

Virtual Reality? Investment Consultants' Claims About Their Own Performance

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

ABSTRACT

Investment consultants market their services by claiming their fund manager recommendations add significant value. Using non-public data sourced from investment consultants and the UK regulator, we find no such evidence, but identify several practices that explain their exaggerated claims: comparisons to benchmarks instead of peers, inclusion of simulated and backfilled returns, use of investment horizons that allow losers to be forgotten, and unexplained exclusions of products from the analysis. Consultants do not fully disclose their methodology to investors, who therefore cannot verify or reliably compare their performance.

July 2021

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

JEL classification: G20, G23, G24, L84

* Gordon Cookson (gpcookson@gmail.com) is with Ofcom; 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; Jose Vicente Martinez (jose.v.martinez@uconn.edu) is with the University of Connecticut, School of Business. We thank Greg Jackson and Anish Thakrar for their assistance over the course of the research project, and we thank Ryan Pratt (discussant), Steven Utke, and seminar participants at Cass Business School, NHH Bergen, University of Bath School of Management, Analytics, Research Affiliates Investment Retreat, and the Financial Intermediation Research Society (Savannah) conference for helpful comments and discussions.

I. INTRODUCTION

Investment consultants are the kingmakers of institutional asset management. By assigning recommendations to fund managers, or withdrawing those recommendations, they have the power to influence which managers are hired and fired by plan sponsors such as pension funds, insurance companies, and endowments.¹ Their influence, however, seems at odds with the lack of evidence that their recommendations add value (Jenkinson, Jones, and Martinez, 2016). Some have pointed to an agency problem, with plan sponsors using consultants to shield themselves from blame in case their chosen managers perform badly (see Lakonishok, Shleifer, and Vishny, 1992, and Jones and Martinez, 2017). We approach the question from another direction, by analysing how investment consultants themselves present the performance of their recommendations. Using non-public data for the UK, we find that consultants exaggerate their own performance in largely non-transparent ways, and that each consultant does so differently, making it very challenging for plan sponsors to choose between them on this measure.

While disclosures by mutual funds, institutional products, and stock analysts are standardized (by regulation or industry practice), performance reporting by investment consultants is not.² Consultants' performance disclosures are also voluntary and essentially unverifiable. In

¹ The investment consulting sector has estimated assets 'under advisement' of some \$41 trillion (Pensions and Investments, 2019).

² Mutual funds offered in the U.S. are required to report performance according to rules set by the Securities and Exchange Commission (SEC); institutional funds in the U.S. and other major capital markets generally abide by the Global Investment Performance Standards (GIPS); and stock analysts in the U.S. are required to report performance according to the rules of the U.S. Financial Industry Regulatory Authority (FINRA), which are also in practice followed in other major capital markets in which U.S. brokers operate. In the U.S., investment consultants are regulated as 'investment advisers' under the U.S. Investment Advisers Act of 1940, and owe a fiduciary duty to their clients (SEC 2013), but there are no rules on performance disclosure. In the U.K. there is no standard form in which consultants are required to disclose information on the performance of past recommendations.

making claims about the performance of their past recommendations, investment consultants typically reveal only the output of their own analysis, but little about their methodology, and nothing of the raw recommendations themselves.³ Nonetheless, investment consultants go to some lengths to highlight the performance of their past manager recommendations. The leading investment consulting firms in the UK share such information with actual and prospective clients, and most of the largest consultants make it publicly available on their websites.⁴

Since investment consultants' recommendations of asset managers are not publicly available, independent analysis of these recommendations and of the claims consultants make about them is difficult. Our research 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 three leading investment consultants, of the claims made by them about these recommendations, and of their own analysis underlying these claims. These consultants have a combined market share of around 51% worldwide (Pensions and Investments, 2019) and of some 60% in the U.K. (FCA, 2016).

All three consultants in our sample claim that the products they recommend produce significant excess returns, ranging between 1.64% and 2.51% per year, but each uses a different methodology and does not make the underlying data public, which means that the basis and

³ By withholding these, consultants avoid disclosing proprietary data which may be valuable to competitors (see Cassar, Gerakos, Green, and Hand (2018) and Guo, Lev, and Zhou (2004), on the competitive pressures felt by hedge funds and biotech firms, respectively, to limit disclosures), but they also make it impossible for their claims to be independently verified.

⁴ The majority of the investment consulting firms included in the Financial Conduct Authority's (FCA) Asset Management Market Study (2017) and the Competition and Markets Authority's (CMA) Investment Consultants Market Investigation (2018) share this information in tenders and/or marketing materials and the largest also disclose it on their websites (see Internet Appendix).

comparability of these claims are opaque to investors. In our analysis we find large differences between consultants' claims on the one hand and our independent analysis of their performance on the other. We find that, on average, consultant-recommended investment products perform no better than other products available to institutional investors: over our 10-year sample period, the portfolio of all products recommended by investment consultants delivered average returns gross of management fees 0.20% per annum lower than the average returns from a matched sample of available products classified in the same investment category (e.g. "U.S. Large Cap Value Equity"). Thus, the weighted average of these claimed excess returns of 1.73% per year exceeds our estimation of their performance by 1.93%. A similar discrepancy, of 1.95% per year, is found after allowing for management fees. For equities, the discrepancy between consultant-reported figures, at 2.21%, and our estimates of their performance of -0.22%, is even higher.

