RiXtrema Research

Significant Fee Waste in Retirement Plans – New Study Using Quantitative Methods

Working Paper

by Daniel Satchkov, CFA,
Yon Perullo, CFA,
Val Kolesov,
Max Grishkov

Significant Fee Waste in Retirement Plans – New Study Using Quantitative Methods

Abstract: Fee inefficiency is pervasive in 401(k) plans and pressure has been building up on plan sponsors to address this problem. And with the new DOL Fiduciary Rule this pressure is now applied to the whole industry of financial advice. However, there has been surprisingly little quantitative evidence to show how much retirees are overpaying in 401(k) plans and other defined contribution plans. We analyzed 52,529 retirement plans from the DOL EFAST database. Using quantitative methods we estimate that participants could save .25% a year on a weighted average basis by switching into lower cost investments that are quantitatively very similar to those they already hold. Similarity is defined as a combination of category filters, together with historical and forward looking predicted correlation based on a multi-factor model. As of March 2015 total defined contribution plan assets stood at $6.8 trillion. This means that savings of approximately $17.07 billion annually are being wasted in the retirement industry. To address arguments about infeasibility of passive investment in aggregate we ran the same analysis while excluding index funds and ETF funds from available list of alternatives. The results do not change significantly on a weighted average basis. We make every effort to be conservative in our methods, so our estimate is likely near lower bound of available savings.

Background

It has been widely argued that 401(k) and other retirement plan participants are poorly served by plan menu choices filled with expensive mutual funds. These expensive funds tend to underperform over time due to higher fees and lack of consistent alpha. Defined contribution plans held $6.8 trillion of fund assets at the end of March of 2016 and a great deal is at stake for participants in this debate. However, there is surprisingly little quantitative evidence regarding fee inefficiency in retirement plans. Estimates of waste in plans varies; for instance the DOL used an average figure of 11.3 basis points as an estimate of waste in defined contribution plans.[1]

We could find only one rigorous quantitative study of 401(k) fee overpayment; see Ayres and Curtis [2015]. The study used 3,500 plans to conclude that menu restrictions in an average plan led to additional cost of seventy eight basis points, above low cost index fund basket. This study, while thorough and illuminating, can be challenged on the basis of an argument that is frequently used as a stopgap in any discussions of excessive fees, namely it could be argued that high fee funds held by participants cannot be directly compared to the low cost funds due to unique return and correlation profile. Our study overcomes this difficulty by making sure that the low fee alternatives are chosen to resemble the active fund being replaced, both qualitatively in terms of fund category and quantitatively in terms of past and holdings based behavior.

---

[1] Assessing the Department of Labor's Assumption that 401(k) Plan Participants Pay Fees that are Higher than Necessary, Investment Company Institute
https://www.ici.org/policy/regulation/fees/ci.08_dol_401k_disclosure_com_app.print

Our ultimate goal is to find a low fee replacement for high fee funds, but only where we can prove that the replacement is not materially changing the risk/return profile offered to participants in the current menu[2]. Additional (and very strict) criteria that we implemented in this study is that any lower fee alternative should also have better 10 year performance. Our goal is to find those high-fee funds that exhibit a great degree of similarity to low cost alternatives, measure that similarity and offer evidence-based low fee replacements. This will give us the ability to estimate potential savings to retirement plan participants. We will define similarity more precisely later in the text, but generally we define similarity as a correlation between two funds subject to condition that funds are in the same category. In addition to this analysis, we also ran the same dataset, but excluded index and ETF funds from the available universe. In fact, the results are very similar even when excluding such frequent forms of passive investment.

Materials & Methods: Plan Data

We have studied 52,529 retirement plans from the DOL EFAST database. Certain 401(k) and 403(b) plans have to submit a long form 5500 which includes a schedule of assets, also called Schedule H. This is the plan data we used. There are some important issues that need to be solved in order to read that data. For example, there are usually no tickers for funds and often the share class is not reported. In addition, the names are not standardized and every schedule seems to use its own abbreviations, sometimes extremely imaginative. Typos are also prevalent. These are probably some of the reasons for the lack of widespread studies on this topic.

