Investment Consultants' Claims about Their Own Performance: What Lies Beneath?

Gordon Cookson, Tim Jenkinson, Howard Jones, and Jose Vicente Martinez*

ABSTRACT

Investment consultants market their services by claiming that their fund manager recommendations add significant value. Using detailed data from the leading investment consultants we find no such evidence. A forensic analysis of consultants' disclosures reveals a number of practices that explain their claims: comparisons to benchmarks rather than to peers, inclusion of simulated and backfilled returns, use of rating survivorship conditions, and unexplained exclusions of products from the analysis. We find that recommended products have similar return and risk characteristics to products that are not recommended, but deviate less from their benchmarks.

July 2018

Key words: Investment consultants, recommendations, institutional asset managers, voluntary performance disclosures

JEL classification: G20, G23, G24, L84

* Gordon Cookson (gordon.cookson@fca.org.uk) is with the U.K. Financial Conduct Authority, 12 Endeavour Square, London E20 1JN, U.K.; Tim Jenkinson (tim.jenkinson@sbs.ox.ac.uk) and Howard Jones (howard.jones@sbs.ox.ac.uk) are with the Saïd Business School, University of Oxford, Oxford OX1 1HP, U.K.; Jose Vicente Martinez (jose.v.martinez@uconn.edu) is with the University of Connecticut, School of Business, 2100 Hillside Road, Storrs, CT 06268, USA. The views expressed are those of the authors and not the Financial Conduct Authority. We thank Greg Jackson and Anish Thakrar for their assistance over the course of the research project, and we thank Rick Di Mascio for helpful comments and discussions.

1. Introduction

Investment consultants are the kingmakers of institutional asset management. By assigning recommendations to fund managers, or withdrawing those recommendations, they have substantial influence over which managers are hired and fired by the institutional investors (such as pension funds, insurance companies, endowments, and other large investors)[1] who follow their advice. The investment consulting sector has estimated assets 'under advisement' of some $36 trillion (Pensions and Investments, 2016), with the market share of the top five firms amounting to around 60%. Consultants are used by the majority of plan sponsors in the U.S. (Goyal and Wahal, 2008) and this usage translates into very significant asset flows, although there is a lack of evidence that their recommendations add value (Jenkinson, Jones, and Martinez, 2016). The question therefore arises why plan sponsors follow the advice of investment consultants as much as they do. One explanation is that they are shielding themselves from blame in case their chosen managers perform badly (see Lakonishok, Shleifer, and Vishny 1992; Jones and Martinez 2017). A further possibility is that the way in which consultants present their own ability to pick fund managers is so favorable that it becomes difficult for plan sponsors to ignore their recommendations. In the present paper we explore this latter possibility by comparing the value added by consultants' recommendations with the claims consultants themselves make about them.

In general, investment consultants' recommendations of asset managers are not publicly available, so that independent analysis of their recommendations and of the claims they make about them is difficult. Our analysis is based on a unique data set sourced by the U.K. regulator, the Financial Conduct Authority ('FCA'), which provides detailed records of the institutional asset managers recommended by each of six leading investment consultants between 2006 and 2015. This allows us to identify, for each consultant, when an investment product was first recommended and the period for which it remained so. The six consultants in the sample include three of the largest firms worldwide, with combined market shares of around 45% (Pensions and Investments 2016). Each of these three consultants also produce an analysis of their own performance of the products they recommend, to which we have access.

These data allow us to address the central question posed in this paper: whether the disclosures that consultants present to institutional investors are reliable guides to their past

[1] We refer to such decision-makers as plan sponsors.

performance. This is a critical question, as evidence presented by consultants that their research and due diligence add value is likely to be one of the most important factors that determine whether they are hired. We calculate the weighted average excess returns over benchmarks claimed by the three consultants for the products they recommend to be an impressive 1.73% per year. All three consultants claim significant excess returns, but each uses a different methodology and does not make the underlying data available, so that institutional investors must take on trust the basis for and comparability of these claims.[2] In our analysis we explore whether this trust is well placed.

Before evaluating consultants' claims about how their recommended products perform, we conduct our own performance analysis. We find that, on average, consultant-recommended investment products perform no better than other products available to institutional investors. Our results indicate that, over our 10-year sample period, the portfolio of all products recommended by investment consultants delivered average returns gross of management fees of 5.40% per year (5.11% after management fees). These returns are, on average, 0.30% per annum lower than the returns obtained by other products available to plan sponsors but not recommended by consultants. When comparing recommended products to a matched sample of non-recommended products classified in the same investment category (e.g. "U.S. Large Cap Value Equity"), which allows a more homogenous, risk-matched comparison between recommended and not recommended products, we find that recommended products still trail non-recommended products by 0.21% per year (or by 0.23% per year after management fees). These differences are not statistically significant. The same is true of the difference between the excess return over (manager chosen) benchmarks of recommended and non-recommended products in the matched sample. Our results are robust to a number of variations in the analysis: using a larger sample of consultants, assuming fixed holding periods for products recommended by plan sponsors, and considering negative recommendations in our analysis. Therefore, at least in aggregate returns, investment consultants appear, on average, to have no systematic skills in manager selection.

When we focus on the three large consultants for which we have access to the claims they make to institutional investors, we find large differences between those claims and our independent analysis of their performance. For this sample, we again find no outperformance of recommended

[2] The challenges faced by institutional investors in understanding the strengths and weaknesses of investment consultants, including their track records, have, in fact, spawned an additional layer of advisors that help plan sponsors select investment consultants.

managers, whereas the consultants themselves claim significant out-performance. The three large consultants in our reduced sample calculate and disclose simple average excess returns over benchmarks for all recommended products in an investment category. The weighted average of these claimed excess returns of 1.73% per year exceeds our estimation of their performance by 1.94% or 1.95% per year depending on whether we do the calculation before or after management fees.

In order to explain this significant difference between our estimates of consultants' manager selection skills and their own claims, we undertake a forensic analysis of the performance disclosures made by each of the consultants. This reveals a number of practices that help explain the discrepancy. First, consultants compare the returns of recommended products to manager-chosen benchmark returns rather than to the returns of other products available in the same investment category. The latter is a more appropriate measure of the value added by consultants, since the former also captures any out-performance over benchmarks by the institutional asset management industry as a whole.[3] Second, consultants typically retain simulated and backfilled returns in their performance analysis, which potentially generate an upward bias in the sample of product returns and excess returns over benchmarks. Third, some consultants use restrictions in drawing up the sample they use when they assess their own performance (e.g., requiring the recommended product to remain in their recommendation lists for at least a certain number of full years), leading to survivorship bias in the results. Finally, some consultants tailor the sample used in their analysis (e.g., excluding certain individual investment products or categories of products) in a way that appears largely arbitrary except that it opens a further gap between their analysis and our own calculations.

Some of these practices, like the tendency to compare products against manager-chosen benchmarks rather than other available products, are transparent (even if their relevance may still

[3] As an example, if institutional products, on average, outperform their benchmarks by 50bps gross of manager fees, even a consultant with no manager selection skills would be expected to select products that outperform benchmarks by that margin (also gross of fees). A more appropriate measure of the consultant's contribution would be the extent to which its recommended products outperform benchmarks on top of those 50bps (or on top of what non-recommended products return, the metric we use in this paper). The practice of using product excess returns over benchmarks is particularly misleading if, as Gerakos, Linnainmaa, and Morse (2016) and the present paper show, institutional products, on average, significantly outperform their benchmarks.

be difficult to assess for outside observers). Others, like the exclusion of certain products and investment categories from the analysis or the inclusion of backfilled returns, are opaque to their clients, who are not privy to detailed information about the performance calculations.

These practices, and possibly others, also affect another claim sometimes made by consultants: the percentage of their recommended products that outperform benchmarks over different horizons. Two of the consultants in our sample report this statistic, claiming in one case that 88% of their recommended products outperform their benchmarks over five-year horizons and in the other that 76% of their recommended products do so. There are two concerns about this claim. First, like the claims about performance discussed above, it does not allow for the fact that institutional products as a whole outperform benchmarks. Secondly, survivorship bias is mechanically introduced into such a comparison, since calculations are necessarily limited to recommended products with available five-year forward returns. When we calculate the percentage of consultant recommended products outperforming the median product in their investment category over a five-year period, which addresses the first of these concerns but not the second, we find that around 52% of recommended products outperform the median product.

Our analysis suggests that consultants' claimed performance is not just wide of the mark, but also that their disclosures are poor guides to their relative performance. A comparison between the performances of consultants' recommendations based on their own claims shows that the relationship between their claims and their actual performance is not monotonic.

In our analysis we rely on two kinds of risk adjustment to compare recommended and non-recommended products: we compute benchmark-adjusted product returns, and we benchmark recommended products using non-recommended products classified in the same narrow eVestment category (there are in total 289 such categories). These seem reasonable alternatives given the way products are classified, compared, and monitored by plan sponsors. However, it is possible that benchmarks and investment categories fail fully to capture some differences in risk between recommended and non-recommended products. For instance, plan sponsors may not be able to invest in multiple recommended or non-recommended products. If so, they may be concerned not just about the average performance of recommended and non-recommended products (and their systematic risk), but also about their idiosyncratic risk.

We find no consistent differences between recommended and non-recommended products in terms of return volatility or betas with respect to manager-chosen benchmarks. This suggests

that there are no significant differences in terms of total and systematic factor risk between the two sets of products, or any significant differences in leverage between the two. However, there is a significant, albeit small, difference between recommended and non-recommended products in terms of tracking error volatility. Recommended products consistently exhibit lower tracking error volatility than non-recommended products both pre- and post-recommendation, suggesting that recommended managers, on average, deviate less from their benchmarks than non-recommended ones.