To explain the remarkable gap between consultants' claims about their manager selection skills and our own estimates, we undertake a forensic analysis of the performance disclosures made by each of the consultants. This reveals a number of practices that partially explain the discrepancy. First, consultants compare the returns of recommended products to passive benchmark returns used by the asset management industry rather than to the returns of other institutional products available in the same investment category. This is not an innocent choice, because by doing this, consultants are implicitly taking credit for the fact that the institutional asset management industry as a whole outperforms benchmarks on a gross basis.⁵ Second, some consultants retain simulated

⁵ See Gerakos, Linnainmaa, and Morse (2019). 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 the returns of all available products, the metric we use in this paper).

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 passive benchmarks rather than other available products, are transparent (even if their relevance may 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 flatter 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. Like the claims about performance discussed above, this claim does not allow for the fact that institutional products as a whole outperform benchmarks. Moreover, 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 not only that consultants' claimed performance is wide of the mark, but also that their disclosures are difficult to compare, making them poor guides even to their relative performance. A comparison between the performance of consultants' recommendations based on their own claims shows that the relationship between claimed and actual performance is not monotonic. Therefore, investors cannot just discount the claims of consultants as being generally inflated but ordinally informative when making comparisons of their performance.

In our analysis we rely on two kinds of risk adjustment to compare recommended products against all products available to institutional investors: first we compare product returns against manager-chosen benchmarks, and, secondly, we benchmark recommended products using available products classified in the same narrow eVestment category. 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 available products. For instance, plan sponsors 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. However, there is a small, but significant, 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. However, even when using information ratios (a measure employed by practitioners and defined as the ratio of excess returns over benchmarks to tracking error volatility) we find that recommended products do not outperform other available products (and certainly not by as much as their disclosures seem to suggest): their tracking error is lower but so are their excess returns over benchmarks in a matched sample, with the two measures offsetting each other.

An analysis of consultants' negative manager recommendations, that is, recommendations to withdraw funds from a fund manager, shows that, on average, these products also outperform benchmarks and that they do it by as much as their (positively) recommended products do. These results underline the folly of comparing recommended products to benchmarks (when the average product outperforms). They also suggest that, although following investment consultants' recommendations leads to lower tracking error, it does not specifically limit downside risk.

Research on investment consultants has been sparse, despite their very significant influence on asset flows. Goyal and Wahal (2008) study the hiring and firing decisions of plan sponsors partly as a function of whether they employ consultants or not; Jenkinson, Jones, and Martinez (2016) analyze consultants' recommendations, their drivers and performance; while Chava, Kim, and Weagley (2018) investigate the screening activity of investment consultants. The present study complements these papers by offering a more thoroughgoing analysis of consultants' recommendations (positive and negative) and by comparing the performance of these recommendations with the claims consultants make about them.

Our analysis of how consultants represent their past performance is related to research elsewhere on the self-reporting of performance. Patton, Ramadorai, and Streatfield (2015) analyze performance reporting by hedge funds to financial databases along with subsequent revisions, while Ljungqvist, Malloy, and Marston (2009) study financial analysts' reporting of their stock recommendations and later revisions to their recommendation records in a leading database. Both these studies and ours investigate how financial intermediaries exaggerate and misreport their past performance. More generally our paper adds to the literature on the reliability of non-verifiable voluntary disclosures in other settings (e.g. Price, 1999, Lang and Lundholm, 2000, Leone, Rock, and Willenborg, 2007, Michels, 2012). We show that agents can make misleadingly favorable disclosures even without outright falsification of evidence, so long as a lack of standardization affords them license to present their performance as they please.

II. DATA

Our main data source is a record of the institutional asset managers recommended by each of three investment consultants between 2006 and 2015. The data was gathered by the FCA as part of their Asset Management Market Study which was conducted between November 2015 and June 2017. This was a one-off exercise, as investment consultants are not required to make any filings or disclosures of their recommendations. The Asset Management Market Study investigated the entire asset management industry in the U.K., and investment consultants were included, among other reasons, because their recommendations are considered influential in the appointment of asset managers. The identity of the consultants in the sample cannot be revealed but each of them is in the top six worldwide, with a combined share of 'assets under advisement' of 45% worldwide (Pensions and Investments, 2016) and some 60% in the U.K. (FCA, 2016).

The recommendations issued by these consultants relate to investment products, that is, to the investment styles/categories in which an asset manager offers its services.⁶ ‘Recommendation’ and ‘recommended’ are shorthand for ‘positive recommendation’ and ‘positively recommended’ unless otherwise stated. ‘Non-recommended’ products include products rated neutral or negative as well as products not rated at all by the consultant. Investment consultants’ recommendations are widely used by plan sponsors as the basis for hiring, firing, and monitoring fund managers. These recommendations are highly valued by plan sponsors and they drive very significant asset flows (see Jenkinson, Jones, and Martinez, 2016). In the U.K., regulation requires trustees of pension funds (the largest category of institutional asset owners) to take the advice of an expert, which in most cases means hiring a firm of investment consultants,⁷ but even pension funds with this expertise in-house tend to use consultants given consultants’ broad coverage of investment products (CMA, 2018).⁸ Our data set 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.

For each of the consultants we also have access to the claims they make about the performance of the products they recommend as well as the detailed analysis underlying these claims. This self-analysis, carried out in 2016 but covering several years, forms the basis of

⁶ A product may be offered to investors in different vehicles, such as a pooled fund or a segregated account.

