

README: Replication Package for Estimation and Inference in High-Dimensional Panel Data Models with Interactive Fixed Effects

Maximilian Rucker¹
Ulm University

Oliver Linton³
University of Cambridge

Michael Vogt²
Ulm University

Christopher Walsh⁴
Newcastle University

Overview

This replication package allows researchers to reproduce the Simulation Study (Supplementary material) and the Empirical Study (Section 8) from the paper "Estimation and Inference in High-Dimensional Panel Data Models with Interactive Fixed Effects" in R. The replication package contains two main folders *MonteCarloSimulations* and *EmpiricalApplication* which accommodate the files `monteCarlo_main.R` and `empiricalApplication_main.R` respectively. The `[...].R` files generate all plots and tables presented in the paper.

Data Availability and Provenance Statements

Summary of Availability

- All data (*imputed_data.csv*) are publicly available.

Public use data

The *imputed_data.csv* data used to support the findings of empirical application have been deposited in the GitHub repository <https://github.com/RueckerM/hdcce-ReplicationFiles> and is

¹Address: Institute of Statistics, Department of Mathematics and Economics, Ulm University, Helmholtzstrasse 20, 89081 Ulm, Germany. Email: maximilian.ruecker@uni-ulm.de.

²Corresponding author. Address: Institute of Statistics, Department of Mathematics and Economics, Ulm University, Helmholtzstrasse 20, 89081 Ulm, Germany. Email: m.vogt@uni-ulm.de.

³Address: Faculty of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD, UK. Email: ob120@cam.ac.uk.

⁴Address: Newcastle University Business School, 5 Barrack Road, Newcastle upon Tyne, NE1 4SE, UK. Email: chris.walsh@newcastle.ac.uk.

publicly available. The data were collected and kindly provided by Shaoran Li.

Data description

We compile a comprehensive set of 102 stock-level characteristics from Green et al. (2017), utilizing the SAS code provided by the authors while extending the sample period to March 2022. The sample is aligned in calendar time, with characteristics calculated at the end of month $t - 1$. Annual accounting data is incorporated if the firm's fiscal year ended at least six months prior, while quarterly data is included if the fiscal quarter ended at least four months prior. The data from I/B/E/S and CRSP are synchronized using the I/B/E/S statistical period date and CRSP's monthly or daily end date.

Computational requirements

Software Requirements

- The main files contain headers to install all dependencies and set up the necessary directory structure.
- R 4.4.1 Race For Your Life
 - `this.path` (2.7.0)
 - `glmnet` (4.1.8)
 - `matrixStats` (1.5.0)
 - `mvtnorm` (1.3-1)
 - `parallel` (1.0.17)
 - `doRNG` (1.8.6.2)
 - `doParallel` (1.0.17)
 - `latex2exp` (0.9.6)
 - the dependencies are set as a header to the main files (need to remove the commands for the corresponding lines)

Controlled Randomness

- Random seed is set at line 64 of program `monteCarlo_main.R`.

Memory, Runtime, Storage Requirements

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

- 3-14 days

Approximate storage space needed:

- 25 MB - 250 MB

Details The code was last run on a **8-core Intel Core i7-10510U CPU 1.80GHz based laptop with Linux Ubuntu 22.04.5 and 16 GB RAM.**

Description of programs/code

- The `monteCarlo_main.R` will generate all results from the Simulation Study. The Figures are stored in *MonteCarloSimulations\plots* and Tables in *MonteCarloSimulations\tables*. The script takes around ten days on the 8-core Intel-based machine. In order to reproduce the exact results, 1000 bootstrap simulations have to be set in line 1653 and following.
 - The files `hdcce_estimator.R` and `hdcce_inference.R` are helper files for the main script `monteCarlo_main.R` and sourced therein. For given data, the functions return the estimated coefficients and confidence bounds respectively.
- The `empiricalApplication_main.R` will generate all results from the Empirical Application. The outputs are stored in *EmpiricalApplication\plots*. The script runs in approximately two minutes.

List of tables and programs

The provided code reproduces:

- All numbers provided in text in the paper
- All tables and figures in the paper