It is unclear if maintaining low tracking errors, but not low return volatility, is beneficial to plan sponsors. Roll (1992) notes that lower tracking errors may make performance monitoring easier, and he also raises the possibility (about which he is skeptical) that they help in a decentralized portfolio management context.[4] However, even when using information ratios (a measure sometimes employed by practitioners and defined as the ratio of excess returns over benchmarks and tracking errors) we find that recommended products do not outperform: their tracking error is lower but so are their excess returns over benchmarks in a matched sample, with the two measures offsetting each other.

Some of these results echo the findings of Jenkinson, Jones, and Martinez (2016); however, the present paper expands the analysis in several directions. First and foremost, we are able to compare consultants' recommendation performance with the way in which that performance is represented by the same consultants in their communications with (current and prospective) clients, and we explore the reasons behind the differences. Second, we look not only at the average performance of products recommended and not recommended by consultants, but also at their distribution and at the amount of idiosyncratic risk incorporated in these products. In doing so, we explore the possibility that consultants may be intentionally avoiding idiosyncratic risk in their recommendations. Third, we look at the performance of positively recommended products but also analyze negatively recommended ones. This allows us to explore the possibility that consultants, even if unable to identify top performing products, may still be able to identify and steer their clients away from poorly performing asset managers. Finally, our analysis includes a significantly

[4] For discussions of decentralized portfolio management see Sharpe (1981), van Binsbergen, Brandt, and Koijen (2008), and Blake, Rossi, Timmermann, Tonks, and Wermers (2013). The desirability of staying close to benchmarks can be called into question by the findings that funds that deviate more from their benchmarks or group of peers tend to outperform those that do not (see, e.g., Cremers and Petajisto (2009) and Sun, Wang, and Zheng (2011)).

broader sample of asset classes (not just U.S. equities), makes use of more precise data (for a small number of consultants), and benefits from more accurate information on backfilled and simulated returns.

The paper is also related to the work of Goyal and Wahal (2008), who study the hiring and firing decisions of plan sponsors partly as a function of whether they employ consultants or not. Goyal and Wahal (2008) find that, although consultants add value to hiring decisions on average (consultant-advised decisions have higher post-hiring returns than non-advised decisions), they destroy value in advising large plan sponsors. Unlike Goyal and Wahal (2008), we do not benchmark consultants against plan sponsors (in their comparison, consultants' manager selection ability is better than that of small plan sponsors but worse than that of large ones), but directly benchmark consultant recommended products against non-recommended products.[5] One way to reconcile our results with theirs is to argue that consultants are not necessarily good at identifying superior performance, but are better than small plan sponsors who make decisions on their own and worse than large plan sponsors.

Finally, our analysis of the way in which consultants represent their past performance to clients and prospective clients is related to Patton, Ramadorai, and Streatfield (2015), who analyze hedge funds reporting and subsequent revisions of their own past performance to financial databases, and Ljungqvist, Malloy, and Marston (2009), who study financial analyst reporting of their stock recommendations and subsequent changes to their recommendation records in a leading database. Investment consultants do not report their past performance to databases, but they share it directly with current and prospective clients. However, both those studies and ours investigate the ways in which financial intermediaries exploit opportunities to present their past performance in an unduly favorable light.

[5] Another difference between our study and Goyal and Wahal (2008) is that we do not restrict our analysis to a window or period following firing and hiring decisions.

2. Data

Our main data source is a record of the institutional asset managers recommended by each of six investment consultants between 2006 and 2015.[6] More precisely, the recommendations relate to 'investment products', that is, the investment styles/categories in which an asset manager offers its services. A product may be offered to investors in different vehicles, such as a pooled fund or a segregated account. This data, which was obtained by the U.K. Financial Conduct Authority in connection with a regulatory market study in 2016, allows us to identify, for each consultant, when an investment product was first recommended, the period during which it remained recommended and, if applicable, when the product was downgraded to a non-recommended rating. The recommendations cover a range of asset classes, investment styles, and regions.

The six consultants in the sample include three of the top six worldwide; their identity cannot be revealed but the combined worldwide share of 'assets under advisement' for the largest three in our sample is 45% according to Pensions and Investments (2016). The other three are smaller, U.K.-focused firms which were included so that possible differences in the recommendation patterns of large and small consultants were reflected in our analysis. For each of the three large consultants we also have access to their own analysis of the performance of the products they recommend. This self-analysis forms the basis of consultants' marketing to asset-owners and in Section 3 we compare it with our own analysis of their performance. In the case of two of these three large consultants we have a record of negative as well as positive recommendations; negatively recommended products are those from which consultants recommend disinvestment, as opposed to being merely not recommended. The data requested from investment consultants also includes information on the charges they make to asset owners. Based on this we estimate that consultants' average annual fees to asset owners for advisory services (which are not limited to providing asset manager recommendations) range from 9 basis points for assets under management of less than $70m to 2 basis points for assets under management of around $1bn. All in all, the detailed, consultant-by-consultant record we have of recommendations

[6] For the years 2008–2010 the number of consultants in the sample is actually seven. During 2010 two of the seven merged, and one of these two did not provide data from before 2008.

makes our data set distinct from those of previous studies of investment consultants, which have not had access to consultants' recommendations at all or have relied on aggregated data.[7]

Our second data source is eVestment, a leading data provider to the asset management industry, which collates self-reported data from institutional asset managers with aggregate assets under management worth more than $37 trillion as at the end of 2015.[8] From this database we derive, for the period 2006–15 and for each institutional product, the quarterly returns together with the returns of a manager-specified benchmark. For each product the database also provides cross-sectional information (as at the end of 2015) on the asset class of the product and pro-forma charges (eVestment does not hold historical charges data).[9] The asset class is designated at two levels in the eVestment database: there is a broad classification into one of six categories and, within each of these, a more granular classification reflecting, among other things, investment style and geographical focus. The pro-forma charges, which we have for around three-quarters of the observations, are based on an investment of $50 million.[10] The returns data are composite, meaning that individual returns earned by each client invested in that product may deviate from the composite returns, but we have been informed that deviations are typically small. Composite returns are net of trading costs but gross of investment management fees.

Data on products which have been discontinued (e.g. because they have been acquired or closed) are retained in the database by eVestment, making the database free of survivorship bias. The database has flags indicating backfilled returns and simulated returns. The 'backfill' flag relates to products for which the inception date is earlier than the date at which the product was

[7] For example, Goyal and Wahal (2008) have data on whether plan sponsors use a consultant or not; and Jenkinson, Jones and Martinez (2016) and Jones and Martinez (2017) have aggregated data on consultants' recommendations, but without a breakdown by consultant.

[8] eVestment does not collect data on some smaller investment products, notably infrastructure funds, nor on funds self-managed by asset owners.

[9] Jenkinson, Jones and Martinez (2016) compare the latest fees in the eVestment database with the fee data available from IIS (Informa Investment Solutions), a firm which does record historical fee data. They find no significant difference between these data sources for the final year in their sample, and they find in IIS very little time-series variation in fees. This provides justification for applying the latest eVestment fees for each product to earlier years when computing net returns.

[10] These are in most instances the fees charged for separate accounts; if a product is unavailable in a separate account we use the fees charged for comingled or mutual funds.

added to the eVestment database; in such cases the returns in the intervening period are flagged. The 'simulated' flag indicates returns which a manager backdates from before the inception date of a product. We use these flags to control for possible bias in the reported returns, as we explain in detail in section 3. This is especially important in the case of returns with a 'backfill' flag, which represent 42.1% of our sample. Only 0.4% of the returns are simulated, and we exclude these from our analysis. We include in the sample only products which report gross returns to eVestment and which are classed by eVestment as institutional or alternatives.[11]

Table I and II contain descriptive statistics on our sample. As Table I shows, the mean number of available products during the sample period is 12,706, managed by 1,682 asset managers, making 7.6 products per manager; these numbers remained relatively stable during the sample period. As for recommendations, in our full sample of six consultants there were an average of 1,736 recommendations per year; this rose steadily during the period, from 931 in 2006 to 2,302 in 2015. Divided by the number of consultants in our full sample, this corresponds to an average of 282 recommendations per consultant. We show separately the reduced sample of the three consultants whose own performance self-analysis we compare with our own in Section 3. Reflecting their large size, the reduced sample averages 1,614 recommendations per year, or 91.5% of the total sample, with each consultant averaging 538 recommendations a year.

Table II shows the breakdown of products and recommendations by asset class. The five classes of Equity, Fixed Income, Balanced/Multi Asset, Real Estate, and Alternatives/Hedge Funds reflect the classification of eVestment, except that we have merged the numerically small Alternatives and Hedge Funds classes. For each asset class we indicate the number of investment subcategories included in that class; there are 289 of these in total. Equities is the most important asset class by number of investment subcategories, products (60.3% of all products), asset managers, and recommendations (58.6% of all recommendations). The next largest class is Fixed Income, and together Equities and Fixed Income account for 89.1% of all products and 93.1% of all recommendations. Table II also shows the percentage of observations in the sample for which asset manager fee data is available – around three-quarters for Equities and Fixed Income and between 30% and 60% for the other asset classes.

[11] This means that we exclude Exchange Traded Funds (ETFs), Separately Managed (or 'Wrap') Accounts (SMAs), and unclassified accounts; together these represent only 6% of total observations in the original sample.

In contrast to retail mutual funds that are rated by firms such as Morningstar, products rated by investment consultants do not have unique and widely-used identifying codes such as an ISIN or Sedol, so matching the data sets for the present paper was done manually. Given the considerable variation among consultants in their naming conventions, and to minimise the possibility or error, each consultant was requested to identify the product in the eVestment database to which each recommended product corresponded.[12]

3. The Performance of Consultant-Recommended Investment Products

A. The actual and advertised performance of consultant recommendations

In our study we have access to information (with varying degrees of detail) about how three leading investment consultants present the performance of their manager recommendations. All three consultants in our sample calculate simple average excess returns over benchmarks for all recommended products in a number of investment categories defined by themselves (e.g. "U.S. equity"), and they present value added results for each of those investment categories in their marketing materials.[13] Some consultants use a simple average of these investment category excess returns to arrive at composite or broad asset class results, such as all equities, all fixed income, etc.

