Reproducible Package for the Paper 'Impact of Temperature Uncertainty on Firm Growth: A Grid-Level Analysis

Overview

This package provides a reproducible analysis workflow for the paper "Impact of Temperature Uncertainty on Firm Growth: A Grid-Level Analysis." Below is the structure of the project, with each folder and file briefly explained.

Repository Structure

The package contains three main subfolders:

```
├── README.PDF                      # Overview of the package, explaining
purpose, usage, and setup
├── 1.Weather data construction  # Folder for weather data processing and
conversion
│   ├── code
│   │   └── Weather_data_processing_WB_vetting_final(spline).R
│   ├── final data (included in the package)
│   └── source data
├── 2.Firm-level data construction and merging  # Folder for firm data
processing and merging with weather data
│   ├── code
│   │   ├── 1.Location_matching_indexing_WB_vetting.R
│   │   ├── 2.Weather_and_Location_Merging_max_WB_vetting_new(spline).R
│   │   ├── 3.Data_processing_ver3_WB_vetting_new_spline.R
│   │   └── 4.Data_presave.R
│   └── source data
└── 3.Analysis                      # Folder for data analysis and figure/table
generation
    ├── code
    │   └── 1.Data_exploration_final_WB_vetting_ver_final_spline.R
    ├── data
    │   ├── Location_Weather_merged_mean(Oct10)_max_spine.csv (included)
    │   ├── orbis_summ_nace.csv (only structure included)
    │   └── orbis_summ.csv (only structure included)
    └── output_temp
```

Directory Details

The package consists of three main sections, described below:

1. Weather Data Construction (`1.Weather data construction`)
- This step starts with the raw BERKEARTH data in `.nc` format, stored in `source data`.
- The R script `Weather_data_processing_WB_vetting_final(spline).R` converts the `.nc` data to `.csv` format and further processes it, saving the final data (`max_1by1_result_new(spline_Oct10.csv)`) in `final data`.

2. Firm-Level Data Construction and Merging (`2.Firm-level data construction and merging`)
- The script `1.Location_matching_indexing_WB_vetting.R` collects and processes missing firm location information, creating location files such as `1-Location_city_merged_Google.csv`, `2-Duplicates_Summary.csv`, and `3-Location_city_merged_Google_no_duplicates_index.csv`. It also creates intermediary indexing datasets (`index1.csv`, `index2.csv`, and `index3.csv`) to be used for merging the weather and location datasets.
- The script `2.Weather_and_Location_Merging_max_WB_vetting_new(spline).R` merges the processed location data with weather data, resulting in the dataset `Location_Weather_merged_mean(Oct10)_max_spine.csv`, which is saved in the data folder of the `3.Analysis` section.
- `3.Data_processing_ver3_WB_vetting_new_spline.R` loads raw Orbis data, cleans it, and merges it with the location-weather data. The resulting dataset includes firm financial metrics, location information, and corresponding temperature metrics, providing the final dataset used for the analysis.
- The `4.Data_presave.R` script generates aggregated datasets (`orbis_summ.csv` and `orbis_summ_nace.csv`) for reproducibility vetting for the main figures and tables, as the original firm-level data cannot be included due to proprietary restrictions.

Note: Firm-level Orbis data files (e.g., `Orbis_dat_mlvl_new.csv`, `industry_code_m.csv`, `industry_code.csv`) are proprietary and not included in this package, therefore this part of the process won't run.

3. Analysis (`3.Analysis`)
- The script `1.Data_exploration_final_WB_vetting_ver_final_spline.R` produces all tables and figures in the paper based on the intermediate data files produced in sections 1 and 2 (`orbis_summ.csv`, `orbis_summ_nace.csv`, and `Location_Weather_merged_mean(Oct10)_max_spine.csv`). All output files are saved in the `output_temp` folder.

Note: The final processed Orbis datasets used for generating the final firm-level data (e.g., `orbis_cut_final_max_new_spine(Oct10).csv`) are not included in the package. In addition, according to Orbis license intermediate data can't be share so this part of the code won't run either. This was only shared with the PRWP verification team for replication purposes but it's not included in the package.

Structure `orbis_summ.csv` (2001-2019)

COUNTRY Index CLOSDATE_year mean variance std uncertainty1 uncertainty2 q100 FIAS_growth_log_mean FIAS_growth_log_med CAP_tan_mean CAP_tan_mean_log Income1 n Index_numeric (structure is saved in the folder 3.Analysis/data for reference on what to consult in Orbis data)

Structure `orbis_summ_nace.csv` (2001-2019)

nace2_main_section Index CLOSDATE_year mean variance std uncertainty1 uncertainty2 q100 FIAS_growth_log_mean FIAS_growth_log_med CAP_tan_mean CAP_tan_mean_log Income1 n Index_numeric (structure is saved in the folder 3.Analysis/data for reference on what to consult in Orbis data)

Data Availability Statement

This project involves five types of data:

1. Temperature Gridded Data
- **Source:** Berkeley Earth Foundation dataset (BERKEARTH), via the Climate Data Store (Copernicus Climate Change Service).

- **Link:** https://cds.climate.copernicus.eu/datasets/insitu-gridded-observations-global-and-regional?tab=download

- **Details:** Original `.nc` files (2000–2019) were converted to `.csv` format for processing. Final data saved as `max_1by1_result_new(spline_Oct10).csv` in final data.

2. Firm-Level Data (not shared)
- **Source:** Orbis data, via Bureau van Dijk (WRDS license, University of Waterloo).

- **Details:** Proprietary data; cannot be shared.

3. Location Data
- **Details:** Missing firm locations filled using the Google Maps API. Final dataset: `3-Location_city_merged_Google_no_duplicates_index.csv`.

4. Miscellaneous Data

- **GDP Deflator:** World Bank (Link).

- **Income Levels:** World Bank classifications (Link).

5. Merged Grid-Level Dataset

- **Details:** Aggregated datasets (`orbis_summ.csv`, `orbis_summ_nace.csv`) used for analysis. Firm-level data cannot be shared due to confidentiality.