⁷ The Occupational Pension Schemes (Investment) Regulations 2005, regulation 2(2a).

⁸ The consultant services used by pension funds vary depending on the type of the pension scheme. However, while defined benefit (DB) schemes tend to demand a wider range of services (including, for example, liability-driven investment and fiduciary management) than defined-contribution (DC) schemes, both types of scheme use consultants’ manager search and selection services.

consultants' marketing to asset owners.⁹ 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.

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. 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).¹⁰ 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

⁹ The material made available to the FCA comprised comprehensive marketing packs, from which consultants would select parts according to the needs of the particular asset owner. For example, when presenting to the trustees of a pension fund with a fully retired set of members, consultants would be likely to focus on fixed income, but if the scheme had a young set of members still largely on the payroll, the consultant would be more likely to focus on asset allocation and equities.

¹⁰ 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.

observations, are based on an investment of \$50 million.¹¹ 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 added to the eVestment database; in such cases the returns in the intervening period are flagged. The ‘simulated’ flag indicates situations where a new product launches and the manager generates a past performance history by simulating (hypothetical) returns for that 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. We include in the sample only products which report gross returns to eVestment and which are classed by eVestment as institutional or alternatives.

Tables 1 and 2 contain descriptive statistics on our sample. As Table 1 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, the consultants in our sample issued an average of 1,614 recommendations per year; this rose steadily during the period, from 853 in 2006 to 2,116 in 2015.

¹¹ 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.

The increase in the number of recommendations per consultant was the result of organic growth as well as growth from acquisitions.¹² Divided by the number of consultants in our full sample, this corresponds to an average of 538 recommendations per consultant per year. Managers sometime receive recommendations from multiple consultants and are sometimes recommended across multiple products. These recommendations typically last for several quarters.

Table 2 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 (59.4% of all recommendations). The next largest class is Fixed Income, and together Equities and Fixed Income account for 89.1% of all products and 94% of all recommendations. Table 2 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.

In contrast to retail mutual funds, institutional 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 of error, each consultant was

¹² The investment consultancy market grew significantly during the period we analyze, as did the fiduciary management market where big consultants are prominent (see CMA, 2018). The growth in fiduciary management has driven additional manager recommendation research, as fiduciary management selections/products typically rely on highly-rated managers.

requested to identify the product in the eVestment database to which each recommended product corresponded.

III. RECOMMENDATION PERFORMANCE AND CONSULTANT DISCLOSURES

3.1 The Claimed and Actual Performance of Consultant Recommendations

All three consultants in our sample calculate and disclose the performance of their manager recommendations. They do so by computing simple average excess returns over benchmarks for all (positively) recommended products in a number of investment categories defined by themselves, and they present value added results for each of those investment categories in their marketing materials. For instance, one of the consultants in our sample reports excess returns over benchmarks for their recommendations of U.K. equity products (2.5% per year over the last ten years), U.K. small cap equity (4.9% per year), Emerging Market Debt (0.2% per year), and 29 other investment categories. 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.¹³

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 excess returns for these broad asset classes, we compute them ourselves using the simple average of the

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.

consultant-reported excess returns over benchmarks for all investment categories covered by the consultant in the asset class. The figures 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 our dataset in that asset class. We then weight each consultant’s performance values based on the total 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%.

We compare this (aggregate) claim to our own assessment of the performance of these consultants’ recommendations. For that we use four alternative measures of performance recommendation, all of which involve comparing the returns, or excess returns over benchmarks, of recommended products against those of available institutional products. As we explain in Section 3.2, these are more appropriate measures 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. They are also simple and easily implementable: they allow 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 products and all available institutional products. Formally, for each consultant and quarter t :

$$\sum_i \frac{1}{N_R} r_{i,j,t} I_{i,t-1} - \sum_i \frac{1}{N} r_{i,j,t} \quad (1)$$

where $r_{i,j,t}$ is the return of product i , belonging to investment category j in quarter t , $I_{i,t-1}$ is an indicator variable that equals one when the product is recommended in quarter $t-1$ and zero otherwise, N is the number of available institutional products in quarter t , and N_R is the number of recommended products in quarter t . 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 eVestment 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 available 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 asset manager-chosen benchmarks for recommended and all available institutional products. Formally, for each consultant and quarter t :

$$\sum_i \frac{1}{N_R} (r_{i,j,t} - b_{i,t}) I_{i,t-1} - \sum_i \frac{1}{N} (r_{i,j,t} - b_{i,t}) \quad (2)$$

where $b_{i,t}$ is the return of an asset manager-chosen benchmark for product i in quarter t . We obtain this measure following the same procedure 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 chosen by the asset manager which may or may not be appropriate for the product in question.¹⁴

The third measure we use is similar to the first but uses a matched sample of recommended and available institutional products classified in the same eVestment category (j) for comparison. Formally, for each consultant and quarter t :

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

where N_j is the number of available institutional products in category j and quarter t . In this case risk adjustment is achieved by comparing recommended products against all other available products of similar risk and characteristics according to eVestment. More precisely, the matching is achieved by pairing every recommended product in each quarter with an equal-weighted portfolio of all available institutional products in the same eVestment investment category in the same quarter. In this way, we compare recommended and available products within narrowly defined investment categories.¹⁵ 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 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. Research by Heater (2019) suggests this problem is also present in institutional funds.