Our data was obtained and cleaned in the following way. We downloaded PDF files containing Schedule H from the DOL EFAST website. We then parsed the data containing fund names and market values as of the end of the reporting period to csv files using Adobe Acrobat Pro XI with PaperCapture image recognition. We then use text search algorithms to assign tickers to every fund based on its name by cross checking across our database containing 31,589 funds available for sale in the US. When share classes were not available, we assumed share class that is most beneficial for the plan sponsor i.e. we picked the lowest cost share class of the fund, as long as the plan could satisfy the minimum required for that share class. This certainly means that many unreported A shares were mapped by us to R3 or other cost efficient retirement share classes, thereby giving every benefit of the doubt to the plan. This also means that the weighted (by plan size) average savings of 25 basis points that we later estimate, is likely the lower bound of available savings for the participants.

Only plans for which more than 80% of the investments reported in Schedule H could be identified are included in this study

[2] We realize that current menus may be imperfect in risk/return profiles as well, but that is the question that is not solved by this study.

Materials & Methods: Fund Data

For calculation of Similarity we have the following building blocks.

- History of returns for 31,589 funds used for calculation of historical correlations (source is Bloomberg, Yahoo and Quandl API)
- Fund objective (source is Bloomberg)
- Position data for mutual funds and ETFs (from SEC EDGAR and ETF provider websites respectively)
- Multi-factor multi-asset class factor model used to forecast correlation for those funds where holdings are available

Materials & Methods: Similarity Calculation

In order to calculate potential savings from switching to lower cost alternatives, we need to select an investable universe that plan menus can choose from. The broadest possible such universe would be an Open Architecture arrangement, under which a retirement plan can select from a large variety of funds available for sale in the United States. Thus, we assume an Open Architecture with 31,589 funds. Our next goal is to calculate Similarity measure which we will use to select the most similar lower expense fund to replace a more expensive fund.

Definition of Similarity:

We define Similarity between any two funds as:

$$Similarity = \frac{\rho_H + \rho_P}{2}$$

Where:

ρ_P - Holdings based correlation based on fund's holdings and use of multi-factor model (where holdings are available, for more detail see Appendix B for holdings data preparation and Appendix C for holdings based correlation calculation)

ρ_H - Historical correlation based on past returns (for more detail, see Appendix D)

Materials & Methods: Replacing Expensive Funds

For every fund that is currently in the plan, we find all funds with the same Bloomberg Objective[3] . We then find that satisfy the following criteria:

- $\rho_H > .9$

- $Similarity \geq .9$

Some funds do not have holdings data, and only historical correlation is used for them. We sort the results of these filters in the ascending order by the Net Expense ratio to find the lowest cost similar funds. We then remove those alternatives that had inferior 10 year performance relative to the incumbent fund in the plan. The logic behind these conditions is relatively straightforward. We are looking for funds that would be qualitatively (category) and quantitatively (past and holdings based correlation) similar to the fund in the plan, while making sure that our alternatives have a better track record. Holdings based correlation is used as an additional variable in those cases when holdings are available. Because holdings based correlation looks at recent fund positioning the similarity call is call even more reliable. As will be shown, many high fee funds will have similarity greater than .9 to low cost alternatives.

Materials & Methods: Calculating Participant Savings

Any fund in the plan with at least one alternative that is less expensive can be replaced with that alternative without significantly changing risk/return profile of the plan menu. Thus, the difference in net expense represents savings for participants. Obviously, typical low cost replacement will be an index fund of one type or another. But in this study we added criteria to ensure that only quality passive investments pass as alternatives and also added another run excluding most of the passive investments altogether. The results show that even with application of such rigorous criteria, defined contribution plans are likely wasting close the $17.07 billion per year in fund expenses.

There is one more important point that needs to be accounted for in order to calculate the savings correctly. Many plans, especially smaller ones, employ revenue sharing arrangements. These arrangements, while increasingly frowned upon due to their opacity, are still legal. Under such arrangement a certain portion of expense from higher expense funds gets back to the plan and is used to pay for admin or advice expenses. The arrangement is less than ideal for participants, because the actual fees paid for fund expense and administration of the plan become hard to decipher. However, if we are switching to lower expense fund, likely an index fund, there will be no revenue sharing. We have to account for that by approximating revenue sharing, since it is not reported clearly on form 5500. We assumed that revenue sharing is equal to 12b-1 fee of the fund. This is a reasonable assumption, though

[3] Using Morningstar Category does not make any substantial difference.

in many cases plans or third-party administrators will negotiate special revenue arrangements that could be better than the 12b-1 fee.

