hd-var: Lasso and Related Tools for High-Dimensional Vector Autoregression

Maintained by: Jesper Riis-Vestergaard Sørensen, University of Copenhagen

---

Table of Contents

Overview

This repository contains an `R` implementation of (weighted) Lasso and its variations [such as least squares refitting following Lasso selection (post-Lasso) and square-root Lasso (sqrt-Lasso)] with a data-driven and theoretically justifiable tuning parameter selection method designed for high-dimensional (HD) vector autoregression (VAR).

The code in this repository was developed for and used in the paper: "***Data-Driven Tuning Parameter Selection for High-Dimensional Vector Autoregressions***," authored by Anders Bredahl Kock, Rasmus Søndergaard Pedersen, and Jesper Riis-Vestergaard Sørensen (henceforth: KPS), available at arXiv:2403.06657. See the paper for the theoretical justification of the method.

Prerequisites

To run the scripts in this repository, you will need the following:

- **R**: Version 4.0 or higher. Key R Packages: `glmnet` and `ggplot2`.
- **Matlab**: Required only for pre-processing the FRED-MD dataset (optional).

Architecture

The repository is organized as follows:

- **Root Directory**: Contains the main scripts and auxiliary functions for fitting VAR models using Lasso and its variations.
- **Simulations Subfolder**: Contains scripts and data for running simulations and generating figures for the main paper and its supplementary appendices.

- **Application/FRED Subfolder**: Contains scripts and data for the empirical illustration using the FRED-MD dataset.

File Descriptions

Root Directory

- `lassoVAR.R`: Fits a VAR using (weighted) Lasso with optional refitting (post-Lasso) using the KPS tuning parameter selection method.
- `sqrtLassoVAR.R`: Fits a VAR using (weighted) square-root Lasso using the KPS tuning parameter selection method.
- `icLassoVAR.R`: Fits a VAR using Lasso with an information criterion (Akaike, Bayes, or Hannan-Quinn) to determine the penalty level.
- `helper_functions.R`: Contains helper functions for data unpacking, least squares fitting, and Lasso application.

Simulations Subfolder

- `simData.R`: Functions for simulating data based on designs in Section 6 of the paper.
- `runSim_v03.R`: Runs the simulations reported in Section 6.
- `simulations_workspace_1000_MC_100_to_1000_n_16_to_128_p_with_num_upd.RData`: Workspace produced by `runSim_v03.R`.
- `createFigs_v07.R`: Produces Figure 6.1 and Figure H.1 (relative and raw average estimation errors, respectively).
- `markupDependence_v01.R`: Runs simulations for mark-up dependence (Figure H.2).
- `markup_dependence_workspace_1000_MC_200_to_1000_n_16_to_128_p_diagonal_only.RData`: Workspace produced by `markupDependence_v01.R`.
- `markupDependenceFigs_v02.R`: Produces Figure H.2 (raw average estimation errors).

Application/FRED Subfolder

- `FREDMD_preprocess.m`: Pre-processes the raw FRED-MD dataset.
- `prepare_missing.m`: Handles missing data in the FRED-MD dataset.
- `remove_outliers.m`: Removes outliers from the FRED-MD dataset.
- `FRED-MD_2022-05_preprocessed.csv`: Pre-processed FRED-MD dataset used in the empirical illustration.
- `FRED-MD_forecasting_v02.R`: Conducts the forecasting exercise in Section 7.
- `application_workspace_N_120_qmax_12_with_num_upd.RData`: Workspace produced by `FRED-MD_forecasting_v02.R`.
- `FRED-MD_figures_v02.R`: Produces Figure 7.1 (average and 95th percentile inverse variance weighted square forecast error).

Reproducibility Workflow

Simulations

1. **Run the main simulations**:

- Execute `runSim_v03.R` to run the simulations described in Section 6. (Execute from the root directory. Use `setwd("..")` to back up, if necessary.)
- Save the workspace manually if not using Linux (via `save.image(...)`).

2. **Run the mark-up dependence simulations**:
   - Execute `markupDependence_v01.R` (from the root) to run the Supplementary Appendix H simulations investigating mark-up dependence.
3. **Generate figures**:
   - Run `createFigs_v07.R` to produce Figure 6.1 and Figure H.1.
   - Run `markupDependenceFigs_v02.R` to produce Figure H.2.

Empirical Illustration

1. **(Optional) Pre-process the raw FRED-MD dataset (using Matlab)**:
   - Run `FREDMD_preprocess.m`, which calls upon `prepare_missing.m`, and `remove_outliers.m` in sequence.
2. **Conduct the forecasting exercise**:
   - Execute `FRED-MD_forecasting_v02.R`. (Execute from the root directory. Use `setwd("../..")` to back up, if necessary.)
   - The script will load the pre-processed data from `application/FRED/data` and save the workspace automatically (if in Linux).
3. **Generate figures**: Run `FRED-MD_figures_v02.R` to produce both parts of Figure 7.1.

Figures and Tables

The following figures are reproduced by the scripts in this repository:

- **Figure 6.1**: Relative average estimation errors (produced by `createFigs_v07.R`).
- **Figure H.1**: Raw average estimation errors (produced by `createFigs_v07.R`).
- **Figure H.2**: Mark-up dependence (produced by `markupDependenceFigs_v02.R`).
- **Figure 7.1**: Average and 95th percentile inverse variance weighted square forecast error (produced by `FRED-MD_figures_v02.R`).

Citation

If you use this repository, please cite:

- Kock, A. B., Pedersen, R. S., & Sørensen, J. R.-V. (2024). "Data-Driven Tuning Parameter Selection for High-Dimensional Vector Autoregressions." [arXiv:2403.06657](https://arxiv.org/abs/2403.06657).