¹⁵ eVestment classifies all products into 289 different categories, e.g., U.S. large cap value equity, U.K. small cap growth equity, Euro high yield fixed income.

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

$$\sum_i \frac{1}{N_R} \left((r_{i,j,t} - b_{i,t}) - \sum_{i \in j} \frac{1}{N_j} (r_{i,j,t} - b_{i,t}) \right) I_{i,t-1} \quad (4)$$

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).

We employ two versions of the above performance measures, based on returns gross and net of asset manager fees. 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 3 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 average returns obtained by all products available to plan sponsors. When we risk-adjust returns using manager-chosen benchmarks, recommended products obtain an average excess return before management fees of 0.45% per year (-0.32% per year after

management fees), once annualized.¹⁶ This is indistinguishable from the 0.37% per year (-0.29% per year after management fees) obtained, on average, by available products. The return difference between recommended and available products in the matched sample (a more homogenous comparison of recommended and available institutional products in the same investment categories), at -0.20% per year (or -0.22% per year after management fees) is not statistically significant either. The same is true of the difference between the excess return (over manager-chosen benchmarks) of recommended products and the excess return of available products in a matched sample (differences of -0.18% before fees and -0.22% after fees).¹⁷

The fact that the difference between recommended and available products returns, or excess returns over benchmarks, is in most cases negative and small suggests that investment consultants have, on average, no real ability to pick investment products. This contrasts with consultants' claimed performance, which we have seen is substantial, averaging 1.73% per year for the sample of consultants analyzed. The differences between claimed and observed performance are large and statistically significantly positive, ranging between 1.64% and 2.02% per year depending on the specification and whether the calculations are before or after management fees.¹⁸

¹⁶ Differences between gross and net results reflect not only the effect of asset management fees but are also affected by different sample compositions. As shown in Table 2, not all products report management fees and the absence of this information is particularly prevalent in some asset classes, e.g., alternatives and hedge funds.

¹⁷ Similar results obtain when (i) using a larger sample of consultants including those for which we do not have detailed information about their claims and (ii) measuring returns over longer holding periods, as discussed in the Internet Appendix.

¹⁸ 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.

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 4) ranges from 1.64% per year to 2.49% per year (using the same weighting principles described previously), yet their actual performance according to our calculations is significantly lower and varies between -0.92% per year and 0.29% per year depending on the consultant and the method used to assess that performance (excess returns over benchmarks or matched samples). Table 4 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 in these three aggregates, with the highest claims associated with the equity class (the largest and most important class) at 2.21%, and the lowest for fixed income at 0.93% per year.¹⁹ 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 (except in one case) statistically significant.

3.2 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 sample. We identify four practices employed by consultants in assessing their own performance which together explain most of the discrepancy with our own calculations: comparison of recommended products to benchmarks rather than to other products

¹⁹ When broken down by more granular investment categories, the consultants in our sample claim to outperform in almost all of them. See Internet Appendix for more details.

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 a benchmark index 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 for each consultant in Table 5, where we also quantify the cumulative impact of these practices on the performance claimed by consultants.

It is common practice for consultants to compare the returns of their recommended products against the returns from a benchmark index, and this is done by all of the consultants in our sample. It is, however, a questionable practice in that it does not capture only the value added by consultants, but the joint added value of both 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 institutional investment products which could be recommended. Such a benchmark is not the same as the asset manager-chosen benchmark index for the products.

The practice of using product excess returns over a benchmark index is particularly problematic because, as Gerakos, Linnainmaa, and Morse (2019) show, and our results confirm,

institutional products, on average, outperform their benchmarks.²⁰ It is then no great feat for consultants to select asset managers that on average outperform them, as that would happen even if consultants did not have any manager selection skill. In such cases, a comparison with index returns 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 5 this practice has a significant impact on results, as the performance of consultants appears significantly better using manager-chosen benchmarks: it improves by 0.61% per year for the first consultant (IC1), 1.55% for the second (IC2), and 0.64% for the third (IC3) – the improvement is even stronger when using consultant-chosen benchmarks.²¹ 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

²⁰ 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, Evans and Fahlenbrach, 2012, and Gerakos, Linnainmaa, and Morse, 2019) provide evidence 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). Another 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).

²¹ Some consultants in our sample use their own benchmarks while others rely on benchmarks that are an aggregate of eVestments' manager-chosen benchmarks. In the Internet Appendix we show that consultant-chosen benchmarks are, if anything, marginally more generous than manager-chosen benchmarks. The difference between the two is not statistically significant, however.

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, the case for excluding backfilled returns can 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, recommended products that perform poorly are likely to be under-represented in the backfilled sample. The obvious solution to avoid any biases is to exclude all backfilled returns.

Table 5 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 second 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 leaves a significant gap unexplained.

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).²² This approach, which we refer to as the opportunity to forget losers, 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.

Two of the consultants in our sample (IC1 and IC2) do not include all products and investment categories 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 81.6% of recommended products closed to new investments but actively reporting to eVestment are included in these consultants' own analyses. In contrast, only 51.1% of previously recommended but now 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 include 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%

²² 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.

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 earlier which are shared with the other consultants.