In order to summarize the advertised recommendation performance of the consultants in our sample into a single measure, we first aggregate the composite excess returns reported by each of these consultants for each of three broad asset classes (equity, fixed income, and a class which we designate 'other' and which includes balanced/multi-asset, alternatives/hedge funds, and real estate) into a single per-consultant performance measure. If a consultant does not report composite excess returns, we compute them ourselves using the simple average of the consultant-reported excess returns over benchmarks for all investment categories covered by the consultant in the asset class. The results we use are those made available by consultants as of January–March 2016. For each consultant we obtain an aggregate performance value, by weighting the consultant's performance within each asset class by the number of products recommended by the consultant in

[12] For a small number of negative recommendations, we conducted the match manually.

[13] Excess returns are usually based on returns gross of asset manager (and consultant) fees. While most consultants report excess return averages, some also report medians. Some consultants calculate a geometric time series average across these average returns, for different time horizons.

our dataset in that asset class. We then weight each consultant's performance values based on the number of recommendations issued by that consultant, to arrive at a weighted average performance value across all consultants. For our sample of consultants this results in a weighted average excess returns over benchmarks for the recommended products of 1.73%. Equally weighting composites and consultants produces similar results.

We assess the accuracy of this claim using four alternative measures. All four measures involve comparing the returns, or excess returns over benchmarks, of recommended products against those of non-recommended products. This is a more appropriate measure of the value added by consultants than the simple excess returns over benchmarks for recommended products, which captures the joint contribution both of consultants' recommendations and of the institutional asset management industry. It is also simple and easily implementable: it allows the performance of any product to be evaluated and is not limited to equities, which is particularly important because there are many asset classes in our sample for which there are no well-established risk models.

The first measure we use is the difference between the average returns on recommended and non-recommended products. Formally, for each consultant and quarter *t*:

$$\sum_i r_{i,j,t} I_{i,t-1} - \sum_i r_{i,j,t}\left(1 - I_{i,t-1}\right) \qquad (1)$$

where $r_{i,j,t}$ is the return of product *i,* belonging to investment category *j* in quarter *t*, and $I_{i,t-1}$ is an indicator variable that equals one when the product is recommended in quarter *t* and zero otherwise. To obtain this measure we create equal-weighted portfolio returns of recommended and non-recommended products available in each quarter. The recommended portfolio includes each product as many times as it is recommended. In this analysis, we use products that report returns gross of asset manager fees and that are classed as institutional or alternative by eVestment (these are the products aimed at institutional investors, and the ones which our recommendations overwhelmingly match to). We also exclude simulated and backfilled returns, using two separate eVestement flags for this purpose. In forming the two portfolios we rely on the status of each product at the end of the quarter preceding the return measurement period. In this way, we avoid capturing returns preceding the initiation of a recommendation, which can happen anywhere in the

quarter. The returns thus obtained, and the resulting difference, are not risk-adjusted in any way, so that differences in returns between recommended and not recommended products may reflect differences in risk between the two groups as much as differences in underlying quality.

The second measure we use is the difference between the average excess return over manager chosen benchmarks for recommended and non-recommended products. Formally, for each quarter *t*:

$$\sum_i (r_{i,j,t} - b_{i,t}) I_{i,t-1} - \sum_i (r_{i,j,t} - b_{i,t})(1 - I_{i,t-1}) \qquad (2)$$

where $b_{i,t}$ is the return of a manager chosen benchmark for product *i* in quarter *t*. We obtain this measure following the same procedure as described above but using returns in excess of a manager chosen benchmark, available in the eVestment database, rather than raw returns for both sets of products. Unlike the previous measure, this one includes a measure of risk adjustment, but may lend itself to gaming by asset managers as it relies on benchmarks chose by the asset manager which may or may not be appropriate for the product in question.

The third measure we use is similar to the first but uses a matched sample of recommended and not recommended products classified in the same eVestment category for comparison. Formally, for each quarter *t*:

$$\sum_i \left(r_{i,j,t} I_{i,t-1} - \sum_{i \in j} r_{i,j,t} (1 - I_{i,t-1}) \right) \qquad (3)$$

In this case risk adjustment is achieved by comparing recommended products against other (non-recommended) products of similar risk and characteristics according to eVestment (which classifies them into 289 different categories). More precisely, the matching is achieved by pairing every recommended product in each quarter with an equal-weighted portfolio of all non-recommended products in the same eVestment investment category in the same quarter. In this way we compare recommended and not recommended products category-by-category. This

matching approach mirrors the way in which plan sponsors choose managers: first deciding on the category of product to invest in, and then on the product within that category.[14]

The fourth measure also uses a matched sample of recommended and non-recommended products, but in this case we base the comparison on returns in excess of manager-chosen benchmarks rather than raw returns. Formally, for each quarter *t*:

$$\sum_{i}\left((r_{i,j,t}-b_{i,t})I_{i,t-1}-\sum_{i\in j}(r_{i,j,t}-b_{i,t})(1-I_{i,t-1})\right) \qquad (4)$$

This may offer a more granular way of risk-adjusting returns, as, even within a given eVestment category, managers may regard different benchmarks to be more appropriate for their product, but it could allow asset managers to game their benchmark choices (which are manager chosen, unlike the group of peers selected by eVestment).[15]

We employ two versions of the above performance measures, one based on gross returns and another based on net returns. To compute net returns-based measures, we subtract one-quarter of the annual pro forma fee based on a $50 million investment from the product's quarterly return, using the fee information from eVestment. We do not make any adjustment for investment consultant fees.

The results in Table III indicate that, over our 10-year sample period the portfolio of all products recommended by investment consultants delivered average returns gross of management fees of 5.40% per year (5.11% after management fees). These returns are, on average, 0.30% per annum lower than the returns obtained by other products available to plan sponsors, which are not recommended by consultants. When we risk-adjust returns using manager chosen benchmarks, recommended products obtain an average excess return before management fees of 0.45% per year

[14] This approach is conceptually similar to the active peer benchmark approach of Hunter, Kandel, Kandel, and Wermers (2014), except that we rely on a more granular classification of investment products and implicitly assume a beta of one with respect to this group.

[15] While asset managers' revealed preferences could be a good indication of their chosen peer group, Sensoy (2009) provides evidence that in the context of mutual funds the free choice also induces fund managers to choose easy-to-beat benchmarks that do not match the risk profile of their investment product. In the context of institutional products one would expect this problem to be attenuated (but not eliminated) by the penalty of increased tracking error, which, as Del Guercio and Tkac (2002) argue, institutional investors pay attention to.

(-0.32% per year after management fees), once annualized. This is indistinguishable from the 0.33% per year (-0.32% per year after management fees), obtained by non-recommended products. The return difference between recommended and non-recommended products in the matched sample (a more homogenous comparison of recommended and not recommended products in the same investment categories), at -0.21% per year (or -0.23% per year after management fees) is not statistically significant either. The same is true of the difference between excess return over manager chosen benchmarks of recommended and not recommended products in a matched sample (differences of -0.21% before fees and -0.26% after fees).

The difference between recommended and non-recommended products returns or excess returns over benchmarks (whether in a matched sample or not) is in most cases negative but not statistically significant and not very different from zero. This suggests investment consultants have, on average, no real ability to pick investment products. On the other hand, claimed performance as we have seen is quite substantial, at 1.73% per year for the sample of consultants analyzed. Claimed minus actual figures are computed as the difference between the claimed average recommended product excess returns over benchmarks across consultants and each of our four measures of recommendation performance. These differences are large and statistically significantly positive.[16] The difference between claimed and actual ranges between 1.61% and 2.03% per year depending on the specification and whether we do the calculations before or after management fees.

This pattern is not driven by a single consultant or a single asset class, but seems to be present across the board to varying degrees. When looking at the consultants individually, their claimed performance (reported in Table IV) ranges from 1.64% per year to 2.51% per year (using the same weighting principles described previously), yet their actual performance according to our calculations is significantly lower and varies between -0.95% per year and 0.30% per year depending on the consultant and the method used to asses that performance (excess returns over benchmarks or matched samples). Table IV also shows the results of splitting the sample into three major asset class groups: equity, fixed income, and other (balanced/multi-asset, alternatives/hedge funds, real estate). All three consultants claim to deliver positive excess returns over benchmarks

[16] We assess the statistical significance of the difference between claimed and observed performance by taking advantage of the time series variation of observed excess returns and treating claimed performance as fixed.

in these three aggregates, with the largest average for the equity class (the largest and most important class) at 2.21%, and the lowest for fixed income at 0.93% per year. We do not, however, find any evidence of outperformance in any of the three classes, independently of the method used. The differences between claimed and observed performance are always positive, and frequently statistically significant.

B. What is behind the differences between the claimed and observed performance of consultant recommendations?

In order to understand the differences between the claimed and observed performance of consultants' recommendations, we undertake a forensic analysis of the performance disclosures made by each of the consultants in our reduced sample. We find that the following practices employed by consultants in assessing their own performance explain the discrepancy with our own calculations: comparison of recommended products to benchmarks rather than to other products available in the same investment category; inclusion of simulated and backfilled returns in the performance analysis; use of restrictions leading to survivorship bias in drawing up the recommendation sample used in the assessment; and arbitrary exclusions of products and investment categories in calculations. Some of these practices, like the tendency to compare products against manager chosen benchmarks rather than other available products, are transparent but still difficult to assess for outside observers. Others, like the exclusion of certain products and investment categories from the analysis, are opaque to asset owners, who are not privy to detailed information about the performance calculations. We list the prevalence of these practices in our reduced sample of consultants in Table V, where we also quantify the cumulative impact of these practices on the performance claimed by consultants.

