Empirical Appendix of

IS THE ACTIVE FUND MANAGEMENT INDUSTRY CONCENTRATED ENOUGH?

David Feldman, Konark Saxena, Jingrui Xu

Latest Revision May 22, 2018

Table of Contents

1   U.S. Active Equity Mutual Funds - Sample Selection

Our sample period is from January 1979 to December 2014. We use monthly observations for the active equity mutual fund data.

First, we download U.S. mutual fund data from Morningstar, which provides fund share class level data. The datasets we download include "United States Mutual Funds" (30,782 share classes) and "United States Closed-End Funds" (565 share class).[1]

Second, we screen out share classes from funds of funds and index funds. We drop a fund share class if it has a value of "Yes" in any of the variables "Fund of Funds," "In House FOF," "Enhanced Index," and "Index Fund."

Third, to screen out share classes of non-equity funds, we drop the funds if their "Global Broad Category Group" value is not "Equity." We keep the funds in the "Equity Funds" category because these funds invest only in the domestic equity market. Note that we do not need to use the "Primary Prospectus Benchmark" to screen out non-equity funds as do Pastor, Stambaugh and Taylor (2015; PST). This is because, in our sample, funds with a value "Equity" in the "Global Broad Category Group" variable do not contain any keywords in "Primary Prospectus Benchmark" that would be defined as non-equity funds in PST.

Fourth, we calculate fund returns as the net asset value weighted average of fund share class returns. All the monthly share class returns are net of administrative and management fees and other costs taken out of fund assets. We identify that a fund share class observation belongs to a particular fund using the fund identifier variable "Fund ID." To calculate fund returns, accounting for missing data, we use an algorithm that ensures the following: in each month, if a share class's return is missing, its net asset value is not used in calculating the asset value weighted average return. If, for a fund, all share classes have return observations but all their net asset values are missing, the fund returns are equally weighted averages of all the share classes' returns. If, for a fund, all share classes have return observations but some have missing net asset values, then almost all the weight is allocated to the share classes with net asset value observations. The algorithm follows:

---

[1] We also download "Unit Investment Trust" (14,881 share classes); however, all funds in this dataset have neither long enough lives nor records of net asset values. So, all funds in the "Unit Investment Trust" dataset are screened out by our criteria.

1. Define a new variable "NA" equal to the Morningstar "Net Asset Share Class (Monthly)" field containing assets at the fund share class level.

2. In each month, if the observation of the "Monthly Return" variable (share class return variable) is not missing, but "NA" is missing, set "NA" to $1.

3. In each month, if the observation of the "Monthly Return" variable is missing, set this month's "NA" to missing.

4. In each month, calculate a variable "TA" (fund's total assets) as the sum of "NA" of share classes with the same "Fund ID" value.

5. In each month, calculate a variable "TA (post return)" as the sum-product of "NA" and "Monthly Return" of share classes with the same "Fund ID" value. This variable measures the total assets of the fund post realization of monthly returns (before considering flows).

6. Calculate the variable "Fund Return" as "TA (post return)" divided by "TA".

The fund data sample obtained after this step is defined as our U.S. active equity fund sample.

2   Estimation of PC-Adjusted Fund Net Alphas

To evaluate robustness of our results to measures of fund net alpha, we use an alternative method to measure them: principal component adjusted fund net alphas ("PC-Adjusted Fund Net Alphas"). Our method addresses the possibility that traded index funds do not capture unobserved common risk factors that drive excess returns. We follow Connor and Korajczyk (1988; CK; pp. 259–260) to calculate asymptotic principal components of fund net alphas. This methodology is described below.