In order to quantify the potential for distortion of each of the four practices we identify, regardless of how many consultants follow them, we next analyze the individual (rather than cumulative) impact of each of them. Although we question the choice by consultants of comparing recommended products against passive benchmarks rather than against other available products, the practice is widespread in the industry, and to reflect this, we display our results both including and excluding the effect of this practice. Table 6 shows that each of the performance measurement practices significantly contributes to enhance the performance figures, except for the seemingly arbitrary inclusion and exclusion of some products and categories. The highest and most significant effect comes from comparing to benchmark indices, rather than to other products in the same category: recommendation performance increases by 0.65% per year in this way (this number is even higher if we base our analysis on consultant-chosen benchmarks, as shown in the Internet Appendix). The use of investment horizons that provide the opportunity to forget losers, leading to survivorship bias, significantly affects excess returns over benchmarks (by 0.48% per year on average) and peers (by 0.38% per year on average). While the use of simulated and backfilled returns has a significant impact on excess returns over benchmarks (the measure used by consultants), at 0.34% per year, it does not significantly affect the excess returns of recommended products over peer products. This is because in this latter case, the bias is introduced in both comparison groups rather than one, and largely cancels out. Arbitrary exclusions seem to have a noticeably larger effect on excess returns over benchmarks rather than excess returns over available peer products, but this effect is not statistically significant.

We have explained a significant part of the discrepancy between our numbers and those produced by consultants (71% for IC1, 67% for IC2, and 54% for IC3). The residual difference is the result of a number of additional factors: different return samples used (eVestment vs. proprietary datasets or dataset obtained from other vendors), different benchmarks (eVestment's asset manager-chosen benchmarks vs. consultant-chosen benchmarks), mismatches in asset class definitions and performance measurement periods (for some asset classes), and differences in weighting strategies (asset class, product, and/or recommendation).²³ In the Internet Appendix we explore some of these factors, which in principle seem more innocuous or less strategic than those highlighted in the paper, and show that they help explain part of the differences between our analysis and consultants', even when individually they do not have a big impact.

Overall, we find significant differences between consultants' claimed performance and our own calculations. The heterogeneity in reporting practices and the resultant lack of monotonicity between consultants' claims and our own calculations suggest that consultants' disclosures are a poor guide not just to their absolute performance but also to their relative performance. It is not surprising that, as the U.K. Competition and Markets Authority survey indicates, many trustees find it difficult to compare the investment track record of rival consultants when tendering for investment consultancy services (CMA, 2018).

²³ In our 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. Not all 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).

3.3 Other Performance Disclosures

Another statistic sometimes reported by consultants is the percentage of their recommended products that outperform benchmark indices 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 over-represented 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 7 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. Consequently, 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). Following this procedure, we find that the proportion of recommended products that outperform benchmarks increases steadily from 63.8% over a one-year horizon to 74.3% over a five-year horizon. This number is lower than that claimed by the consultants in our sample, the difference reflecting arbitrary

exclusions of products and the ratings survivorship bias outlined in Section 3.2. 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 7 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 only marginally better than a random draw, and falls far short of the 76% to 88% claimed by the consultants.

IV. VOLATILITY, TRACKING ERROR, AND THE RISK PROFILE OF RECOMMENDED PRODUCTS

So far, our comparison of recommended and available products has been based on two types of risk adjustment: we have computed benchmark-adjusted product returns, and we have benchmarked recommended products against available 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 available products.

Table 8 reports the average return volatility, tracking error volatility, and benchmark beta of recommended and available 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 available products. Whereas return volatilities tend to be larger for recommended products (although significantly so only in the pre-recommendation period), they are on average marginally smaller when we use a matched sample based on the same eVestment category. In such a sample, recommended products exhibit an average volatility of 12.66% per year and available products 12.71% per year. Betas with respect to manager-chosen benchmarks for recommended and available products are equally indistinguishable. In the same matched sample they stand at 0.94 for recommended product versus 0.98, on average, for all available products in the two year period following the recommendation. These results suggest that there are no significant differences in total or systematic factor risk between the two sets of products, nor any significant differences in leverage. 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 available products is in tracking error volatility. On average, recommended products exhibit lower tracking error volatility than other available products both pre- and post-recommendation. Thus, in the post-recommendation matched sample the average tracking error volatility of recommended products stands at 3.29% per year, compared to 3.54% per year for all available products. This difference is not large but is statistically significant, suggesting that, on average, recommended managers deviate less aggressively from their benchmarks than their peers.

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 fund sponsors are statistically sure that a manager is adding (or subtracting) value. In this context, low tracking error volatility may be desirable because fund sponsors 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.²⁴

Whatever the reason, results in Table 8 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 by some practitioners on the use of information ratios, defined as the ratio of excess returns over benchmarks and tracking errors, in their analysis.²⁵

Consistent with these results, a separate 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. For this analysis, we assign recommended and non-recommended products to different deciles depending on their gross return ranking relative to other products in the same eVestment investment category or asset class. Returns are measured over a calendar year and matched to recommendations at the end of the previous year.

²⁴ 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.

²⁵ It is worth noting that, even when information ratios are used, recommended products in our sample do not outperform: although tracking error is lower (Table 8), this is offset by their lower excess returns over benchmarks (Table 3).