Thus, the plan savings from switching to low fee funds with high Similarity are as follows:

Given:

A_f - total assets in a given fund f

x_H - expense ratio of the high expense fund minus the 12b-1 fee

x_L - expense ratio of the low expense fund

a - additional admin fee (if any) or revenue sharing from a given fund (if any)

v - additional advisor fee (if any)

N - number of years

r - assumed return

replacement fund.

$\%P / A$ - is Total Plan Savings expressed in percent per annum

$$\%P / A = \sum_{f=1}^{Z} w_f * (x_H - x_L - a - v)$$

Where:

Z - total number of funds in the plan

Results

Exhibit 1 below presents the results of our analysis for the overall universe as well as plans stratified by size into buckets where %P/A is percentage savings available in annual terms.

Exhibit 1 – % Per Annum Savings From Switching To Lower Expense Funds With High Similarity & 10 Year Outperformance of the Incumbent Fund

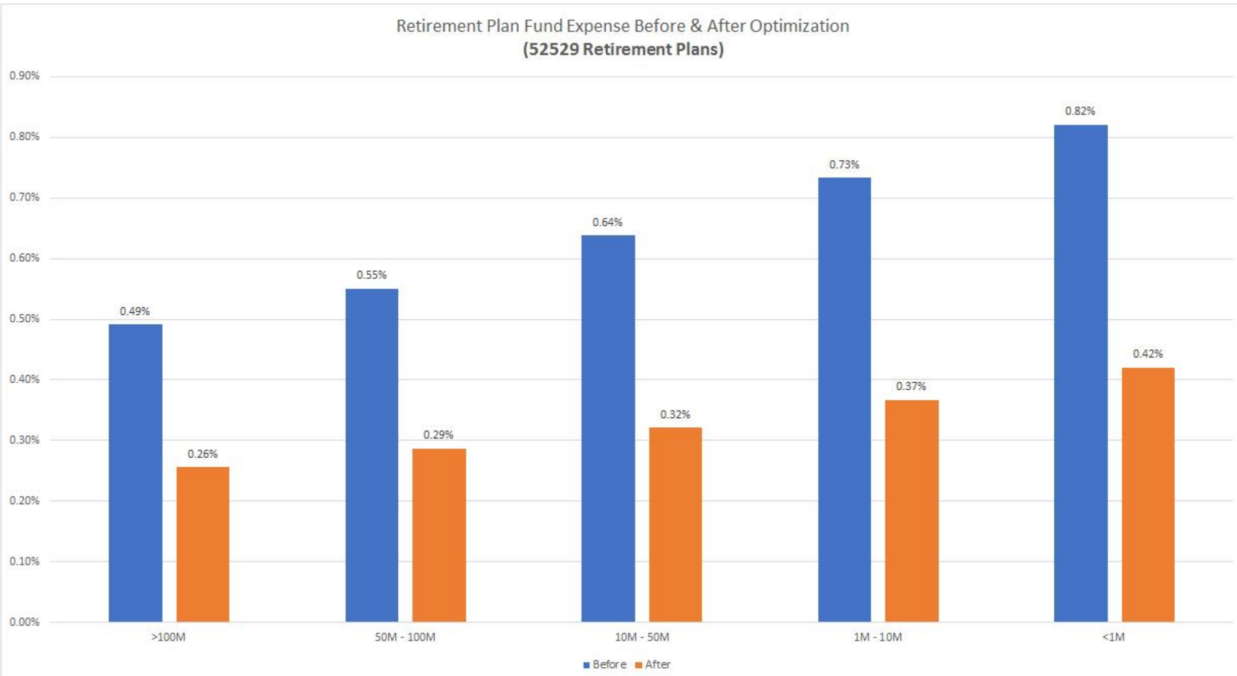


Exhibit 1 shows that larger plans are indeed less inefficient with average net expense less the 12b-1 fee equal to about .42% vs .82% for the smallest plans. But as is clear from the exhibit, quantitative algorithms can find highly similar less expensive alternatives for the larger plans, reducing expense from .49% to .26%. This is not suprprising considering the wealth alternatives available to large plan sponsors, including insitutional shares and various trust investments. Overall, the weighted average plan savings are equal to .25%, which suggests that the overall savings available to defined contribution plans are $17.07 billion per year only on the investment expense of the funds.

