Data in Brief 8 (2016) 858–862

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Real-world datasets for portfolio selection and solutions of some stochastic dominance portfolio models

Renato Bruni [a], Francesco Cesarone [b,*], Andrea Scozzari [c], Fabio Tardella [d]

[a] Dip. Di Ingegneria Informatica, Automatica e Gestionale, Sapienza Università Di Roma, Rome, Italy
[b] Dip. di Studi Aziendali, Università di Roma Tre, Rome, Italy
[c] Facoltà di Economia, Università degli Studi Niccolò Cusano, Rome, Italy
[d] Dip. Metodi e Modelli per l'Economia, il Territorio e la Finanza, Sapienza Università di Roma, Rome, Italy

ARTICLE INFO

Article history:
Received 26 February 2016
Received in revised form
9 June 2016
Accepted 21 June 2016
Available online 28 June 2016

ABSTRACT

A large number of portfolio selection models have appeared in the literature since the pioneering work of Markowitz. However, even when computational and empirical results are described, they are often hard to replicate and compare due to the unavailability of the datasets used in the experiments.

We provide here several datasets for portfolio selection generated using real-world price values from several major stock markets. The datasets contain weekly return values, adjusted for dividends and for stock splits, which are cleaned from errors as much as possible. The datasets are available in different formats, and can be used as benchmarks for testing the performances of portfolio selection models and for comparing the efficiency of the algorithms used to solve them. We also provide, for these datasets, the portfolios obtained by several selection strategies based on Stochastic Dominance models (see "On Exact and Approximate Stochastic Dominance Strategies for Portfolio Selection" (Bruni et al. [2])). We believe that testing portfolio models on publicly available datasets greatly simplifies the comparison of the different portfolio selection strategies.

* Corresponding author. Phone: +39 06 57335744.
E-mail addresses: francesco.cesarone@uniroma3.it (F. Cesarone).
URL: http://host.uniroma3.it/docenti/cesarone

http://dx.doi.org/10.1016/j.dib.2016.06.031

Specifications Table

Subject area	*Economics and Finance*
More specific subject area	*Portfolio selection, Portfolio optimization, Asset allocation*
Type of data	*Tables, text files, excel files, matlab files, figures*
How data was acquired	*Thomson Reuters Datastream, Fama & French Data Library*
Data format	*Processed, filtered, analyzed*
Experimental factors	*When necessary, the assets prices are filtered to check and to correct missing or inaccurate data*
Experimental features	*All data sets provided consist of weekly assets returns readily usable in Portfolio Selection models*
Data source location	*N/A*
Data accessibility	*Data is within this article*

Value of the data

- The datasets provided here can be used as benchmarks by researchers willing to implement and to compare portfolio selection models on publicly available data.
- If different researchers use the same publicly available data, the comparison of different approaches would be more easy and fair.
- The data are filtered to remove possible errors in the original source. This allows researchers to perform more accurate and realistic simulations and evaluations.
- For our datasets we also provide the solutions to several portfolio selection models. Such solutions can be used by other researchers to compare the efficiency of their algorithms and the quality of their solutions.
- Availability of data and solutions can stimulate contacts among researchers working in this area for future collaborations and projects.

1. Data

We provide weekly returns time series for assets and indexes belonging to several major stock markets across the world. Weekly returns data are computed from prices values obtained from *Thomson Reuters Datastream* (http://financial.thomsonreuters.com/) and from daily returns obtained from *Fama & French Data Library* (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). The data are filtered to check and to correct missing or inaccurate values. The data provided can be used as input for several types of portfolio selection models to compare on both efficiency and performance (for references on portfolio selection approaches see, e.g., [3]). For the above datasets, we also include as benchmarks the portfolios obtained by using several selection strategies based on both exact and approximate Stochastic Dominance models (described in [2]).

2. Experimental design, materials and methods

Asset allocation aims at selecting a portfolio over N available assets in an investment universe $A=\{1,\ldots,N\}$ according to specific choice criteria under uncertainty. More precisely, we must decide how much of each asset $i\in A$should be purchased in the selected portfolio. The portfolio is denoted by $x=\{x_1,\ldots,x_N\}$, where x_i is the fraction of the given capital invested in asset $i\in A$.