

Details on the updated Active Share data file

(c) Antti Petajisto, 2013-05-27

Variable definitions

wfincn	Wharton financial institution code number; unique fund identifier
rdate	Report date (effective date) of fund's holdings
fundno	Fund number in Thomson Reuters mutual fund holdings data; supposedly unique fund identifier but occasionally reused in later periods
crsp_fundno	Fund number in CRSP mutual fund data; unique identifier for one share class of a fund; here we include the fund number of the largest share class of a fund
index	Official benchmark index disclosed by the fund in its prospectus (inferred from holdings when prospectus benchmark not available)
index_min	Alternative benchmark index that produces the lowest Active Share for a fund as of each report date
activeshare	Active Share of a fund relative to its official benchmark index ("index") as of each report date
activeshare_min	Active Share of a fund relative to the alternative benchmark index (index_min) as of each report date
trackingerror	Tracking error volatility of a fund, computed based on the prior 6 months of daily returns; expressed as an annualized number
trackingerror_min	Tracking error volatility of a fund relative to its alternative benchmark index (index_min)
indexfund	Indicator variable for a pure index fund
enhanced_index	Indicator variable for an enhanced index fund
TNA	Total net assets of a fund's all share classes based on CRSP data
ticker	Ticker symbol of the largest share class of the fund
fund_name	Name of the largest share class of the fund

Other details

All variables are defined and the sample of funds is selected as in Petajisto (2013). The benchmark indices included in the calculations were 5 indices from S&P, 12 indices from Russell, and 2 indices from Dow Jones / Wilshire. For example, "S5" means S&P500, "W4" means Wilshire 4500, and "R2G" means Russell 2000 Growth. This sample covers the years 1980-2009 as in Petajisto (2013).

Petajisto (2013) uses the official benchmark index of each fund as stated in its prospectus. The associated variables are *index*, *activeshare*, and *trackingerror*. However, in a few cases the official prospectus benchmark is misleading, so I am also reporting the index that produces the lowest Active Share (*index_min*), along with Active Share and tracking error numbers associated with it (*activeshare_min*, *trackingerror_min*).

The methodology is identical to that followed by Cremers and Petajisto (2009), except for the differences listed in the data section of Petajisto (2013).