Exhibit 2 – Histogram of % Per Annum Savings For Universe of 52,529 Plans

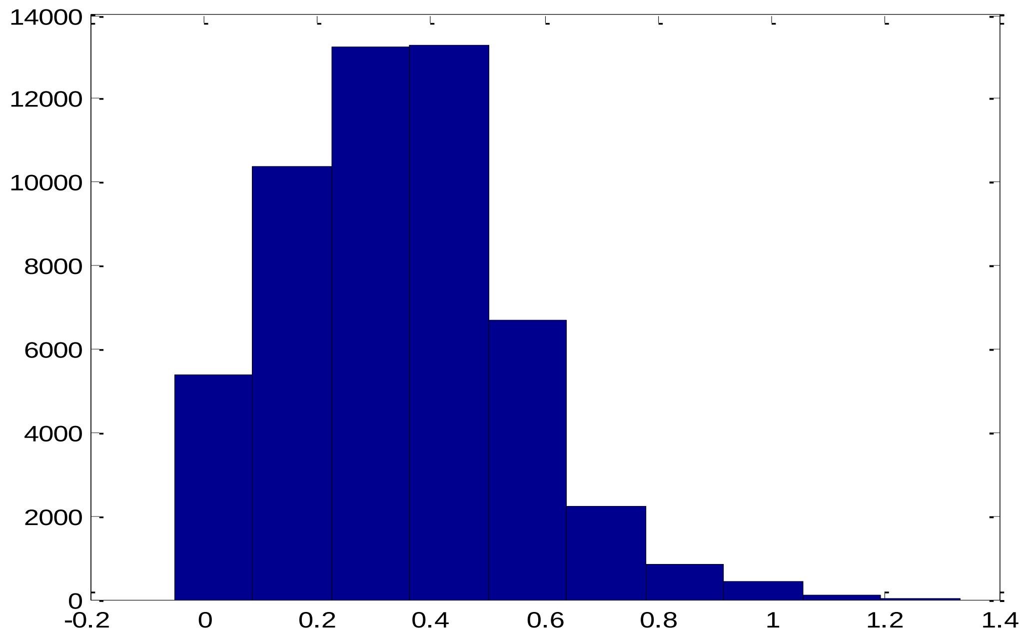
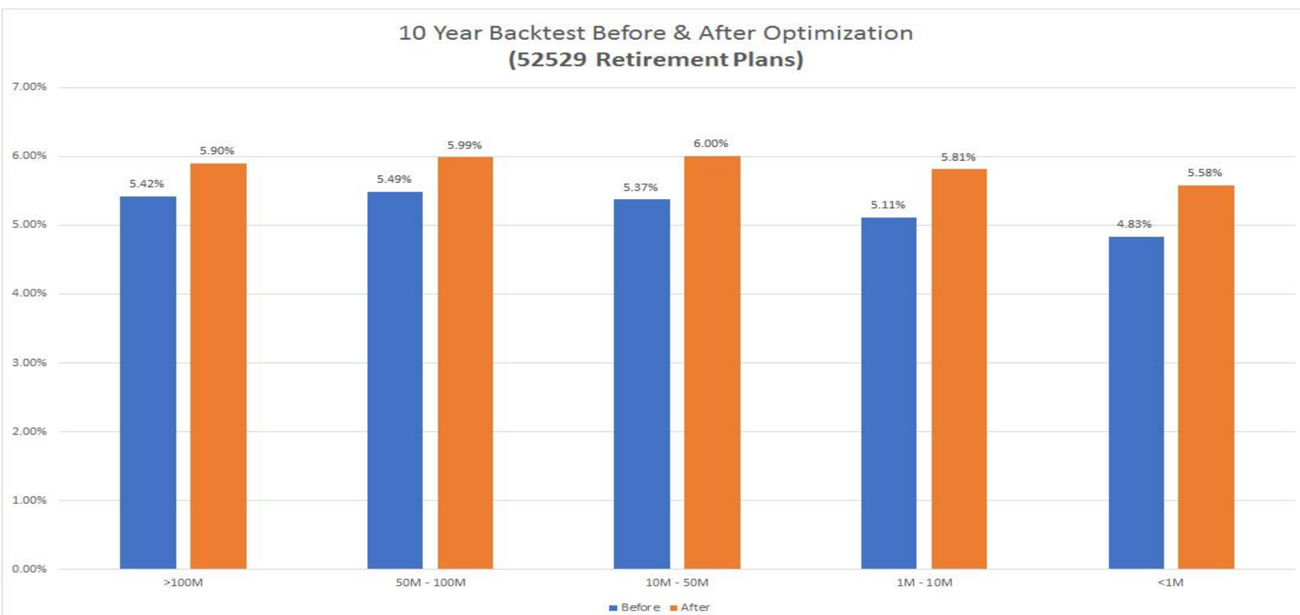