Figure/Table	Program	Line	Output file	Note
Figure S.1	monteCarlo_main.R	673	<i>FigS1\T_dim15gamma1HAC1p7.pdf</i>	L-T panel
Figure S.1	monteCarlo_main.R	744	<i>FigS1\T_dim15gamma1HAC1p10.pdf</i>	L-M panel
Figure S.1	monteCarlo_main.R	798	<i>FigS1\T_dim15gamma1HAC1p13.pdf</i>	L-B panel
Figure S.1	monteCarlo_main.R	861	<i>FigS1\T_dim50gamma1HAC1p16.pdf</i>	R-T panel
Figure S.1	monteCarlo_main.R	945	<i>FigS1\T_dim50gamma1HAC1p31.pdf</i>	R-M panel
Figure S.1	monteCarlo_main.R	1000	<i>FigS1\T_dim50gamma1HAC1p46.pdf</i>	R-B panel
Figure S.2	monteCarlo_main.R	699	<i>FigS2\T_dim15gamma1HAC1p31.pdf</i>	L-T panel
Figure S.2	monteCarlo_main.R	771	<i>FigS2\T_dim15gamma1HAC1p151.pdf</i>	L-M panel
Figure S.2	monteCarlo_main.R	823	<i>FigS2\T_dim15gamma1HAC1p301.pdf</i>	L-B panel
Figure S.2	monteCarlo_main.R	899	<i>FigS2\T_dim50gamma1HAC1p91.pdf</i>	R-T panel
Figure S.2	monteCarlo_main.R	975	<i>FigS2\T_dim50gamma1HAC1p451.pdf</i>	R-M panel
Figure S.2	monteCarlo_main.R	1026	<i>FigS2\T_dim50gamma1HAC1p901.pdf</i>	R-B panel
Figure S.3	monteCarlo_main.R	722	<i>FigS3\T_dim50gamma1HAC1p901.pdf</i>	L panel
Figure S.3	monteCarlo_main.R	923	<i>FigS3\T_dim50gamma1HAC1p3001.pdf</i>	R panel
Figure S.4	monteCarlo_main.R	1266	<i>FigS4\[/.../T_dim15gamma1HAC1p13.pdf</i>	L-T panel
Figure S.4	monteCarlo_main.R	1290	<i>FigS4\[/.../T_dim15gamma1HAC1p151.pdf.pdf</i>	L-M panel
Figure S.4	monteCarlo_main.R	1246	<i>FigS4\[/.../T_dim15gamma1HAC1p901.pdf</i>	L-B panel
Figure S.4	monteCarlo_main.R	1326	<i>FigS4\[/.../T_dim15gamma1HAC1p46.pdf</i>	R-T panel
Figure S.4	monteCarlo_main.R	1350	<i>FigS4\[/.../T_dim15gamma1HAC1p451.pdf</i>	R-M panel
Figure S.4	monteCarlo_main.R	1382	<i>FigS4\[/.../T_dim15gamma1HAC1p3001.pdf</i>	R-B panel
Figure S.5	monteCarlo_main.R	1509	<i>FigS5\[/.../p13.pdf</i>	L-T panel
Figure S.5	monteCarlo_main.R	1533	<i>FigS5\[/.../p151.pdf.pdf</i>	L-M panel
Figure S.5	monteCarlo_main.R	1489	<i>FigS5\[/.../p901.pdf</i>	L-B panel
Figure S.5	monteCarlo_main.R	1569	<i>FigS5\[/.../p46.pdf</i>	R-T panel
Figure S.5	monteCarlo_main.R	1593	<i>FigS5\[/.../p451.pdf</i>	R-M panel
Figure S.5	monteCarlo_main.R	1625	<i>FigS5\[/.../p3001.pdf</i>	R-B panel
Table S.3	monteCarlo_main.R	1111	<i>InferenceT_dim15gamma1HAC1.csv</i>	L Panel
Table S.3	monteCarlo_main.R	1111	<i>InferenceT_dim50gamma1HAC1.csv</i>	R Panel
Table S.4	monteCarlo_main.R	1111	<i>InferenceT_dim15gamma1HAC1.csv</i>	L Panel
Table S.4	monteCarlo_main.R	1111	<i>InferenceT_dim50gamma1HAC1.csv</i>	R Panel
Table S.5	monteCarlo_main.R	1111	<i>InferenceT_dim15gamma1HAC1.csv</i>	L Panel
Table S.5	monteCarlo_main.R	1111	<i>InferenceT_dim50gamma1HAC1.csv</i>	R Panel
Table S.6	monteCarlo_main.R	1155	<i>FactorsT_dim15gamma1HAC1.csv</i>	L Panel
Table S.6	monteCarlo_main.R	1155	<i>FactorsT_dim50gamma1HAC1.csv</i>	R Panel
Table S.7	monteCarlo_main.R	1155	<i>FactorsT_dim15gamma3.16[/.../HAC1.csv</i>	L Panel
Table S.7	monteCarlo_main.R	1155	<i>FactorsT_dim50gamma3.16[/.../HAC1.csv</i>	R Panel
Table S.8	monteCarlo_main.R	1155	<i>FactorsT_dim15gamma1HAC2.csv</i>	L Panel
Table S.8	monteCarlo_main.R	1155	<i>FactorsT_dim50gamma1HAC2.csv</i>	R Panel
Table S.9	monteCarlo_main.R	1155	<i>FactorsT_dim15gamma1HAC2.csv</i>	L Panel
Table S.9	monteCarlo_main.R	1155	<i>FactorsT_dim50gamma1HAC2.csv</i>	R Panel
Table S.10	monteCarlo_main.R	1155	<i>FactorsT_dim15gamma1HAC2.csv</i>	L Panel
Table S.10	monteCarlo_main.R	1155	<i>FactorsT_dim50gamma1HAC2.csv</i>	R Panel
Table S.11	monteCarlo_main.R	1155	<i>FactorsT_dim15gamma1HAC3.csv</i>	L Panel
Table S.11	monteCarlo_main.R	1155	<i>FactorsT_dim50gamma1HAC3.csv</i>	R Panel
Table S.12	monteCarlo_main.R	1155	<i>FactorsT_dim15gamma1HAC3.csv</i>	L Panel
Table S.12	monteCarlo_main.R	1155	<i>FactorsT_dim50gamma1HAC3.csv</i>	R Panel
Table S.13	monteCarlo_main.R	1155	<i>FactorsT_dim15gamma1HAC3.csv</i>	L Panel
Table S.13	monteCarlo_main.R	1155	<i>FactorsT_dim50gamma1HAC3.csv</i>	R Panel
Figure 8.1	empiricalApplication_main.R	394	<i>VarianceImportanceMeasure.pdf</i>	
Figure 8.2	empiricalApplication_main.R	172	<i>CIandEstimates95.pdf</i>	

References

Green, J., Hand, J. R., and Zhang, X. F. (2017). *The characteristics that provide independent information about average us monthly stock returns. The Review of Financial Studies*, 30(12):4389–4436.