Author Contributions Checklist Form

This form documents the artifacts associated with the article (i.e., the data and code supporting the computational findings) and describes how to reproduce the findings.

Part 1: Data

□ This paper **does not** involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

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

Abstract

FRED-MD is a large macroeconomic database designed for the empirical analysis of "big data." The datasets of monthly and quarterly observations mimic the coverage of datasets already used in the literature, but they add three appealing features. They are updated in real-time through the FRED database. They are publicly accessible, facilitating the replication of empirical work.

Availability

☒ Data **are** publicly available
□ Data **cannot be made** publicly available

If the data are publicly available, see the *Publicly available data* section. Otherwise, see the *Non-publicly available data* section, below.

Publicly available data

☒ Data are available online at: https://research.stlouisfed.org/econ/mccracken/fred-databases/

□ Data are available as part of the paper's supplementary material.
□ Data are publicly available by request, following the process described here:

□ Data are or will be made available through some other mechanism, described here:

Non-publicly available data

Discussion of lack of publicly available data:

Description

File format(s)

☒ CSV or other plain text: current.csv
☐ Software-specific binary format (.Rda, Python pickle, etc.):
☐ Standardized binary format (e.g., netCDF, HDF5, etc.):
☐ Other (described here):

Data dictionary

☒ Provided by the authors in the following file(s): reproduction_code/Real_Data/Pre-Process/current.csv
☐ Data file(s) is (are) self-describiing (e.g., netCDF files)
☐ Available at the following URL:

Additional information (optional)

Part 2: Code

Abstract

We provide the code for reproduce our simulation and real data application. See workflow for more details.

Description

Code format(s)

☒ Script files
- ☒ R ☐ Python ☐ Matlab
- ☐ Other:

☐ Package
- ☐ R ☐ Python ☐ MATLAB toolbox
- ☐ Other:

☐ Reproducible report
- ☐ R Markdown ☐ Jupyter notebook
- ☐ Other:

☐ Shell script

☐ Other (described here):

Supporting software requirements

Version of primary software used

Rstudio 1.2.5001

Libraries and dependencies used by the code

Glmnetc (4.1-3), SIS (0.8-8), MASS (7.3-54)

Supporting system/hardware requirements (optional)

Parallelization used

☒ No parallel code used
☐ Multi-core parallelization on a single machine/node
Number of cores used:
☐ Multi-machine/multi-node parallelization
Number of nodes and cores used:

License

☒ MIT License (default)
☐ BSD
☐ GPL v3.0
☐ Creative Commons
☐ Other (described here):

Additional information (optional)

Part 3: Reproducibility workflow

Scope

The provided workflow reproduces:
☐ Any numbers provided in text in the paper
☒ The computational method(s) presented in the paper (i.e., code is provided that implements the method(s))
☒ All tables and figures in the paper
☐ Selected tables and figures in the paper, as explained and justified here:

Workflow details

Location

The workflow is available:
☐ As part of the paper's supplementary material
☒ In this Git repository: https://github.com/qooyqpqy123/FARM_Reproduction/tree/main/reproduction_code
☐ Other:

Format(s)

☐ Single master code file
☐ Wrapper (shell) script(s)
☐ Self-contained R Markdown file, Jupyter notebook, or other literate programming approach
☐ Text file (e.g., a readme-style file) that documents workflow
☐ Makefile
☒ Other (more detail in 'Instructions' below)

Instructions

The folder named `Simulation', reproduces all methodology of simulation results in the paper. Including Figure 1, Figure 2, Table 1, Table 2 and Prediction Results of Table 1-4 in Section B of Appendix.

To be more specific, we reproduces all three figures in Figure 1 using the code inside 'Gaussian', Uniform' and 'Heavy-tail' respectively in folder 'Simulation/Consistency_Figure 1", respectively.

As for figure 2, we reproduce it using the file inside 'Simulation/Figure2"

For Table 1 and Table 2, we reproduce it using the file under 'Simulation/PCR_adequate_Table1' and 'Simulation/Sparse_Adequate_Table2', respectively. The Gaussian(or Uniform)_indp means the data are generated with independent measurements. In addition, Gaussian(or Uniform)_mix means the data are generated with dependent measurements. These match with the two settings in our paper.

For Prediction Results, we reproduce it using the file under the folder of 'Simulation/ Prediction_Section_B.1" There contains codes for reproduce Table 1-4 in Appendix B.

As for real data. We provide the reproduction code for it under 'Real_Data'.

To be more specific, for pre-processing, we put the original data as well as the code under 'Real_Data/Pre-process'.

We then reproduce Prediction results (Table 3 and Plot 7) using the code under 'Real_Data/ Prediction_Table3_Plot7'

Finally, we justify the results in Table 4 using code in 'Real_Data/PCR_Adequate_Table4' and 'Real_Data/Sparse_Adequate_Table4'.

Expected run-time

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

□ <1 minute
□ 1-10 minutes
□ 10-60 minutes
□ 1-8 hours
☒ >8 hours
□ Not feasible to run on a desktop machine, as described here:

Additional documentation (optional)

Notes (optional)