Exhibit 3 – Number of Plans in Sample within Each Bucket

Universe	Total
>100M	2855
50M - 100M	2527
10M - 50M	15109
1M - 10M	27759
<1M	4279
Total	52529

As discussed above, we constructed our similarity mapping to pick only funds that have a better 10 year track record than the incumbent (if the track record was less, then performance had to be better over whatever history the incumbent has). Exhibit 4 confirms that indeed the performance of our optimized significantly less expensive plan menus is consistently better than the current ones. In Exhibit 4 we show performance of incumbent plan menu vs. the proposed optimized plan menu backtested for 10 years. This means that we take both menus with current weight allocations and measure the annualized performance of that plan. For micro plans with less than $1 million in assets, the annualized 10 year performance has gone from 4.83% to 5.58%, a .75% improvement. Note, that the pure expense savings were only .4%, so the aggregate outperformance is not just due to savings, but the superior gross performance of lower fee alternatives as well. Likewise the performance gap for all other plan size buckets is larger than the expense differential itself. For $1-$10 million, the fee differential is .73% vs .37% in Exhibit 1, whereas the performance differs by .7% (5.81% vs 5.11%). For largest plans the fee differential is .23% (.49% vs .26%), but the performance difference is more than twice as large at .48% in Exhibit 4. Thus, the quality of the lower fee menus is actually higher. To calculate our approximate $17 billion in annual savings for retirement plans we used average fee savings of .25%, but if we are to look at actual performance over the past 10 years – then the actual benefit to participants could be twice as large. So, our $17 billion savings estimate is quite conservative based on these empric findings.

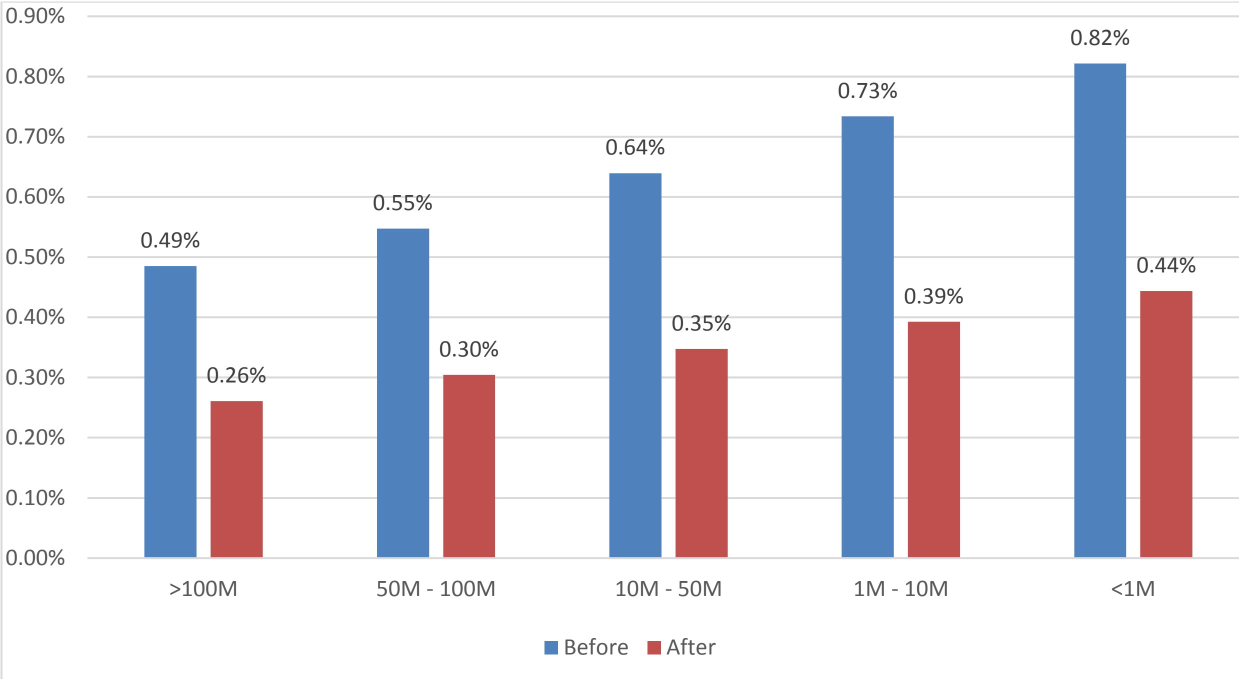
Exhibit 4 – 10 Year Annualized Net-of-Fees Performance Backtest Before & After Optimization for 52,529 Plans