Results presented in Table 9 show that when ranking products within narrowly defined investment categories, 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). The same tendency is present, although in a much weaker form, when ranking within more heterogeneous and broadly defined asset classes (equity, fixed income, etc.). This is consistent with the results presented in Table 8 and may reflect consultants' preference for (more conservative) products with low tracking error volatility: lower tracking errors with respect to narrowly defined benchmarks (consistent with the granular category the products are assigned to) would result in returns that are closer to the median of the distribution. The results could also indicate an ability on the part of consultants to avoid the worst performers. To see whether consultants specifically avoid the worst performers (limit downside risk), or just more volatile products (limit risk), within an investment category, we next turn our attention to negatively recommended products.

V. NEGATIVE RECOMMENDATIONS

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 (as 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 10 shows the average performance of consultants' negatively rated products and the difference between these and all other products in our sample. On average, negatively recommended products outperform their benchmarks by 0.50% per year on a gross of management fees basis.²⁶ This outperformance is as high as that of (positively) recommended products (Table 3). When comparing negatively and non-negatively rated products in a matched sample (that is, category by category), we find that negatively rated products do marginally worse than other products. The differences, however are all small and within the margins of statistical error, which suggest that consultants may not be able to reliably identify future underperformers. The same conclusions obtain if we look at the less informative unmatched sample. 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. Put together with the results of the previous section,

²⁶ This reflects the fact that, as discussed in Section 3.2, the universe of products rated by consultants outperforms, on average, passive indices and is indicative of the problems of benchmarking consultant recommendations in that way.

these results suggest that consultants avoid tracking risk, but that in doing so do not specifically limit downside risk (at least no more than upside risk and very little in total terms).

VI. CONCLUSIONS

Investment consultants tend to overstate their ability to select fund managers. A detailed analysis of the claims they make about the performance of the products they recommend reveals a number of practices that lead to this exaggerated view: comparison of recommended products to selected benchmark indices (which are on average beaten by institutional products, not just those that are recommended) rather than peer products; the inclusion of simulated and backfilled returns in the analysis; the use of assessment practices that provide the opportunity to forget losers, leading to survivorship bias; and seemingly arbitrary exclusions of products and asset classes in their analysis. The distortions generated by these practices – which vary between consultants and are not always disclosed – mean that institutional investors face significant challenges in assessing not just the absolute but also the relative performance of investment consultants.

REFERENCES

- Blake, D., A. G. Rossi, A. Timmermann, I. Tonks, and R. Wermers. 2013. Decentralized investment management: Evidence from the pension fund industry. *The Journal of Finance* 68, 1133-1178.
- van Binsbergen, J. H., M. W. Brandt, and R. Koijen. 2008. Optimal decentralized investment management. *The Journal of Finance* 63, 1849-1895.
- Cassar, G. J., J. J. Gerakos, J. R. Green, and J. R. M. Hand. 2018. Hedge Fund Voluntary Disclosure, *The Accounting Review* 93, 117–35.
- Chava, S., S. Kim, and D. Weagley. 2018. Revealed Heuristics: Evidence from Investment Consultants' Search Behavior. Working paper available at SSRN 3277424.
- Cohen, R. B., P. A. Gompers, and T. Vuolteenaho. 2002. Who underreacts to cash-flow news? Evidence from trading between individuals and institutions. *Journal of financial Economics* 66, 409-462.
- CMA. 2018. Investment consultants market investigation, final report, December. Available at: https://assets.publishing.service.gov.uk/media/5c0fee5740f0b60c8d6019a6/ICMI_Final_Report.pdf
- Cremers, M., and A. Petajisto. 2009. How active is your fund manager? A new measure that predicts performance. *The Review of Financial Studies* 22, 3329-3365.
- Del Guercio, D., and P. 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, R. M. 1999. Investor flows and the assessed performance of open-end mutual funds. *Journal of Financial Economics* 53, 439-466.
- Evans, R. B., and R. Fahlenbrach. 2012. Institutional investors and mutual fund governance: Evidence from retail–institutional fund twins. *The Review of Financial Studies* 25, 3530-3571.
- Fama, E. F., and K. R. French. 2010. Luck versus skill in the cross-section of mutual fund returns. *The journal of finance* 65, 1915-1947.
- FCA. 2016. *Asset Management Market Study Interim Report*, <https://www.fca.org.uk/publication/market-studies/ms15-2-2-interim-report.pdf>

- 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, J., J. T. Linnainmaa, and A. Morse. 2019. Asset Managers: Institutional Performance and Factor Exposures. *The Journal of Finance*, forthcoming.
- Goyal, A., and S. Wahal. 2008. The selection and termination of investment management firms by plan sponsors. *The Journal of Finance* 63, 1805-1847.
- Guo, R-J., B. Lev, and N. Zhou. 2004. Competitive costs of biotech IPOs. *Journal of Accounting Research* 42, 319–355.
- Heater, J. C. 2019. Intended benefits and unintended consequences of improved performance disclosure. Duke University Working Paper.
- James, C., and J. Karceski. 2006. Investor monitoring and differences in mutual fund performance. *Journal of Banking & Finance* 30, 2787-2808.
- Jenkinson, T., H. Jones, and J.V. Martinez. 2016. Picking winners? Investment consultants' recommendations of fund managers. *The Journal of Finance* 71, 2333-2370.
- Jones, H., and J. V. Martinez. 2017. Institutional investor expectations, manager performance, and fund flows. *Journal of Financial and Quantitative Analysis* 52, 2755-2777.
- Jorion, P. 2003. Portfolio optimization with tracking-error constraints. *Financial Analysts Journal* 59, 70-82.
- Lakonishok, J., A. Shleifer, and R. W. Vishny. 1992. The Structure and Performance of the Money Management Industry. Brookings Papers on Economic Activity, 339–391.
- Lang, M. H., and R. J. Lundholm. 2000. Voluntary disclosure and equity offerings: reducing information asymmetry or hyping the stock? *Contemporary accounting research* 17, 623-662.
- Leone, A. J., S. Rock, and M. Willenborg. 2007. Disclosure of intended use of proceeds and underpricing in initial public offerings. *Journal of Accounting Research* 45, 111-153.
- Ljungqvist, A., C. Malloy, and F. Marston. 2009. Rewriting history. *The Journal of Finance* 64, 1935-1960.
- Malkiel, B. G., and A. Saha. 2005. Hedge funds: Risk and return. *Financial Analysts Journal* 61, 80-88.
- Michels, J. 2012. Do unverifiable disclosures matter? Evidence from peer-to-peer lending. *The Accounting Review*, 87, 1385-1413.

