

# Replication files for “Information Theoretic Approach to High Dimensional Multiplicative Models: Stochastic Discount Factor and Treatment Effect”

Chen Qiu and Taisuke Otsu

Last updated: June 2021

We use R to generate all empirical results. The following R packages should be installed first: “glmnet”, “lbfgs”, “moments” and “MASS”. They can be installed from R CRAN by using code *install.packages*.

All datasets in the zip file were directly downloaded from Kenneth data library:

[https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

Before running each R script, make sure to change “...” from first line code *setwd* to your working directory and save the R scripts.

## Instructions for replicating empirical results in low dimensional case

1. To replicate results in Tables 1 and 2, first save the following datasets to the working directory:
  - rf.CSV: monthly risk free rate from 1926-07 to 2017-06;
  - F-F\_\_Factors.CSV: monthly returns of Fama-French three factors from 1963-07 to 2010-12;
  - 25FFm.CSV: monthly returns of Fama-French 25 portfolios from 1926-07 to 2017-06;
  - 25\_Portfolios\_5x5.CSV: monthly returns of Fama-French 25 portfolios from 1963-07 to 2010-12.
2. Before running each R script, make sure to change “...” from first line code *setwd* to your working directory and save the R scripts.
3. Run “25FF.KL.R”. This produces two outputs: “25ff.sdf.est.KL.csv” containing extracted out-of-sample SDF using KL divergence without trimming; “Table2.A.SDF.KL.csv” containing results on tests of cross-sectional predictability using extracted SDF.
4. Similarly, running each of “25FF.PSN.1.R” and “25FF.PSN.2.R” will produce two csv files, one on extracted SDF and the other on tests of cross-sectional predictability.
5. Run “25FF.KL.trim.R”. This produces two outputs: “25ff.sdf.est.KL.trim.csv” containing extracted out-of-sample SDF using KL divergence with trimming; “Table2.B.SDF.KL.trim.csv” containing results on test of cross sectional predictability using extracted SDF.
6. Similarly, running each of “25FF.PSN.1.trim.R” and “25FF.PSN.2.trim.R” will produce two csv files, one on extracted SDF and the other on tests of cross-sectional predictability.
7. Table 1 can be replicated from running “Table1.R”.
8. To replicate Table 2, first run “Table2.3Factors.R”, which will produce the benchmark result for the Fama French three factors. The other results in Table 2 can be directly seen from “Table2.A.SDF.KL.csv”, “Table2.A.SDF.PSN.1.csv”, “Table2.A.SDF.PSN.2.csv”, “Table2.B.SDF.KL.trim.csv”, “Table2.B.SDF.PSN.1.trim.csv” and “Table2.B.SDF.PSN.2.trim.csv”.

## Instructions for replicating empirical results in high dimensional case

1. To replicate results in Figure 1 and Table 3, first save the following datasets to the working directory:
  - rf.CSV: monthly risk free rate from 1926-07 to 2017-06;
  - F-F\_\_Factors\_s.CSV: monthly returns of Fama-French three factors from 1993-07 to 2010-12;
  - 100\_Portfolios\_ME\_INV.CSV: monthly returns of 100 portfolios formed on size and investment from 1963-07 to 2017-06;
  - 100\_Portfolios\_ME\_OP.CSV: monthly returns of 100 portfolios formed on size and operating profitability from 1963-07 to 2017-06;
  - 100\_Portfolios.CSV: monthly returns of 100 portfolios formed on size and book-to-market from 1963-07 to 2017-06;
  - 300\_Portfolios\_fortest.CSV: monthly returns of the 300 portfolios above from 1993-07 to 2017-06;
  - F-F\_Momentum\_Factor.CSV: monthly returns of the momentum factor from 1993-07 to 2017-07.
2. Before running each R script, make sure to change “...” from first line code *setwd* to your working directory and save the R scripts.
3. Run “300FFm.KL.R”. This produces two outputs: “300ff.KL.sdf.csv” containing extracted out-of-sample SDF using KL and penalties  $\alpha_n$  from 0 to 2 with 0.05 increments; “table3.SDF.KL.csv” containing cross sectional predictability results for each penalty levels.
4. To save computational time extracting SDF using PSN1 divergence with 300 portfolios, we split the task into four R scripts to allow parallel computation: “300FFm.PSN.1.R”, “300FFm.PSN.2.R”, “300FFm.PSN.3.R” and “300FFm.PSN.4.R”. Each of them has exactly the same codes except that their penalty levels are different. After running all of them, we get extracted SDF using PSN1 and penalties  $\alpha_n$  from 0 to 1 with 0.025 increments by combining: “300ff.PSN.sdf.1.csv”, “300ff.PSN.sdf.2.csv”, “300ff.PSN.sdf.3.csv” and “300ff.PSN.sdf.4.csv”.<sup>1</sup>
  - (a) After getting “300ff.PSN.sdf.1.csv”, “300ff.PSN.sdf.2.csv”, “300ff.PSN.sdf.3.csv” and “300ff.PSN.sdf.4.csv”, run “assettest.300m.PSN.R”, which produces cross sectional regression results “table3.SDF.PSN.csv”.
5. Run “assettest.benchmark.R” to get cross sectional predictability results for Fama-French three factors, saved in “table3.FF.csv”.
6. Run “Figure1.R” to replicate Figure 1.
7. Results in Table 3 can be directly found in “table3.SDF.KL.csv”, “table3.SDF.PSN.csv” and “table3.FF.csv”.
8. Run “Table4.Figure2.R” to replicate results in Table 4 and Figure 2. This will produce two csv files “Table4.A.csv” and “Table4.B.csv”, and an in-app graph that can be saved separately.

