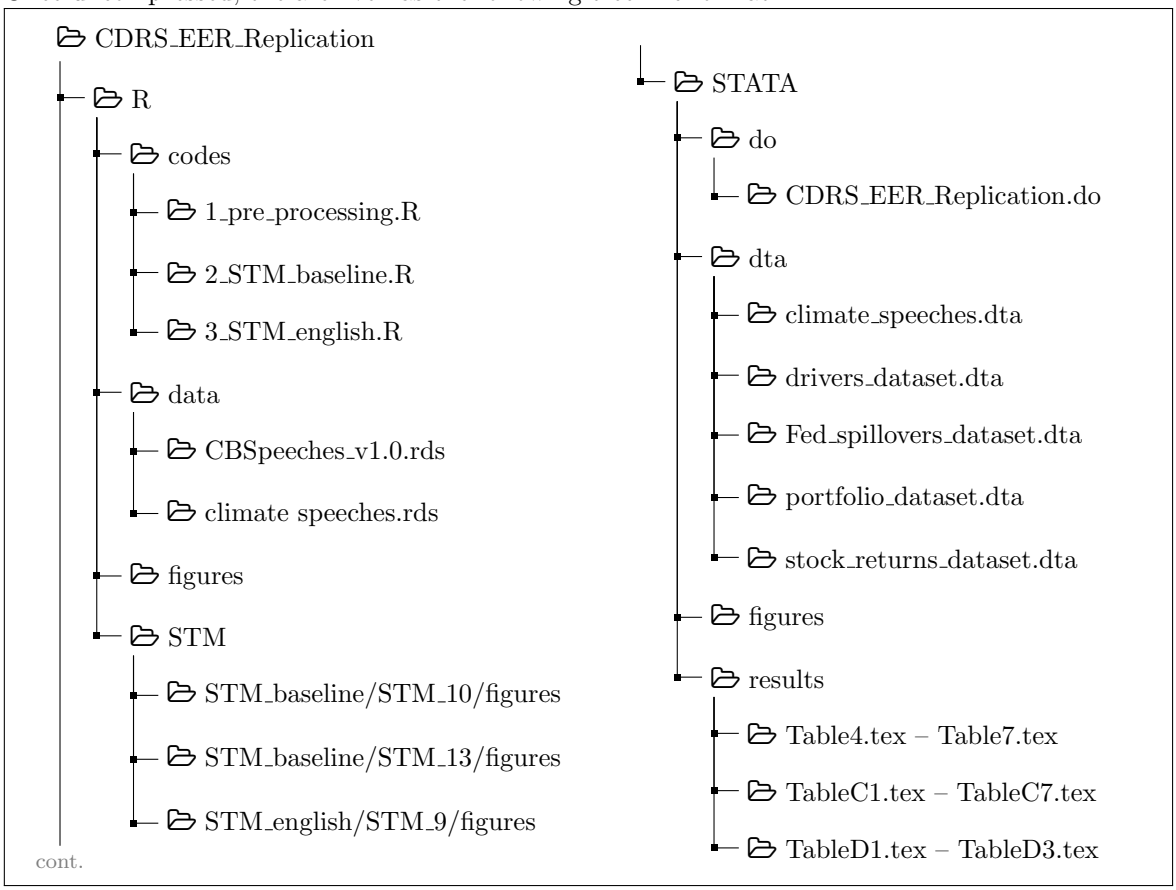
The archive contains all the code and non-proprietary data necessary to replicate the main results, figures, and tables presented in the paper. Specifically, it includes:

- The corpus of central bank speeches.
- Scripts (in R) for analysing and pre-processing the corpus.
- Scripts (in R) to run the Structural Topic Model (STM).
- Scripts (in Stata) to run the empirical analysis.
- Summary datasets to be used in the regressions.

Please note that the archive **does not include the firm-level balance sheet data and the measures of firms' 'greenness'**, as these are based on proprietary commercial databases (LSEG, formerly Refinitiv) and cannot be shared publicly.