- Newey, W. K. and K. D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Patton, A. J., T. Ramadorai, and M. Streatfield. 2015. Change you can believe in? Hedge fund data revisions. *The Journal of Finance* 70, 963-999.
- Price, R. 1999. Who reports earnings when reporting is optional? The market for new franchises. *Journal of Accounting and Economics*, 28, 391-423.
- Roll, R. 1992. A mean/variance analysis of tracking error. *Journal of Portfolio Management*, 18, 13-22.
- SEC. 2013. Regulation of Investment Advisers by the U.S. Securities and Exchange Commission (https://www.sec.gov/about/offices/oia/oia_investman/rplaze-042012.pdf) last accessed 18 July 2019.
- 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, Z, A. Wang, and L. Zheng. 2011. The road less traveled: Strategy distinctiveness and hedge fund performance. *The Review of Financial Studies* 25, 96-143.

Table 1**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 positive recommendations of institutional products issued by the investment consultants in our main sample, and the average number of positive recommendations per investment consultant.

	Number of Products	Number of Asset Managers	Average Number of Products per Asset Manager	Number of Recommendations	Average Number of Recommendations per Investment Consultant
2006	11,413	1,685	6.8	853	284.3
2007	12,060	1,703	7.1	1,122	374.0
2008	12,546	1,720	7.3	1,336	445.3
2009	12,853	1,736	7.4	1,477	492.3
2010	13,151	1,745	7.5	1,611	537.0
2011	13,300	1,706	7.8	1,771	590.3
2012	13,308	1,679	7.9	1,895	631.7
2013	13,197	1,656	8.0	1,938	646.0
2014	12,869	1,620	7.9	2,025	675.0
2015	12,359	1,570	7.9	2,116	705.3
Mean	12,706	1,682	7.6	1,614	538.1

Table 2**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 positive recommendations in each asset class and the percentage of observations in the sample for which we have asset manager fee data available.

	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
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 3**Claimed and Actual 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 consultants' (positively) recommended products, the performance of all products available to institutional investors, and the difference between the two. Product performance is measured using returns and returns in excess of asset manager-chosen benchmarks. The first two columns show results for the full sample of recommended products and assuming products are held only for the period they are recommended (or not) by consultants. The third and fourth columns show results for recommended products and a matched sample of available institutional products classified in the same eVestment category, thus effectively comparing recommended and available 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 (Recommended minus All). *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)
All Products	5.70% (1.17)	0.37% (1.30)	5.67% (1.11)	0.61% (2.41)**
Recommended minus All	-0.30% (-0.52)	0.08% (0.43)	-0.20% (-1.05)	-0.18% (-0.99)
Claimed minus Actual	2.02% [<0.00]***	1.64% [<0.00]***	1.93% [<0.00]***	1.90% [<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)	-0.33% (-0.84)
All Products	5.33% (1.11)	-0.29% (-1.07)	5.41% (1.05)	-0.11% (-0.44)
Recommended minus All	-0.22% (-0.31)	-0.03% (-0.11)	-0.22% (-1.03)	-0.22% (-1.00)
Claimed minus Actual	1.95% [0.01]**	1.75% [<0.00]***	1.95% [<0.00]***	1.95% [<0.00]***

Table 4
Claimed and Actual 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 available products' excess returns over their asset manager-chosen benchmarks and as the difference in returns between recommended products and a matched sample of available institutional 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%	2.49%	1.65%
Actual: Recommended minus All - Average Excess Ret. Over Benchmark	0.29% (0.55)	0.26% (0.58)	0.05% (0.30)
Claimed minus Actual	1.35% [0.02]**	2.23% [<0.00]***	1.61% [<0.00]***
Actual: Recommended minus All - Average Ret. Difference in a Matched Sample	0.10% (0.20)	-0.92% (-1.38)	-0.26% (-1.43)
Claimed minus Actual	1.54% [<0.00]***	3.41% [<0.00]***	1.91% [<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 All - Average Excess Ret. Over Benchmark	0.00% (0.00)	0.34% (1.19)	0.31% (0.34)
Claimed minus Actual	2.21% [<0.00]***	0.59% [0.04]**	0.80% [0.39]
Actual: Recommended minus All - Average Ret. Difference in a Matched Sample	-0.22% (-0.91)	-0.09% (-0.35)	-0.57% (-0.99)
Claimed minus Actual	2.43% [<0.00]***	1.03% [<0.00]***	1.69% [<0.00]***