It is commonplace among practitioners to assess the recommendation performance of investment consultants by comparing the returns of recommended products against the returns from manager-chosen benchmarks, and this practice is followed by all of the consultants in our reduced sample. It is, however, an unsatisfactory practice in that it does not capture the value added by consultants but the joint contribution to investors of consultants and the institutional asset management industry. As a parallel, when studying the performance of a mutual fund, segregated account, or other investment product, it seems reasonable to use as a benchmark a portfolio of stocks, bonds, or other assets that could be included in that portfolio. By the same reasoning, the

appropriate benchmark for consultants, who make recommendations from among investment products rather than stocks or bonds, is a portfolio of all institutional investment products which could be recommended. Such a benchmark is not the same as the manager chosen benchmark.[17]

The practice of using product excess returns over benchmarks is particularly problematic if, as Gerakos, Linnainmaa, and Morse (2016) show (see also Busse, Goyal, and Wahal (2010) and Jenkinson, Jones, and Martinez (2016)), institutional products, on average, outperform their benchmarks.[18] In that case a comparison with manager-chosen benchmarks rather than with other institutional products would misleadingly imply that consultants add value when in fact their recommended products may do no better than the average products available to plan sponsors.

As we show in Table V this practice has a significant impact on results, as the performance of consultants appears significantly better using manager-chosen benchmarks: it improves by 0.57% per year for the first consultant (IC1), 1.58% for the second (IC2), and 0.72% for the third (IC3). However, this alone does not explain the difference between claimed and observed performance, because even if we were to measure consultants' performance using the excess returns of products over manager-chosen benchmarks, the differences from the results claimed by consultants would remain large and significant, ranging from 0.93% per year for IC1, to 1.88% and 1.27% per year for IC2 and IC3, respectively.

Another source of the discrepancy between our results and those claimed by investment consultants is the inclusion of simulated and backfilled returns in consultants' calculations. While the reasons for excluding simulated returns are obvious, those for excluding backfilled ones can

[17] Our benchmark is not all available products, but all those products not selected by the consultants. Using one or the other should lead to the same results, as all products are just a weighted average of recommended and not.

[18] This positive performance is consistent with institutions being sophisticated investors (Del Guercio and Tkak, 2002, Cohen, Gompers, and Vuolteenaho, 2002). While most evidence suggest that retail mutual funds earn gross alphas close to zero (Fama and French, 2010), some papers that compare institutional and retail products (James and Karceski (2006) and Evans and Fahlenbrach (2012)) provide evidence that suggest that institutional products are able to do better. An advantage of institutional products, and in particular separate accounts, is that they do not suffer a drag (or at least do not suffer it to the same extent) due to the need to provide liquidity to investors that open ended mutual fund provide (Edelen, 1999). A possible reason why institutional products outperform cross-border indices is that some of these indices assume that firms paying dividends to international investors withhold or deduct tax from dividends. However, many international investors are able, thanks to double taxation treaties, to reduce the amount of tax withheld or deducted below the level assumed by the indices; see FCA (2017).

be more subtle. From the perspective of the consultants, it would seem sensible to include in their calculations the backfilled returns of recommended products. Clearly the recommendations themselves were not backfilled, and the returns and excess returns over benchmarks of those recommended products are real. The problem, however, is that databases (including those used by consultants) are usually incomplete and filled in over time, and the new products that are added to them are more likely to report past returns if those returns are good (Malkiel and Saha, 2005). As a result, series of poor returns are likely to be under-represented in the backfilled sample. The problem then is not the inclusion of backfilled returns for recommended products in itself, but the failure to include in the calculations returns of recommended products that were neither originally reported to the database nor backfilled later. The obvious solution to avoid any biases is to exclude all backfilled returns.

Table V shows the cumulative impact of both looking at excess returns over benchmarks and not discarding simulated or backfilled returns. On average, as seen by comparing the second and third rows in the row part of the table, excess returns over benchmarks increase by 0.29% per year for our sample of consultants if we allow backfilled and simulated returns in the calculations. This brings us closer to the figures claimed by consultants (and further from our own measure of performance), but still nowhere close to consultants' numbers.

At least one of the consultants in our sample (IC1) employs a third practice that is likely to generate a significant upward bias in its performance results. IC1's self-assessment takes into account only the excess performance of those managers that were given a top rating for fixed periods of time (typically five years) or longer. Therefore, if IC1 downgraded a manager after, say, four years and nine months of receiving a top rating, that manager would not appear in the five year assessment. This practice, which involves a look-ahead condition, is problematic in light of the tendency of consultants to base their recommendations partly on the past performance of products (Jenkinson, Jones and Martinez, 2016).[19] This approach increases IC1's performance by 0.27% per year so that, together with the other two effects outlined above (shared with IC2 and IC3), the performance of IC1 is boosted by 1.16% per year.

[19] This methodology could also give rise to an incentive to downgrade highly rated managers that perform poorly ahead of the five year milestone, further increasing the survivorship bias.

For two of the consultants in our reduced sample (IC1 and IC2), we have access to the list of products/strategies used in their self-assessments. An analysis of such data reveals that not all products and investment categories are included in their calculations. In principle the exclusions could be random or, even if not random, may not reflect an attempt to improve the performance claimed by the consultant (for instance, a consultant could tailor the disclosure to investors interested in some specific geographic region). However, we find that 81.2% of the recommended products open to investors and actively reporting to eVestment, and 81.6% of recommended products closed to new investments are included in these consultants' own analyses. In contrast, only 51.1% of recommended inactive products (liquidated, merged/acquired, restructured, or no longer providing data to current or prospective investors) are included, which suggests that the samples used by consultants in their self-assessment may be biased against these types of products.

In light of these findings we re-compute these two consultants' results using only the products that they list in their calculations. The matching of products used in marketing materials to eVestment categories is sometimes problematic (especially for consultants that do not use eVestment), but any deviations should be small and random unless the exclusions/inclusions are strategic in nature. After performing these adjustments, the performance of IC1 rises to 1.19% per year and that of IC2 to 1.35% per year. In the case of IC2, this adjustment thus adds 0.41% per year to the claimed performance over and above the effects outlined above which are shared with the other consultants.

We have been able to explain most of the discrepancy between our numbers and those produced by consultants (70% for IC1, 67% for IC2, and 56% for IC3). The residual difference could be the result of a number of factors: different return samples used, mismatches in asset class definitions and performance measurement periods (perhaps strategically chosen), differences in weighting strategies (asset class, product, and/or recommendation), matching of recommendations to products in databases, and product and investment categories exclusions (for the one consultant for which we do not have access to this information).[20] If the residual is small and not statistically significant it should signal that there is no other major source of discrepancy (at least none that is

[20] The matching of recommendations to products in the eVestment database was done by the FCA using input from the consultants. The degree of detail of the material provided varies from consultant to consultant.

strategic).[21] The residual difference between the performance claimed by consultants and the performance we calculate varies between 0.45% and 1.15% per year, that is, even after we have accounted for the various approaches by consultants outlined above. For two of the consultants this residual is significant at the 1% level, suggesting that, at least for them, the practices we have analyzed are not the full story.

Overall, we find significant differences between consultants' claimed performance and our own calculations. Most, but not all, of those differences can be attributed to practices by consultants in their self-assessment which we have identified. A comparison between consultants based on their own claimed performance shows that there is no monotonicity between consultants' claims and their actual performance, suggesting that consultants' disclosures may not be good guides to their relative performance either.

C. Larger consultant sample, fixed holding periods, and negative recommendations

The analysis in the previous two subsections is based on the (positive) recommendations of a reduced sample of three leading consultants for which we have detailed access not just to their recommendations, but also to the way in which they present their recommendation performance to current and prospective clients. In this subsection we extend this analysis in a number of ways.

First we explore the recommendation performance of a larger sample of seven consultants, including some smaller consultants for which we lack information about their own views on their performance. The results of this analysis are presented in Table VI. This table shows the performance of investment products recommended and not recommended by consultants and the difference between the two, using all seven consultants and the same performance metrics as before. The first four columns show results assuming that products are held only for the period they are recommended (or not) by consultants, as in the rest of this study. A comparison with the results obtained for the reduced sample of consultants (shown in Table III) reveals that the

[21] In our performance analysis we lag the recommendation data one period because in principle the product may be recommended any time during the quarter, and it would be inappropriate to assign performance to a consultant based on returns accrued before the product was added to the recommendation list. It is unclear whether consultants do the same. This practice, however, has negligible impact on results (2 basis points per year at most) and may very well be justified in the case of some consultants (if they include the products in the recommendations lists at the beginning of the quarter).

recommending performance of both groups is very similar. In gross returns (shown in the first part of the table) differences in the performance of recommended products between the full and reduced samples of consultants are never larger than 10bps per year. When using returns net of asset management fees, and assuming a notional investment of USD 50 million (but not making any adjustment for investment consultant fees), differences between both groups are also small and never larger than 13bps per year. These results suggest that the reduced sample is quite representative of the investment consulting industry at large, which is perhaps unsurprising, given that it is based on consultants with a combined share of close to 50% of the entire investment consulting market.

In contrast to retail investors, who are prone to long periods of inattention, plan sponsors (asset owners) are unlikely to leave their assets with an asset manager through neglect. Indeed, consultants are there to remind plan sponsors to transition out of products that are no longer recommended. However, transition is costly for plan sponsors, who may therefore be disinclined to switch asset managers, so it is meaningful to explore the performance of recommended products over fixed time horizons. In this vein, the last four columns of Table VI show the returns and excess returns over benchmarks for portfolios of recommended and non-recommended products which are formed and assumed to be held by investors for a fixed period of five years starting each listing month (or until the end of our sample period). A comparison of both sets of columns in Table VI reveals almost no differences in results between these two holding period assumptions.