Data structure and content

Once uncompressed, the archive has the following tree-like format:



- The `R/data` folder contains the text of central bank speeches
  - `R/data/CBSpeeches_v1.0.rds`: Contains the full corpus of central bank speeches.
  - `R/data/climate_speeches.rds`: Contains the subset of climate-related speeches (using an identification threshold of one climate-related keyword).
- The `R/codes` folder contains the codes to estimate the Structural Topic Model (STM)
  - `R/codes/1_pre_processing.R`: Prepares the speech corpus for the STM estimation.
  - `R/codes/2_STM_baseline.R`: Estimates the STM model used in the baseline analysis.
  - `R/codes/3_STM_english.R`: Estimates the STM model run on the subset of English-language speeches.

- The `R/figures/` folder stores the figures generated by the analysis of the dataset. Figures will appear here once the `R/codes/1_pre_processing.R` code will be run.
- The `R/STM/` folder stores the results from the STM estimations. Charts and results will appear here once the STM models are run.
  - `R/STM_baseline/STM_10/figures`: Contains the results for the baseline STM using 10 topics (main specification).
  - `R/STM_baseline/STM_10/figures`: Contains the results for the baseline STM using 13 topics (robustness check).
  - `R/STM_english/STM_9/figures`: Contains the results for the STM run using only English speeches (no translations).
- The `R/tables/` folder stores the tables generated by the analysis of the dataset. The folder and the tables within will be created once the `R/codes/1_pre_processing.R` code is run.
- The `STATA/do` folder contains the file to replicate the figures and tables of the empirical analysis
  - `STATA/do/CDRS_EER_Replication.do`: STATA do-file for running the empirical analysis.
- The `STATA/dta` folder stores all the data files needed to run the empirical analysis
  - `STATA/dta/climate_speeches.dta`: Speech-level dataset with topic scores and metadata.
  - `STATA/dta/drivers_dataset.dta`: Dataset for analysing the determinants of climate communication.
  - `STATA/dta/Fed_spillovers_dataset.dta`: Dataset used to assess international spillovers of Fed communication.
  - `STATA/dta/portfolio_dataset.dta`: Industry-level portfolio returns for green-minus-dirty portfolios.
  - `STATA/dta/stock_returns_dataset.dta`: Firm-level return dataset.
- The `STATA/figures/` folder contains the generated figures used in the paper and appendices.
- The `STATA/results/` folder contains `.tex` files with the scripts used to reproduce the tables included in the paper and appendices.[1]
  - `STATA/results/Table4.tex -- Table7.tex`: Main regression tables.

[1] To generate or update these tables, the Stata replication file `CDRS_EER_Replication.do` must be executed. The `.tex` files are dynamically created during that process, meaning that they recall the output of the Stata code.

- `STATA/results/TableC1.tex -- TableC7.tex`: Robustness checks and extensions (Appendix C).
- `STATA/results/TableD1.tex -- TableD3.tex`: Additional robustness (Appendix D).

Instructions for replication

1. Download the replication package, uncompress it, and store it in a folder of your choice.
2. Open the `R/R.proj` file. This will automatically set the working directory. Alternatively, make sure to manually specify the path to your working directory at the top of the R scripts. This is required for the code to locate the data and output folders correctly.
3. Use the R scripts in `R/codes` to pre-process the speech corpus and estimate the baseline STM topic model:
   - Run `R/codes/1_pre_processing.R` to prepare the speech corpus for the STM estimation.
   - Run `R/codes/2_STM_baseline.R` to estimate the baseline STM.[2]
   - Run `R/codes/3_STM_english.R` to estimate the robustness STM using English-only speeches.
4. Use the STATA do-file `CDRS_EER_Replication.do` to reproduce the figures and tables of the empirical analysis:
   - Run `CDRS_EER_Replication.do` to perform the empirical analysis. At the top of the script, it is possible to define four locals that act as on/off switches for different sections of the replication, corresponding to different sections of the paper. Set local to 1 to run that section; set to 0 to skip it.
   - The resulting tables are stored in the `results/` folder, while the corresponding graphics are saved in the respective `figures/` directories.

Citation

If using any of the material, please cite the paper as:

Campiglio, E., Deyris, J., Romelli, D., & Scalisi, G. (2025). Warning words in a warming world: Central bank communication and climate change. European Economic Review.

[2]The code is set to run the STM with a range of topics going from 5 to 50, so as to identify the best model among them. This process can take several hours. To reduce the running time, modify the command `topic_range <- 5:50` in Section 2.2 of the script, and select the desired range of topics. The baseline model we use in the paper has 10 topics; we also use a 13-topic model as a robustness check.