

Useful R Code Examples and Data They Use

V2 May 28 2016

BTgmv40StocksLoSfm.R

Does back-test of GMV portfolio of first 40 stocks in the “stocks145bonds7.csv” file using both the sample covariance matrix and the statistical factor model covariance matrix computed with **fitSfm**. Result shows improved performance using **fitSfm**.

BTgmv145StocksLoSfm.R

Now working as of today May 28, the SFM covariance with the 145 stocks beats the SP500 nicely. Takes around 10 minutes on my machine.

BTgmvLoBoxWeekly104.R

Uses weekly data **smallcapW** from **mpo**. Back-test comparison of global minimum variance (GMV) long-only and box-constrained [0, 0.2] portfolios, both of which are compared with the market portfolio. Results show that the box constraint is better than the long-only constraint, and both are considerably better than the market.

BTQuLoBox104Lambda20.R

Uses weekly data **smallcapW** from **mpo**. Back-test comparison of quadratic utility optimal long-only and box-constrained [0, 0.2] portfolios for a risk aversion parameter value of $\lambda = 20$. Results show that the box constraint is overall a bit better than the long-only constraint and noticeably so during the recovery from the market crisis starting 2009, and both are considerably better than the market.

covmatRMT3.R

Example of using eigenvalue shrinkage based on random matrix theory (RMT), to get better performance than with sample covariance matrix on DJIA daily returns. You need to install and load the **covmat** package. See the Section 4

of the “covmat vignette 09 28 15.pdf” now available in the Final Project files folder.

eigen145stocks.R

Uses monthly stock returns in “stocks145scores6.csv”. Shows that for 25 years (300 months of returns) all eigen-values (EV’s) of the returns covariance matrix are non-zero, and shows that for 5 years (60 months of returns) there are only 59 non-zero EV’s (just as theory tells you).

ffModIndustryStocks145.R

Uses monthly stock returns and scores in “stocks145scores6.csv”. Fits fundamental factor model with **fitFfm** function using for all 25 years and the SECTOR exposure variable for 10 sectors, with option to do computation for just the 5-year period from 2008 through 2012. A cross-section model is fit for every month in the overall time interval. Monthly R-squared values are computed and displayed (nicer plot of the 5-year segment than for the 25 years), and average R-squared across time is computed. The code makes plots of time series of the 10 factor returns, and makes classical and robust correlations of the 10 factor returns and displays with ellipses plots.

ffModMixedStocks145.R

This code is a modified version of **ffModIndustryStocks145.R** in which the 6 risk indices (“scores”) are also included in the model. This is the typical industry version, i.e., industry and/or sector exposures plus continuous variable exposures.

ffModRiskIndices145.R

This code is a modified version of **ffModMixedStocks145.R** in which the sector exposures are deleted from the model.

fitSfmCovVsDataCov145stocks.R

Shows that the sample covariance matrix and the factor model covariance matrix obtained from the fitted statistical factor model using **fitSfm** are not the same.

plotTsStocks145.R

Uses monthly stock returns in “stocks145scores6.csv”, and provides a function **plot.Ts** to conveniently plot the time series of the 145 stocks on multiple pages. The code currently plots 10 time series per page, and so the results appear on 15 pages in the Rstudio plot window. But the code could be easily changed to use different layouts and correspondingly different number of pages.

tsPlotMP.R

This is a convenience function for plotting multiple time series using lattice graphics (hence requires the package **lattice**), with a number of useful optional arguments. The most useful of the latter is the **scaleType = “free”** default, which will often be most useful when set to **scaleType = “same”**. This function is created at the beginning of the R code script of **ffModIndustryStocks145.R**, as well as the other two fundamental factor model fitting scripts.

Data Sets Used by Preceding Code

stocks145bonds7.csv

Monthly returns from 1/31/1990 to 12/31/2014

145 stocks

7 bonds: 1, 2, 5, 7, 10, 20, 30 year

T-bills: 30D, 90D

Other: CPI (consumer price index), MKT (CRSP “market”), SP500

stocks145scores6.csv

Monthly returns from 1/31/1990 to 12/31/2014

Contains the following columns of data for each stock:

DATE, compustat (number), TICKER, COMPANY, GSECTOR

Six “scores” from Capital-IQ:

SECTOR	(GICS sector name)
ROE	(Return on equity)
BP	(Book to price ratio)
PM12M1M	(Momentum)
SIZE	(Size, log market cap in \$M)
ANNVOL1M	(Annualized monthly volatility)
EP	(Earnings to price ratio)

RETURNS.RAW: returns with missing data and zero's

RETURNS: cleaned returns

rf: risk-free rate

stocks145scores6lagged.csv

This is the same as the above “stocks145scores6lagged.csv”, except that the scores (exposures) are lagged by one month as is needed for a predictive fundamental factor model, i.e., the return is at time t but the exposures are at time $t-1$.

smallcapW

Weekly returns of 20 smallcap stocks from 1997 through 2010, contained in **mpo**.