Fiduciary Conversation is not About Active vs Passive

We also reran this analysis adding one more filter. In this run we removed all funds that are flagged in our database as index funds and ETFs. The reason for this is as follows. There is a very strong argument that taking passive investment to the extreme will result in extreme market instability, when everyone is a passive investor with no price discovery. While the results our studies strongly argue that fee inefficiency is pervasive, we understand the argument that indiscriminant movement to passive investing is dangerous. That is why we reran the whole study excluding many passive investments.

Exhibit 5 – % Per Annum Savings From Switching To Lower Expense Funds With High Similarity & 10 Year Outperformance of the Incumbent Fund (excluding Index Funds & ETFs)



As can be seen from Exhibit 5, removal of index funds and ETFs does not negate the savings available to the contribution plans. The savings for large plans are virtually identical, while for smaller plans some opportunities are removed and optimal fees are a bit higher. For example, optimal fees for plans with less than $1 million in assets goes up from .42% to .44%. There are many relatively efficient non-passive funds from providers like DFA which ensure that removing passive investments still allows for fee efficiency.

Discussion

Weighted average available savings for the analysis corresponding to exhibits 1 & 2 are twenty-five basis points per annum, while meeting very strict criteria of similarity and better 10 year track record. This is lower than the seventy-eight basis points found by Ayres and Curtis [2015]. The reason for the difference is likely twofold. First, their algorithm is based on a sample basket of lower fee funds without regard for the allocation we are trying to replace. But our approach requires a matching similar fund for any fund to be replaced. If there is not a matching lower fee fund, then original fund remains in the plan and savings are zero. If we pick only those funds from the universe that do have a 10 year track record, then out of 31,598 funds in our universe, only 14,315 funds have at least one lower fee alternative with a Similarity greater than .9 and a better 10 year track record. Thus, significant number of funds do not have any outperforming funds that closely resemble them according to our conditions; while in Ayres and Curtis [2015] all expensive funds can be replaced. Second source of the difference is likely in the fact that we used Schedule H data and assumed the lowest fee share class when share class was not reported. What this also suggests is that active management should not be thrown out altogether, but used selectively when data and research warrants it. Thus, if a plan held class A and did not report it, we would likely map it to one of the cheaper priced R (Retirement) share classes or even an institutional class, if a minimum investment requirement was satisfied. Thus, our results are more conservative, but they support the thesis of the Ayres and Curtis [2015] paper, namely that retirement plans are likely wasting a significant amount of participant's resources.

In our research we have made every effort to be conservative in estimates of possible savings. Clearly, moving all investments to passive strategies is neither prudent nor safe for the market as a whole. However, as is evident from our study – most defined contribution plans in the US are very inefficient and retirees would be much better served by adoption of fiduciary best practices in design of the plan menu. The savings available from applying quantitative approaches to fiduciary best practices could be $17 billion annually as a conservative estimate.