Table 5
What is Behind the Differences between Claimed and Actual 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 sample we start with the difference in returns between recommended products and a matched sample of available institutional 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	✓	✓	✓
Opportunity to Forget Losers (Rating Survivorship Conditions)	✓		
Arbitrary Inclusion/Exclusions of Products and Investment Categories	✓	✓	
	Cumulative Impact on Measured Performance		
	Consultant 1	Consultant 2	Consultant 3
Actual Performance	0.10%	-0.92%	-0.26%
Comparison to Benchmarks rather than other Products in the Same Category	0.71%	0.63%	0.38%
	[0.02]**	[<0.00]***	[0.02]**
+			
Inclusion of Simulated and Backfilled Returns	0.89%	0.94%	0.77%
	[0.10]*	[<0.00]***	[<0.00]***
+			
Opportunity to Forget Losers (Rating Survivorship Conditions)	1.16%	-	-
	[0.04]**		
+			
Arbitrary Inclusion/Exclusions of Products and Investment Categories	1.19%	1.35%	-
	[0.03]**	[<0.00]***	
Claimed Performance	1.64%	2.49%	1.65%
Residual	0.45%	1.13%	0.89%
	[0.17]	[<0.00]***	[<0.00]***
Residual (as a percentage of gap)	29%	33%	46%
Explained (as a percentage of gap)	71%	67%	54%

Table 6**Performance Measurement Practices: Individual Impact on Results**

This table shows the individual impact of four different performance measurement practices employed by the consultants in our sample on results. The first column shows the impact on our measure of performance (the difference in returns between recommended products and all available institutional products in the same investment category) of each of the following four practices: comparing to passive benchmarks rather than other available institutional products in the same investment category, including simulated and backfilled returns in the analysis, using rating survivorship conditions in drawing up the sample of recommendations used in the analysis, and performing some arbitrary inclusions/exclusions of products and categories in the analysis (restricted to two consultants in the sample). The second column shows how the last three practices mentioned before affect the excess returns over benchmarks of recommended products. Return (performance) differences are expressed in % per year. Statistics are computed on quarterly returns and annualized by multiplying quarterly returns by four. *p*-values for the differences are reported in square brackets and are based on Newey-West standard errors. ***, **, * denote statistical significance at 1%, 5%, and 10% levels, respectively.

	Recommended products excess returns over all available products in the same investment category (Δ%)	Recommended products' excess returns over benchmarks (Δ%)
Comparison to Benchmarks rather than other Products in the Same Category	0.65% [0.01]**	-
Inclusion of Simulated and Backfilled Returns	-0.04% [0.71]	0.34% [0.03]**
Opportunity to Forget Losers (Rating Survivorship Conditions)	0.38% [0.03]**	0.48% [0.03]**
Arbitrary Inclusion/Exclusions of Products and Investment Categories (two ICs)	0.10% [0.54]	0.28% [0.12]

Table 7
**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 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 8**Return Volatility, Tracking Error Volatility, and Beta of Recommended Products**

This table shows the average return volatility, tracking error volatility, and beta of recommended products, available institutional 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 three columns show results for the full sample of recommended and available institutional products. The last three columns show results for a matched sample of recommended and available institutional 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	13.13%	0.89	3.41%	13.09%	0.88	3.40%
All Products	12.58%	0.95	3.82%	13.13%	0.92	3.59%
Recommended minus All	0.56% [<0.00]***	-0.07 [0.66]	-0.42% [<0.00]***	-0.04% [0.42]	-0.03 [0.82]	-0.19% [<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	12.71%	0.94	3.30%	12.66%	0.94	3.29%
All Products	12.70%	0.93	3.82%	12.71%	0.98	3.54%
Recommended minus All	0.01% [0.96]	0.01 [0.95]	-0.52% [<0.00]***	-0.04% [0.36]	-0.04 [0.78]	-0.25% [<0.00]***

Table 9

Return Distribution of Recommended and Non-Recommended Products

This table shows the distribution of recommended and non-recommended product returns within each investment category and asset class by displaying 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 eVestment investment category. 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 asset class. 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 consider only 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.

Decile	By Investment Category									
	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]***
Decile	By Asset Class									
	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]***

Table 10
Performance of Negatively Recommended Products

This table shows the average performance of consultants' negatively rated products and the difference between these and all products in our sample. Performance is measured using returns and returns in excess of asset manager-chosen benchmarks. The first two columns show results for the full sample of negatively rated products and assuming products are held only for the period they are rated as such by consultants. The third and fourth columns show results negatively rated products and a matched sample of available institutional products classified in the same eVestment category, thus effectively comparing negatively recommended and available 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)
All Products	5.70% (1.16)	0.37% (1.17)	6.55% (1.04)	0.77% (1.96)*
Negatively Recommended minus All	0.60% (0.34)	0.13% (0.51)	-0.22% (-0.91)	-0.28% (-1.22)
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)
All Products	5.33% (1.10)	-0.29% (-0.95)	6.23% (0.97)	-0.12% (-0.31)
Negatively Recommended minus All	0.67% (0.35)	-0.03% (-0.12)	-0.15% (-0.59)	-0.20% (-0.79)