Many investment consultants not only recommend investment products to their clients but also, having performed due diligence on fund managers, assign to some products a rating similar to the buy-hold-sell classification used by equity analysts. In this scheme, a sell or negative rating is understood as advice to divest from a manager or (given the costs of transition) to place that manager on a watch list unless the rating is reversed within a short time. We therefore also analyze negatively rated products. Such an analysis could help us assess consultants' ability to identify underperforming products, even if they lack ability to pick outperformers. Note that non-recommended products in our analysis above (or in Jenkinson, Jones, and Martinez, 2016) include not only negatively rated products but also neutrally rated and unrated products. Analyzing negatively rated products is not without problems, as negative ratings are assigned less frequently than positive ratings. In fact, some consultants assign a negative rating to a product only until their clients divest their holding in that product, at which point they stop rating it. Although negative

ratings were assigned less frequently and for shorter periods of time than positive ratings in our sample (particularly in certain asset classes, such as alternatives and hedge funds) and the quality of the negative ratings data was significantly lower (fewer negatively rated products, lower matching rate to the eVestment data, and shorter history), we found enough instances of negative ratings to warrant an analysis.

Table VII shows the average performance of consultants' negatively rated products and the difference between these and all other products in our sample. In raw return terms, negatively rated products' average gross returns are 0.62% per year higher than the returns of all other products in our sample. That difference does not, however, take into consideration differences in the risk profile of these two sets of products, and is not statistically significant. When comparing excess returns over manager-chosen benchmarks there is hardly anything to separate the performance of negatively and non-negatively rated products, which seems to suggest that consultants' are also unable to identify underperforming products. When comparing negatively and non-negatively rated products in a matched sample (that is, category by category), we find that negatively rated products do worse than other products. The differences, however are all small and within the margins of statistical error. The analysis is based on a sample which, in terms of products and investment categories, is substantially smaller than that used for positive recommendations (the matched sample loses 55.2% of the primary universe/quarter observations compared to the analysis of positive recommendations) and is heavily tilted towards equity and fixed income. The loss of observations, at 94.2%, is particularly marked in the alternative/hedge fund asset class, where there are almost no negative recommendations in our sample, balanced/multi asset (78.8% loss), and real estate (77.7% loss).

D. Other performance disclosures

Another statistic sometimes reported by consultants is the percentage of their recommended products that outperform benchmarks over different horizons. Two of the consultants in our sample report this statistic, claiming in one case that 88% of their recommended products outperform their benchmarks over five-year horizons (100% in some asset classes) and in the other that 76% of their recommended products do so. These statistics present several problems. First, as we have seen, comparing product returns against benchmarks when institutional products on average outperform their benchmarks is misleading. Even more problematic, however,

is the survivorship bias which is mechanically introduced by such a comparison. Calculations are necessarily limited to recommended products with available five-year forward return data at each point in time (or at rating inception). Since better performing products are more likely to survive during such a time horizon, this requirement results in better performing products being overrepresented in the sample used to compute this statistic. Other problems identified in previous sections, such as arbitrary sample selection and inclusion of backfilled and simulated returns, are also likely to affect the results.

In Table VIII we show the percentage of recommended products that outperform their benchmarks over one-, three-, and five-year horizons. We perform these calculations using returns gross of asset manager fees and excluding backfilled and simulated returns, but otherwise following the practices of consultants in producing such statistics. Thus the sample is limited to products with available one-, three-, and five-year forward return data at each point in time (and as a result affected by the survivorship bias mentioned above). We also limit our analysis to recommendations issued by the investment consultants in our reduced sample. Following this procedure we find that 74.3% of recommended products outperform benchmarks over a five-year horizon (63.8% over a one-year horizon). This number is lower than that claimed by the consultants in our reduced sample, the difference reflecting arbitrary exclusions of products and the ratings survivorship bias outlined in Section 3.B. However, it is still remarkably high, and misleadingly so, if consultants' clients are unaware that institutional products generally outperform benchmarks or that the requirement for periods of forward return data introduces survivorship bias.

In Table VIII we also report the percentage of consultant recommended products outperforming the median product in their investment category and, separately, the average of all products in that category, over one-, three-, and five-year horizons. These calculations are based on all products in eVestment categories with five or more members (for comparison). These metrics are not distorted by the average outperformance of institutional products. It is true that they are still affected by the survivorship bias explained above, although here this applies to recommended products as well as to the other products used in the comparison. According to this metric, 52.3% of recommended products outperform the median product in the same eVestment category over a five-year horizon (51% over a one-year horizon), and 53.8% outperform the average of all products in that category in that horizon (52.5% over a one year-horizon). In both

cases the fraction of outperforming funds is marginally over half, but falls far short of the 76% to 88% claimed by the consultants.

4. Volatility, tracking error, and the risk profile of recommended and non-recommended products

So far, our comparison of recommended and non-recommended products has been based on two types of risk adjustment: we have computed benchmark-adjusted product returns, and we have benchmarked recommended products against non-recommended products classified in the same (narrow) eVestment category. These procedures mirror the way in which products are classified, evaluated, and compared by plan sponsors, but it is conceivable that benchmarks and investment categories could fail fully to capture differences in riskiness between recommended and non-recommended products.

Table IX reports the average return volatility, tracking error volatility, and benchmark beta of recommended and non-recommended products, and the difference between the two. Return and tracking error volatility are measured as the standard deviation of, respectively, quarterly product returns and excess returns over manager-chosen benchmarks. For exposition, we annualize return and tracking error volatilities by multiplying quarterly standard deviations by the square root of four. Beta is the regression coefficient in a regression of product returns on product benchmark returns. We compute these statistics both for the two-year (or eight-quarter) period preceding the recommendation (the information in the consultants information set), and, more relevantly for investors, the two-year period following the recommendation.

In terms of return volatility there are no consistent differences between recommended and non-recommended products. Whereas return volatilities tend to be larger for recommended products in the full sample (although significantly so only in the pre-recommendation period), they are on average marginally smaller when we use a matched sample of recommended and non-recommended products classified in the same eVestment category. In such a sample, recommended products exhibit a volatility of 12.67% per year and non-recommended products 12.75% per year. Betas with respect to manager-chosen benchmarks for recommended and non-recommended products are equally indistinguishable. In the same matched sample they stand at 0.94 for recommended product versus 0.93 for non-recommended products in the two year period

following the recommendation. These results suggest that there are no significant differences in terms of total and systematic factor risk between the two sets of products, nor any significant differences in leverage between the two. They also indicate that alternative benchmarking approaches which allow for differences in factor exposures to a single category-specific factor are unlikely to affect our performance results.

Where there is a significant difference between recommended and non-recommended products is in tracking error volatility. Recommended products exhibit lower tracking error volatility than non-recommended product both pre- and post-recommendation. Thus, in the post-recommendation matched sample the tracking error volatility of recommended products stands at 3.29% per year, compared to 3.63% per year for non-recommended products. This difference is not large but it is highly statistically significant, suggesting that, on average, recommended managers deviate less aggressively from their benchmarks than non-recommended ones.

Why consultants or plan sponsors may want to avoid tracking error and not total return volatility, and whether this is a good thing, are questions about which there is no consensus in the literature. Lower tracking errors may make performance monitoring easier. As Roll (1992) notes, asset returns are extremely noisy, and a long time may elapse before the fund sponsor is statistically sure that the manager is adding (or subtracting) value. In this context, low tracking error volatility may be desirable because the fund sponsor could more easily ascertain whether the asset manager is adding value over an index fund alternative.

Lower tracking errors may also help with decentralized portfolio management, in which a plan sponsor faces the problem of coordinating multiple managers to achieve optimal diversification (see, for instance, Sharpe (1981), van Binsbergen, Brandt, and Koijen (2008), and Blake, Rossi, Timmermann, Tonks, and Wermers (2013)). In that context, keeping returns close to a known index (low tracking error) might help with the coordination problem by making the attributes of individual portfolios more predictable. As Roll (1992) notes, however, investment products managed or selected based on tracking error volatility are dominated by other feasible portfolios that have both lower volatility and higher average total return. As a result, for the sponsor's overall portfolio to be optimal, these individual manager disadvantages must be offset by substantially lower cross-manager correlations when they adhere to a tracking error volatility strategy – a possibility about which Roll, together with Sharpe (1981) and Jorion (2003), is skeptical. The desirability of staying close to the benchmark is also called into question by the findings of

Cremers and Petajisto (2009), according to which mutual funds that deviate more from their indices tend to outperform those that do not. Similarly Sun, Wang, and Zheng (2011) find that hedge funds that deviate more aggressively from their peers perform better.

Whatever the reason, our results in Table IX suggest that tracking error may matter to consultants. This is consistent with comments from practitioners and the observations of Roll (1992), Jorion (2003), and others, suggesting that institutional investors commonly impose a limit on the volatility of the deviation of the active portfolio from the benchmark. It is also consistent with the reliance of some practitioners on the use of information ratios, defined as the ratio of excess returns over benchmarks and tracking errors, in their analysis.

It is worth noting that, even when information ratios are used, recommended products do not outperform: although tracking error is lower (Table IX), this is offset by their lower excess returns over benchmarks (Table V). In any case, while we found repeated claims about excess returns over benchmarks in the dissemination materials of our reduced sample of consultants, we found none about tracking errors or information ratios, suggesting that these measures are likely to be a secondary consideration.