## Instructions for replicating empirical results in online appendix

1. The following datasets are needed to replicate Table 5:
  - rf.CSV: monthly risk free rate from 1926-07 to 2017-06;
  - F-F\_\_Factors.CSV: monthly returns of Fama-French three factors from 1963-07 to 2010-12;
  - 10mom.CSV: monthly returns of 10 momentum portfolios from 1927-01 to 2017-06;
  - 25long.CSV: monthly returns of 25 long term reversal and size portfolios from 1927-01 to 2017-06;
  - 10mom for test.CSV: monthly returns of 10 momentum portfolios from 1963-07 to 2010-12;

---

<sup>1</sup>On a laptop with a 1.4 GHz Quad-Core Intel Core i5 processor with 8 GB memory, it took about 14 hours to run “300FFm.PSN.1.R”, an hour and half to run “300FFm.PSN.2.R”, 39 hours to finish “300FFm.PSN.3.R”, and less than 30 min to run “300FFm.PSN.4.R”.

- 25long for test.CSV: monthly returns of 25 long term reversal and size portfolios from 1963-07 to 2017-06.
2. Run “10mom.R”. This will produce several files: “10mom.sdf.est.csv” containing extracted SDF using KL divergence; “table5.A.FF.csv” and “table5.A.SDF.csv” summarizes regression results in Panel A of Table 5.
  3. Run “25long.R”. This will produce several files: “25long.sdf.est.csv” containing extracted SDF using KL divergence; “Table5.B.FF.csv” and “Table5.B.SDF.csv” summarizes regression results in Panel B of Table 5.
  4. The following datasets are needed to replicate Table 6:
    - rf.CSV: monthly risk free rate from 1926-07 to 2017-06;
    - F-F\_\_Factors.CSV: monthly returns of Fama-French three factors from 1963-07 to 2010-12;
    - 25\_long\_size.CSV: monthly returns of 25 long term reversal portfolios from 1931-01 to 2017-06;
    - 25\_mom.size.CSV: monthly returns of 25 momentum portfolios from 1931-01 to 2017-06;
    - 25\_short.size.CSV: monthly returns of 25 short term reversal portfolios from 1931-01 to 2017-06;
    - 75mixed for test.csv: monthly returns of the above 75 portfolios 1963-07 to 2017-06;
    - 49\_Industry.CSV: monthly returns of 49 industry portfolios from 1926-07 to 2017-06;
    - 49\_Industry for test.CSV: monthly returns of 49 industry portfolios from 1963-07 to 2017-06;
    - 100FFm.CSV: monthly returns of 100 size and book-to-market portfolios from 1926-07 to 2017-06;
    - 100\_portfolios.CSV: monthly returns of 100 size and book-to-market portfolios from 1963-07 to 2010-12.
  5. Run “100FFm.R”. This will produce several files: “100ff.sdf.est.csv” containing extracted SDF using KL divergence; “table6.A.FF.csv” and “table6.A.SDF.csv” summarize regression results in Panel A of Table 6.
  6. Run “49ind.R”. This will produce several files: “49ind.sdf.est.csv” containing extracted SDF using KL divergence; “table6.B.FF.csv” and “table6.B.SDF.csv” summarize regression results in Panel B of Table 6.
  7. Run “75.mixed.R”. This will produce several files: “75mixed.sdf.est.csv” containing extracted SDF using KL; “table6.C.FF.csv” and “table6.C.SDF.csv” summarize regression results in Panel C of Table 6.