References:

Kwak, James "Improving Retirement Savings Options For Employees." University of Pennsylvannia Journal of Business Law, 15 (2013), pp. 483-540.

Ayres, Ian, Quinn,Curtis. "Beyond Diversification: The Pervasive Problem of Excessive Fees and 'Dominated Funds' in 401(k) Plans." The Yale Law Journal, Vol. 124, No. 5 (2015), pp. 1476-1552.

Appendix A – Additional Metrics of Fund Similarity

Replacement Funds

An important consideration for replacing expensive funds in any plan is the quality and suitability of suggested replacements. Obviously, just because a cheaper investment option is available does not mean that it is a better alternative. Exhibit 6 below shows examples of selected target funds and cheaper alternatives across various asset categories. All alternative choices outperform the target funds over 1,3 and 5-year investment horizons and have Similarity scores higher than .9. Also note that not all target funds have 5 alternate funds that both outperform the target and have high Similarity scores (see IG Bonds for example).

Exhibit 6 – Expense Ratios of Target funds and Alternatives by Asset Category

	Asset Category								
Fund	**US Large Cap**	**US Small Cap**	**MSCI EAFE**	**REIT Fund**	**US Agg Bond**	**IG Bond**	**TIPS**	**Target Date Fund 2025**	**Balanced Fund**
Target	**1.00%**	**0.37%**	**1.06%**	**0.85%**	**0.45%**	**0.45%**	**0.42%**	**0.70%**	**0.87%**
Alt 1	0.03%	0.09%	0.09%	0.07%	0.05%	0.09%	0.07%	0.15%	0.30%
Alt 2	0.03%	0.09%	0.09%	0.12%	0.05%	0.10%	0.12%	0.27%	0.45%
Alt 3	0.03%	0.15%	0.12%	0.12%	0.07%	0.20%	0.15%	0.36%	0.54%
Alt 4	0.03%	0.20%	0.13%	0.18%	0.07%		0.19%	0.40%	0.71%
Alt 5	0.04%	0.25%	0.13%	0.25%	0.07%		0.20%		

Exhibit 7 shows the similarity scores for the alternative fund choices. Unsurprisingly, funds that are primarily comprised of OTC traded instruments tend to have similarity scores that are lower than funds comprised primarily of exchange traded instruments.

Exhibit 7 – Similarity Score of Fund Alternatives by Asset Category

	Asset Category								
Fund	**US Large Cap**	**US Small Cap**	**MSCI EAFE**	**REIT Fund**	**US Agg Bond**	**IG Bond**	**TIPS**	**Target Date Fund 2025**	**Balanced Fund**
Alt 1	0.99	0.99	0.97	1.00	0.92	0.92	0.93	0.99	0.93
Alt 2	0.98	0.99	0.92	1.00	0.93	0.92	0.93	0.98	0.96
Alt 3	0.99	0.98	0.98	1.00	0.92	0.91	0.97	0.96	0.94
Alt 4	0.98	0.98	0.98	1.00	0.93		0.93	0.98	0.93
Alt 5	0.99	0.98	0.97	1.00	0.92		0.95		0.95

Not all funds have similarity scores above the 0.9 threshold. For example, a typical high yield fund or certain active equity managers that have few holdings or particularly high concentrations in a few securities will not have replacement funds that have similarity scores above 0.9.

REIT Fund Example

For an asset class example, we turn to the REIT Funds, where our target fund with an expense ratio of 85bps has five alternative funds which are all priced less than 25bps and have similarity ratings of 1.0, meaning that the replacement funds are almost identical to the target fund. Reducing fees from 85bps to 7bps for a $1M initial investment pool with gross returns of 6% p.a. will enable an investment in Alternative 1 to grow $127k more over a 10-year period than an investment in our target fund, assuming returns are identical. As Exhibit 7 illustrates, the paths of the two investments have almost identical since the beginning of 2011.

Exhibit 7 – Investment Returns of REIT Funds with Similarity scores of 1.0

As the starting point chosen to measure returns can bias a sample, we also examine rolling periods. Exhibit 8 uses a rolling 252-day period to approximate 1-year of daily returns. While on average, the target fund has done an admirable job of recuperating a significant amount of the fees paid, the after fee net return is still worse *on average* across the board.