Define $\boldsymbol{\alpha}$ as a $M \times T$ matrix of fund net alphas in the market whose row indices represent funds and column indices represent time. First, estimate the $T \times T$ cross-product matrix $\boldsymbol{\Omega} = \boldsymbol{\alpha}^{\mathbf{T}} \times \boldsymbol{\alpha} / M$. CK's method is designed for balanced samples and samples without missing observations. We modify their method to allow for missing values by making sure that we do not count missing values when taking "averages" and instead calculate the matrix element-wise, as $\boldsymbol{\Omega}(s, t) = \sum_{i=1}^{M} \boldsymbol{\alpha}(i, s) \times \boldsymbol{\alpha}(i, t) / m_{s,t}$, where $m_{s,t}$ is the number of funds whose fund net alphas are non-missing at both time s and time t. Connor and Korajczyk (1986) show that the eigenvector matrix of $\boldsymbol{\Omega}$ provides estimates of common factors. We allow for up to 6 omitted common factors and, therefore, restrict our analysis to the first 6 columns of the eigenvector matrix. This $T \times 6$ matrix of eigenvectors is denoted $\mathbf{G}$.

CK suggest a refined procedure in terms of estimation efficiency, building on the Connor and Korajczyk (1986) method described above. Their idea is to scale each return by

the standard deviation of its idiosyncratic returns to improve the estimation efficiency of the asymptotic principal components. We follow this refined procedure, which is described next.

The first step is to estimate the $\mathbf{G}$ matrix, as described above. Next, regress the M fund alphas on the estimated common factors, that is, regress $\boldsymbol{\alpha}^{\mathbf{T}}$ on $\mathbf{G}$ (with a constant term), to obtain a $T \times M$ residual matrix. Next, calculate pair-wise covariance between the residuals of any two funds, and estimate the $M \times M$ residual variance matrix $\mathbf{V}$ element-wise, where $\mathbf{V}(i, j)$ is the pair-wise covariance of the residuals of fund i and fund j.[2] Then, construct a $M \times M$ diagonal matrix $\mathbf{DIAG(V)}$, whose diagonal elements are the same as those in $\mathbf{V}$, and off-diagonal values are zero. After that, estimate the scaled $T \times T$ matrix $\boldsymbol{\Omega}^{*}$ whose eigenvectors are the CK asymptotic principal components. The expression for this $T \times T$ matrix is $\boldsymbol{\Omega}^{*} = \boldsymbol{\alpha}^{\mathbf{T}} \times \mathbf{DIAG(V)^{-1}} \times \boldsymbol{\alpha} / M$. As before, calculate this matrix element-wise, taking into account the number of missing values when calculating each element. The first 6 eigenvectors of $\boldsymbol{\Omega}^{*}$ are denoted as $\mathbf{G}^{*}$, which is a $T \times 6$ matrix. Each column in $\mathbf{G}^{*}$ is an asymptotic principal component that estimates the time series of common factor returns.

In our main tests, we use the first two principal components to calculate our PC-adjusted fund net alphas." For each fund, we regress its time-series of fund net alphas $\alpha_{i,t}$ on the first two principal components PC_t^1 and PC_t^2, without a constant term:

$$\alpha_{i,t} = b_{i,t}^1 PC_t^1 + b_{i,t}^2 PC_t^2 + \alpha_{i,t}^{PC} . \tag{1}$$

Our estimate of PC-adjusted fund net alphas is the estimated residual: $\alpha_{i,t}^{PC}$.

REFERENCES

Connor, G. and Korajczyk, R.A., 1986, Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis. *Journal of Financial Economics* 15(3), 373–394.

Connor, G. and Korajczyk, R.A., 1988, Risk and Return in an Equilibrium APT Application of a New Test Methodology. *Journal of Financial Economics* 21, 255–289.

Pastor, L., Stambaugh, R.F. and Taylor, L.A., 2015, Scale and Skill in Active Management. *Journal of Financial Economics* 116, 23–45.

---

[2] We use pair-wise variance instead of a covariance matrix of the residual matrix because different funds have missing fund net alphas in different months, and calculating the covariance matrix of the residual matrix will only use the observations of the months when all funds have data (in software such as MATLAB). Calculating a pair-wise variance using the observations of the months when two funds have data means that the non-missing observations can be better utilized.