Consistent with the results of Table IX, an analysis of the distribution of recommended and non-recommended product returns shows that non-recommended products are more likely to gravitate towards the tails (positive and negative) of the distribution than recommended products when products are ranked within each narrow investment category but not when they are ranked within each asset class. These results are presented in Table X. In the first part of this table recommended and non-recommended products are assigned to different deciles depending on their gross return ranking relative to other products in the same asset class. In the second part of the table, products are assigned to different deciles depending on their gross return ranking relative to other products in the same eVestment investment category. In both parts of the table returns are measured over a calendar year and matched to recommendations at the end of the previous year.[22]

Results in the first part of Table X indicate that, when ranking products within asset classes (equity, fixed income, etc.) non-recommended products are overrepresented (relative to recommended products) in the top deciles (seven to ten) of the distribution, whereas recommended products are overrepresented in the middle to bottom deciles (two to six). But when products are

[22] For this table we only consider eVestment categories with at least ten products available. Even with this restriction the number of products in each decile is not necessarily constant across deciles.

ranked within each granular investment category, non-recommended products are overrepresented in the tails of the distribution (deciles one, two, and ten) whereas recommended products gravitate towards the middle deciles (three to nine). This is consistent with the results presented in Table IX. Lower tracking errors with respect to narrowly defined benchmarks (consistent with the granular category the products are assigned to) would suggest that recommended products are closer to the median of the distribution than non-recommended products with higher tracking errors. Non-recommended products, however, do not exhibit higher return volatility than recommended products (while having marginally higher mean returns); this could explain why, when sorting products within broad asset classes, they do not consistently occupy the tails of the distribution.

An analysis of return distributions is potentially important because (some) plan sponsors may not be able to invest in multiple recommended or non-recommended products.[23] If so, they may be concerned not just about the average performance of recommended and non-recommended products, but also about their performance volatility or idiosyncratic risk. However, even if plan sponsors are concerned about idiosyncratic risk in individual products, it is unclear how maintaining low tracking errors, but not low return volatility, is beneficial to them.

5. Conclusions

Investment consultants tend to overstate their ability to select fund managers. A forensic analysis of the claims they make about the performance of the products they recommend reveals a number of practices that lead to this distorted view: comparison of recommended products to manager-chosen benchmarks (which are frequently beaten by all institutional products, not just those recommended) rather than peer products; inclusion of simulated and backfilled returns in the analysis; use of assessment practices leading to survivorship bias; and even seemingly arbitrary inclusions and exclusions of products in their analysis. The distortions generated by these practices – which vary between consultants – means that institutional investors face an almost impossible task in assessing not just the absolute, but also the relative performance of investment consultants.

[23] Blake, Rossi, Timmermann, Tonks, and Wermers (2013) observe that it is indeed common for plan sponsors to employ multiple competing managers (rather than single managers).

An analysis of return volatilities and betas with respect to benchmarks of consultant recommended and non-recommended products suggests that there are no significant differences in riskiness between these two sets. Recommended products, however, tend to perform closer to, and deviate less aggressively from, their benchmarks than non-recommended products, although differences again are not very large.

This study highlights a number of concerns with the performance reporting methodologies employed by consultants. We do not find these methodologies set out clearly in materials prepared for clients, and some practices are problematic even if disclosed, as they tend to result in biases that clients may not be able to understand. Our study suggests that, for performance assessment to be meaningful, it should allow for external scrutiny. It also adds to the suspicion that, for lack of information, plan sponsors are unable effectively to assess and monitor investment consultants. This opacity, together with the upward bias in consultants' claims about their performance, may help to explain the considerable impact that consultants' recommendations have on institutional investment flows, whether between different active products or into active, rather than passive, products in general.

References

Blake, David, Alberto G. Rossi, Allan Timmermann, Ian Tonks, and Russ Wermers, 2013, Decentralized investment management: Evidence from the pension fund industry. *The Journal of Finance* 68, 1133-1178.

van Binsbergen, Jules, H., Michael W. Brandt, and Ralph SJ Koijen, 2008, Optimal decentralized investment management. The Journal of Finance 63,1849-1895.

Busse, Jeffrey A., Amit Goyal, and Sunil Wahal. 2010, Performance and persistence in institutional investment management. *The Journal of Finance* 65, 765-790.

Cohen, Randolph B., Paul A. Gompers, and Tuomo Vuolteenaho, 2002, Who underreacts to cash-flow news? Evidence from trading between individuals and institutions. *Journal of financial Economics* 66, 409-462.

Cremers, KJ Martijn, and Antti Petajisto, 2009, How active is your fund manager? A new measure that predicts performance. The Review of Financial Studies 22, 3329-3365.

Del Guercio, Diane, and Paula A. Tkac, 2002, The determinants of the flow of funds of managed portfolios: Mutual funds vs. pension funds. Journal of Financial and Quantitative Analysis 37, 523-557.

Edelen, Roger M., 1999, Investor flows and the assessed performance of open-end mutual funds. *Journal of Financial Economics* 53, 439-466.

Evans, Richard B., and Rüdiger Fahlenbrach, 2012, Institutional investors and mutual fund governance: Evidence from retail–institutional fund twins. *The Review of Financial Studies* 25, 3530-3571.

Fama, Eugene F., and Kenneth R. French, 2010, Luck versus skill in the cross-section of mutual fund returns. *The journal of finance* 65, 1915-1947.

FCA, 2017, *Asset Management Market Study Final Report: Annex 5 – Assessment of third party datasets*, https://www.fca.org.uk/publication/market-studies/ms15-2-3-annex-5.pdf

Gerakos, Joseph, Juhani T. Linnainmaa, and Adair Morse, 2016, Asset Managers: Institutional Performance and Smart Betas. No. w22982. National Bureau of Economic Research.

Goyal, Amit, and Sunil Wahal, 2008, The selection and termination of investment management firms by plan sponsors. *The Journal of Finance* 63, 1805-1847.

Hunter, David, Eugene Kandel, Shmuel Kandel, and Russ Wermers, 2014, Mutual fund performance evaluation with active peer benchmarks. Journal of Financial economics 112, 1-29.

James, Christopher, and Jason Karceski, 2006, Investor monitoring and differences in mutual fund performance. *Journal of Banking & Finance* 30, 2787-2808.

Jenkinson, Tim, Howard Jones, and Jose Vicente Martinez, 2016, Picking winners? Investment consultants' recommendations of fund managers. *The Journal of Finance* 71, 2333-2370.

Jones, Howard, and Jose Vicente Martinez, 2017, Institutional investor expectations, manager performance, and fund flows. *Journal of Financial and Quantitative Analysis* 52, 2755-2777.

Jorion, Philippe, 2003, Portfolio optimization with tracking-error constraints. Financial Analysts Journal 59, 70-82.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1992, The Structure and Performance of the Money Management Industry. Brookings Papers on Economic Activity, 339–391.

Ljungqvist, Alexander, Christopher Malloy, and Felicia Marston, 2009, Rewriting history. *The Journal of Finance* 64, 1935-1960.

Malkiel, Burton G., and Atanu Saha, 2005, Hedge funds: Risk and return. *Financial Analysts Journal* 61, 80-88.

Newey, Whitney K. and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

Patton, Andrew J., Tarun Ramadorai, and Michael Streatfield, 2015, Change you can believe in? Hedge fund data revisions. *The Journal of Finance* 70, 963-999.

Roll, Richard, 1992, A mean/variance analysis of tracking error. Journal of Portfolio Management, 18, 13-22.

Sensoy, Berk A., 2009, Performance evaluation and self-designated benchmark indexes in the mutual fund industry. *Journal of Financial Economics* 92, 25-39.

Sharpe, William F., 1981, Decentralized investment management. The Journal of Finance 36, 217-234.

Sun, Zheng, Ashley Wang, and Lu Zheng, 2011, The road less traveled: Strategy distinctiveness and hedge fund performance. The Review of Financial Studies 25, 96-143.

Table I

Asset Managers, Products, and Recommendations - Descriptive Statistics by Year

This table presents descriptive statistics on the sample of asset managers, institutional products, and investment consultant recommendations used in our study. It shows the number of institutional investment products available each year, the number of asset managers managing those products, the average number of products per asset manager, the number of recommendations of institutional products issued by the investment consultants in our sample, the number of investment consultants in our sample, and the average number of recommendations per investment consultant. The number of recommendations, the number of investment consultants and the average number of recommendations per investment consultant are shown first for the full sample of investment consultants, and then for a reduced sample of three consultants for which we have detailed information of the way they represent their performance before their (prospective) clients.

				Full Sample			Reduced Sample		
	Number of Products	Number of Asset Managers	Average Number of Products per Asset Manager	Number of Recommendations	Number of Investment Consultants	Average Number of Recommendations per Investment Consultant	Number of Recommendations	Number of Investment Consultants	Average Number of Recommendations per Investment Consultant
2006	11,413	1,685	6.8	931	6	155.2	853	3	284.3
2007	12,060	1,703	7.1	1,223	6	203.8	1,122	3	374.0
2008	12,546	1,720	7.3	1,500	7	214.3	1,336	3	445.3
2009	12,853	1,736	7.4	1,662	7	237.4	1,477	3	492.3
2010	13,151	1,745	7.5	1,823	7	260.4	1,611	3	537.0
2011	13,300	1,706	7.8	1,890	6	315.0	1,771	3	590.3
2012	13,308	1,679	7.9	2,036	6	339.3	1,895	3	631.7
2013	13,197	1,656	8.0	2,077	6	346.2	1,938	3	646.0
2014	12,869	1,620	7.9	2,186	6	364.3	2,025	3	675.0
2015	12,359	1,570	7.9	2,302	6	383.7	2,116	3	705.3
Mean	12,706	1,682	7.6	1,763	6	282.0	1,614	3	538.1

Table II

Products and Recommendations by Asset Class

This table presents descriptive statistics on the sample of institutional investment products and investment consultant recommendations used in our study by asset class category. It shows the number of eVestment subcategories, the number and percentage of investment products, and the number of asset managers with products in each asset class. It also shows the number and percentage of recommendations in each asset class and the percentage of observations in the sample for which we have asset manager fee data available. The first part of the table present these results for the full sample of investment consultants, whereas the second part of the table shows the results for a reduced sample of three consultants for which we have detailed information of the way they represent their performance before their (prospective) clients.