Exhibit 7 – Rolling 252 Day return Metrics of Target REIT Fund vs. Alternatives

	Alt 1	**Alt 2**	**Alt 3**	**Alt 4**	**Alt 5**
Average	-0.16%	-0.11%	-0.10%	-0.13%	-0.01%
Median	-0.25%	-0.07%	-0.06%	0.05%	-0.08%
Max	3.99%	3.95%	3.30%	3.68%	4.10%
Min	-5.40%	-3.84%	-3.82%	-4.34%	-5.28%
Std Dev	1.38%	1.15%	1.13%	1.21%	1.37%

For example, the Target REIT Fund is 78bps more expensive than Alternative 1 and should underperform Alt 1 by 78bps if gross returns and expenses were equal. Instead the Target REIT Fund only underperforms by 16bps on average in any given 252-day period. So while the fund does make up a portion of the fee, the investor is still better off with Alternative 1 in general.

Appendix B: Holdings Based Data For Funds

Holdings data for mutual funds comes from forms N-Q and N-CSR (for ETFs from the provider websites). Those forms are parsed from the SEC EDGAR website and Microsoft Full Text Search (SQL Server) is used to map security names to tickers for equity funds. Fixed income instruments are recognized by the issuer, coupon, maturity which are present in all EDGAR forms.

Every instrument is then added to commercially available RiXtrema Multi-Asset Class Risk model. Equities are added by stepwise regression to calculate sensitivities to factors like equity beta, liquidity, size industry. Fixed income instruments are added via pricing mechanisms to calculate sensitivities to different points on the yield curve and sector/rating buckets for credit risk. Out of 31,589 funds in our universe, 21,288 funds have holdings information that was readable.

Appendix C: Forward Looking Correlation

Our goal is to calculate correlations between two funds to determine how similar they are, but do it on a position level. The reason we need to calculate this is that for some funds historical returns are not a good indicator of behavior, because they change their profile often. It is also important to calculate position level correlations for fund without long history.

Let us define:

W - weighted average exposure of the fund to a risk factor (for example, to equity beta, effective duration, liquidity etc.)[4]

$$W = w * H$$

Where:

w - vector of weights of each security in a fund

H - matrix of exposures of all securities to each risk factor in a multi-factor model (number of securities by number of factors)[5]

$STD(A)$ - annualized forecasted standard deviation of the fund

G - is the idiosyncratic relationship matrix for fund A vs. fund B. It has number of columns equal to number of positions in fund A and number of rows equal to number of positions in fund B. Each element of the matrix has either zero if a security listed in a given column is different from the one listed in a row or else it has the idiosyncratic standard deviation[6] of the security. For example, in row 4 and column 12 of the matrix we have the intersection of IBM and MSFT. Since these are different stocks, the element in

[4] For a full list of factors and risk model calculations see RiXtrema GML Whitepaper

[5] Risk model is based on Arbitrage Pricing Theory and advanced extension of Modern Portfolio Theory, for more detail read here: http://www.investopedia.com/terms/a/apt.asp

[6] Idiosyncratic risk or idiosyncratic standard deviation is standard deviation of any asset (particularly a stock) that cannot be explained by systematic risk factors included in the model. This is a standard way to model investment risk.

the matrix will be zero. Another element, say row 4 and column 8 has IBM for both. In that case the element of the matrix will be equal to idiosyncratic risk of the IBM.

Holdings Based Correlation (Fund A vs. Fund B) = $\dfrac{W_A{}^T * C * W_B + w_A^T * G * w_B}{STD(A) * STD(B)}$

Appendix D: Historical Correlation

Historical correlations are calculated as a standard Pearson product-moment correlation coefficient. The formula for correlation between funds A & B is:

$$\rho_{A,B} = \frac{\sum_{i=1}^{T}(r_{A,i} - \overline{r_A}) * \sum_{i=1}^{T}(r_{B,i} - \overline{r_B})}{\sqrt{\sum_{i=1}^{T}(r_{A,i} - \overline{r_A})^2} * \sqrt{\sum_{i=1}^{T}(r_{B,i} - \overline{r_B})^2}}$$

Where:

$r_{A,i}$ - return of a fund A at time i

$\overline{r_A}$ - average return of fund A across time

T - number of periods in the sample