	Number of eVestment subcategories covered	Number of Products	Percentage of Products	Number of Asset Managers	Number of Recommendations	Percentage of Recommendations	Percentage of Observations with AM Fee Data Available
				Full Sample			
Equity	163	10,045	60.3	1,642	2,095	58.6	74.2
Fixed Income	65	4,796	28.8	666	1,238	34.6	74.5
Balanced/Multi-asset	27	1,117	6.7	370	132	3.7	54.7
Alternatives/Hedge Funds	28	459	2.8	228	68	1.9	29.5
Real Estate	6	232	1.4	124	45	1.3	56.8
				Reduced Sample			
Equity	163	10,045	60.3	1,642	1,888	59.4	74.2
Fixed Income	65	4,796	28.8	666	1,100	34.6	74.5
Balanced/Multi-asset	27	1,117	6.7	370	93	2.9	54.7
Alternatives/Hedge Funds	28	459	2.8	228	56	1.8	29.5
Real Estate	6	232	1.4	124	39	1.2	56.8

Table III

Actual and Claimed Performance of Consultants' Recommendations

This table shows the average performance claimed by three leading investment consultants for their recommendations of investment products. It also shows the actual performance of these consultant's recommended and not recommended products, and the difference between the two, according to our calculations. Performance is measured using returns and returns in excess of asset manager chosen benchmarks. The first two column shows results for the full sample of recommended and not recommended products and assuming products are held only for the period they are recommended (or not) by consultants. The third and fourth columns shows results for a matched sample of recommended and not recommended products classified in the same eVestment category, thus effectively comparing recommended and not recommended products category by category. Returns and excess returns are expressed in % per year. Statistics are computed on quarterly returns and annualized by multiplying quarterly returns by four. t-statistics based on standard errors, robust to conditional heteroscedasticity and serial correlation as in Newey and West (1987), are reported in parentheses. Claimed minus actual figures are computed as the difference between the claimed average recommended product excess returns over benchmarks across consultants and each of our four measures of recommendation performance. P-values for the differences are reported in square brackets and are based on Newey-West standard errors. The middle part of the table shows the results obtained using investment products' gross returns whereas the last part of the table shows results obtained using returns net of asset management fees assuming a notional investment of USD 50 million (but not making any adjustment for investment consultant fees). ***, **, * denote statistical significance at 1%, 5%, and 10% levels, respectively.

Recommended Products' Average Excess Return over Benchmark	Claimed Performance Consultants' Recommendations Weighted Avg. across Consultants 1.73%			
	Actual Performance of Consultants' Recommendations - Gross Returns Analysis			
	Avg. Returns	Avg. Excess Ret. over Benchmark	Avg. Returns - Matched Sample	Avg. Excess Ret. over Benchmark - Matched Sample
Recommended Products	5.40% (1.04)	0.45% (1.38)	5.47% (1.05)	0.43% (1.28)
Non-Recommended Products	5.71% (1.17)	0.33% (1.05)	5.68% (1.11)	0.63% (2.58)**
Recommended minus Non-Recommended Products	-0.30% (-0.45)	0.12% (0.58)	-0.21% (-0.96)	-0.21% (-0.98)
Actual minus Claimed	-2.03% [<0.00]***	-1.61% [<0.00]***	-1.94% [<0.00]***	-1.93% [<0.00]***
	Actual Performance of Consultants' Recommendations - Net Returns (Net of Asset Managers Fees) Analysis			
	Avg. Returns	Avg. Excess Ret. over Benchmark	Avg. Returns - Matched Sample	Avg. Excess Ret. over Benchmark - Matched Sample
Recommended Products	5.11% (0.95)	-0.32% (-0.85)	5.18% (0.98)	-0.39% (-0.91)
Non-Recommended Products	5.34% (1.12)	-0.32% (-1.04)	5.41% (1.06)	-0.13% (-0.53)
Recommended minus Non-Recommended Products	-0.24% (-0.30)	0.00% (-0.00)	-0.23% (-0.91)	-0.26% (-0.79)
Actual minus Claimed	-1.96% [0.02]**	-1.73% [<0.00]***	-1.95% [<0.00]***	-1.99% [<0.00]***

Table IV

Actual and Claimed Performance of Consultant's Recommendations by Consultant and Asset Class

The first part of this table shows the performance claimed by three leading investment consultants for their recommendations of investment products and how it compares to their actual performance according to our calculations. We measure performance as the difference between recommended and not recommended products excess returns over their asset manager chosen benchmarks and as the difference in returns between recommended products and a matched sample of not recommended products classified in the same eVestment category. Returns and excess returns are expressed in % per year. Statistics are computed on quarterly returns and annualized by multiplying quarterly returns by four. t-statistics based on standard errors, robust to conditional heteroscedasticity and serial correlation as in Newey and West (1987), are reported in parentheses. Claimed minus actual figures are computed as the difference between the claimed average recommended product excess returns over benchmarks across consultants and each of our two measures of recommendation performance. P-values for the differences are reported in square brackets and are based on Newey-West standard errors. The second part of the table show results disaggregated by asset class, rather than investment consultant. All results are obtained using investment products' gross returns. ***, **, * denote statistical significance at 1%, 5%, and 10% levels, respectively.

	By Consultant		
	Consultant 1	Consultant 2	Consultant 3
Claimed	1.64%	2.51%	1.65%
Actual: Recommended minus not Recommended	0.30%	0.27%	0.08%
Average Excess Ret. Over Benchmark	(0.55)	(0.59)	(0.48)
Claimed minus Actual	1.34%	2.24%	1.57%
	[0.02]**	[<0.00]***	[<0.00]***
Actual: Recommended minus not Recommended	0.15%	-0.95%	-0.34%
Average Ret. Difference in a Matched Sample	(0.27)	(-1.30)	(-1.43)
Claimed minus Actual	1.49%	3.46%	2.00%
	[<0.00]***	[<0.00]***	[<0.00]***

	By Asset Class		
	Equity	Fixed Income	Balanced / Multi Asset, Alternatives / Hedge funds, and Real Estate
Claimed	2.21%	0.93%	1.12%
Actual: Recommended minus not Recommended	0.04%	0.38%	0.34%
Average Excess Ret. Over Benchmark	(0.15)	(1.28)	(0.35)
Claimed minus Actual	2.18%	0.55%	0.78%
	[<0.00]***	[0.07]*	[0.42]
Actual: Recommended minus not Recommended	-0.21%	-0.14%	-0.72%
Average Ret. Difference in a Matched Sample	(-0.70)	(0.46)	(-1.05)
Claimed minus Actual	2.42%	1.07%	1.84%
	[<0.00]***	[<0.00]***	[0.01]**

Table V

What is Behind the Differences between Actual and Claimed Consultants' Recommendation Performance

The first part of this table lists four different performance measurement practices employed by the consultants in our sample that help explain the differences between consultants' claimed performance and our results. The second part of the table shows the impact of these practices on results. For each consultant in the reduced sample we start with the difference in returns between recommended products and a matched sample of non-recommended products (our measure of performance) and sequentially analyse the additive impact of each of the practices we identify in our sample of consultants on estimated performance. The residual is the difference between consultants' claimed performance (reported in the penultimate row) and our estimation of consultants' recommending performance using consultants' methodology and our data. All results are obtained using investment products' gross returns. Returns and excess returns are expressed in % per year. Statistics are computed on quarterly returns and annualized by multiplying quarterly returns by four. P-values for the differences between actual and alternative performance measures based on consultants' reporting practices, and for the residual, are reported in square brackets and are based on Newey-West standard errors. ***, **, * denote statistical significance at 1%, 5%, and 10% levels, respectively.

	Performance Measurement Practices Employed by Consultants		
	Consultant 1	Consultant 2	Consultant 3
Comparison to Benchmarks rather than other Products in the Same Category	✓	✓	✓
Inclusion of Simulated and Backfilled Returns	✓	✓	✓
Rating Survivorship Conditions	✓		
Arbitrary Inclusion/Exclusions of Products and Investment Categories	✓	✓	

	Cummulative Impact on Measured Performance		
	Consultant 1	Consultant 2	Consultant 3
Actual Performance	0.15%	-0.95%	-0.34%
Comparison to Benchmarks rather than other Products in the Same Category	0.71% [0.04]**	0.63% [<0.00]***	0.38% [0.01]**
+ Inclusion of Simulated and Backfilled Returns	0.89% [0.15]	0.94% [0.02]**	0.77% [<0.00]***
+ Rating Survivorship Conditions	1.16% [0.05]*	-	-
+ Arbitrary Inclusion/Exclusions of Products and Investment Categories	1.19% [0.05]**	1.35% [<0.00]***	-
Claimed Performance	1.64%	2.51%	1.65%
Residual	0.45% [0.17]	1.15% [<0.00]***	0.89% [<0.00]***

Table VI

Performance of Recommended and Not Recommended Products - All Consultants

This table shows the performance of consultant recommended and not recommended investment products, and the difference between the two. Performance is measured using returns and returns in excess of asset manager chosen benchmarks. The first two column shows results for the full sample of recommended and not recommended products and assuming products are held only for the period they are recommended (or not) by consultants. The third and fourth columns shows results for a matched sample of recommended and not recommended products classified in the same eVestment category, thus effectively comparing recommended and not recommended products category by category, and under the same holding period assumptions. The last four columns shows the same statistics for portfolios of products built assuming investors hold them for five years starting each listing quarter (or until the end of our sample period). The first part of the table shows the results obtained using investment products gross returns whereas the second part of the table shows results obtained using returns net of asset management fees assuming a notional investment of USD 50 million (but not making any adjustment for investment consultant fees). Returns and excess returns are expressed in % per year. Statistics are computed on quarterly returns and annualized by multiplying quarterly returns by four. The sample period is January 2006 to December 2015. t-statistics based on standard errors, robust to conditional heteroscedasticity and serial correlation as in Newey and West (1987), are reported in parentheses. ***, **, * denote statistical significance at 1%, 5%, and 10% levels, respectively.

	Holding Period for as long as Listed				Five-year Holding Period			
	Avg. Returns	Avg. Excess Ret. over Benchmark	Avg. Returns - Matched Sample	Avg. Excess Ret. over Benchmark - Matched Sample	Avg. Returns	Avg. Excess Ret. over Benchmark	Avg. Returns - Matched Sample	Avg. Excess Ret. over Benchmark - Matched Sample
				Gross Returns				
Recommended Products	5.48% (1.06)	0.43% (1.36)	5.54% (1.07)	0.40% (1.25)	5.64% (1.11)	0.26% (0.68)	5.66% (1.12)	0.25% (0.67)
Non-Recommended Products	5.71% (1.17)	0.33% (1.05)	5.83% (1.14)	0.71% (2.82)***	5.79% (1.19)	0.19% (0.55)	6.01% (1.24)	0.35% -1.14
Recommended minus Non-Recommended Products	-0.23% (-0.35)	0.10% (0.44)	-0.29% (-1.38)	-0.31% (-1.44)	-0.15% (-0.25)	0.07% (0.33)	-0.36% (-1.03)	-0.10% (-0.46)
				Net Returns (Net of Asset Managers Fees)				
Recommended Products	5.23% (0.98)	-0.35% (-1.01)	5.32% (1.02)	-0.38% (-0.99)	5.10% (0.98)	-0.54% (-1.31)	5.10% (0.98)	-0.53% (-1.31)
Non-Recommended Products	5.34% (1.12)	-0.32% (-1.03)	5.61% (1.09)	-0.09% (-0.35)	5.30% (1.10)	-0.48% (-1.46)	5.43% (1.13)	-0.30% (-0.89)
Recommended minus Non-Recommended Products	-0.11% (-0.14)	-0.04% (-0.04)	-0.29% (-1.29)	-0.29% (-1.24)	-0.20% (-0.32)	-0.07% (-0.31)	-0.33% (-0.64)	-0.24% (-1.01)

Table VII

Performance of Negatively Recommended Products

This table shows the average performance of consultants' negatively rated products and the difference between these and all other products in our sample. Performance is measured using returns and returns in excess of asset manager chosen benchmarks. The first two column shows results for the full sample of negatively rated and non-negatively rated products and assuming products are held only for the period they are rated as such by consultants. The third and fourth columns shows results for a matched sample of negatively and non-negatively rated products classified in the same eVestment category, thus effectively comparing negatively recommended and all other products category by category. Returns and excess returns are expressed in % per year. Statistics are computed on quarterly returns and annualized by multiplying quarterly returns by four. t-statistics based on standard errors, robust to conditional heteroscedasticity and serial correlation as in Newey and West (1987), are reported in parentheses. The first part of the table shows the results obtained using investment products' gross returns whereas the second part of the table shows results obtained using returns net of asset management fees assuming a notional investment of USD 50 million. ***, **, * denote statistical significance at 1%, 5%, and 10% levels, respectively.

	Actual Performance of Consultants' Recommendations - Gross Returns Analysis			
	Avg. Returns	Avg. Excess Ret. over Benchmark	Avg. Returns - Matched Sample	Avg. Excess Ret. over Benchmark - Matched Sample
Negatively Recommended Products	6.30% (0.97)	0.50% (1.25)	6.33% (0.99)	0.49% (1.18)
Non-Neg. Recommended Products	5.68% (1.17)	0.37% (1.17)	6.56% (1.05)	0.79% (1.96)*
Negatively Recommended minus Non-Neg. Recommended Products	0.62% (0.34)	0.13% (0.50)	-0.24% (-0.87)	-0.30% (-1.10)
	Actual Performance of Consultants' Recommendations - Net Returns (Net of Asset Managers Fees) Analysis			
	Avg. Returns	Avg. Excess Ret. over Benchmark	Avg. Returns - Matched Sample	Avg. Excess Ret. over Benchmark - Matched Sample
Negatively Recommended Products	6.00% (0.91)	-0.32% (-0.84)	6.08% (0.93)	-0.32% (-0.81)
Non-Neg. Recommended Products	5.30% (1.11)	-0.29% (-0.94)	6.28% (0.98)	-0.08% (-0.19)
Negatively Recommended minus Non-Neg. Recommended Products	0.70% (0.35)	-0.03% (-0.13)	-0.21% (-0.71)	-0.24% (-0.86)

Table VIII

Percentage of Recommended Products Outperforming Benchmarks and other Products in the Same Investment Category

This table shows the percentage of consultant recommended products outperforming their benchmarks, the median product in their investment category, or the average of all products in that category, over one-, three-, and five-year horizons. Performance is measured using returns gross of asset manager fees. The sample is limited to recommendations issued by investment consultants in our reduced sample and products with available one-, three-, and five-year ahead return data at each point in time.

	One-year Holding Period	Three-year Holding Period	Five-year Holding Period
% of Recommended Products Outperforming Benchmarks	63.8%	71.3%	74.3%
% of Recommended Products Out performing the Median Product in their Category	51.0%	52.7%	52.3%
% of Recommended Products Outperforming the Average of all Product in their Category	52.5%	54.3%	53.8%

Table IX

Return Volatility, Tracking Error Volatility, and Beta of Recommended and Non-Recommended Products

This table shows the average return volatility, tracking error volatility, and beta of recommended and not recommended products, and the difference between the two. Return and tracking error volatility are measured as the standard deviation of quarterly product returns and excess returns over manager chosen benchmarks respectively. Beta is the regression coefficient in a regression of product returns on product benchmark returns. In the first part of the table these statistics are computed using product and benchmark returns for the two-year period preceding the recommendation whereas the second part of the table shows results obtained using product and benchmark returns for the two-year period following the recommendation. The first two column shows results for the full sample of recommended and not recommended products. The third and fourth columns shows results for a matched sample of recommended and not recommended products classified in the same eVestment category. Volatilities are expressed in % per year (they are annualized by multiplying quarterly standard deviations by the square root of four). P-values for the differences are reported in square brackets and are based on standard errors clustered at the product level. ***, **, * denote statistical significance at 1%, 5%, and 10% levels, respectively.

	Two-year Period Preceding the Recommendation					
	Return Volatility	Beta	Tracking Error Volatility	Return Volatility - Matched Sample	Beta - Matched Sample	Tracking Error Volatility - Matched Sample
Recommended Products	13.12%	0.89	3.41%	13.06%	0.89	3.39%
Non-Recommended Products	12.48%	0.96	3.90%	13.16%	0.85	3.67%
Recommended minus Non-Recommended Products	0.64% [<0.00]***	-0.08 [0.66]	-0.48% [<0.00]***	-0.09% [0.07]*	0.03 [0.84]	-0.28% [<0.00]***
	Two-year Period Following the Recommendation					
	Return Volatility	Beta	Tracking Error Volatility	Return Volatility - Matched Sample	Beta - Matched Sample	Tracking Error Volatility - Matched Sample
Recommended Products	12.72%	0.94	3.30%	12.67%	0.94	3.29%
Non-Recommended Products	12.70%	0.93	3.91%	12.75%	0.93	3.63%
Recommended minus Non-Recommended Products	0.02% [0.90]	0.01 [0.94]	-0.61% [<0.00]***	-0.08% [0.08]*	0.02 [0.92]	-0.34% [<0.00]***

Table X

Return Distribution of Recommended and Non-Recommended Products

This table shows the distribution of recommended and non-recommended product returns within each asset class or investment category. The table shows the percentages of recommended and non-recommended products in each return decile. In the first part of the table products are assigned to different deciles depending on their return ranking relative to other products in the same asset class. In the second part of the table products are assigned to different deciles depending on their return ranking relative to other products in the same eVestment investment category. The p-value of a t-test of the difference between the percentage of recommended and non-recommended products in each return decile is shown in square brackets. Results are obtained using investment products' gross returns. We only consider eVestment categories with at least ten products available. Returns are measured over a calendar year and matched to recommendations at the end of the previous year. ***, **, * denote statistical significance at 1%, 5%, and 10% levels, respectively.

					By Asset Class					
Decile	1 (Bottom)	2	3	4	5	6	7	8	9	10 (Top)
Recommended Products	9.88%	10.83%	10.56%	10.93%	10.44%	10.42%	9.32%	9.15%	9.68%	8.77%
Non-Recommended Products	9.98%	9.87%	9.92%	9.85%	9.93%	9.96%	10.12%	10.14%	10.07%	10.16%
Difference	-0.09% [0.80]	0.96% [<0.00]***	0.65% [0.08]*	1.08% [<0.00]***	0.51% [0.16]	0.46% [0.21]	-0.80% [0.03]**	-0.98% [<0.00]***	-0.39% [0.28]	-1.39% [<0.00]***
					By Investment Category					
Decile	1 (Bottom)	2	3	4	5	6	7	8	9	10 (Top)
Recommended Products	6.85%	9.01%	10.55%	10.83%	10.75%	11.66%	11.16%	10.45%	10.52%	8.21%
Non-Recommended Products	9.54%	10.26%	10.26%	9.99%	9.72%	10.30%	10.15%	10.06%	10.19%	9.52%
Difference	-2.69% [<0.00]***	-1.25% [<0.00]***	0.29% [0.47]	0.85% [0.03]**	1.03% [<0.00]***	1.36% [<0.00]***	1.01% [0.01]**	0.39% [0.32]	0.33% [0.40]	-1.31% [<0.00]***