How Skilful are US Fixed-Income Fund Managers?

By

Andrew Clare

Keith Cuthbertson

Dirk Nitzsche

Faculty of Finance, Cass Business School,
City, University of London, London. EC1Y 8TZ.

and

Niall O'Sullivan

Cork University Business School, University College Cork, Cork, Ireland

March 2020

Abstract

We develop a performance evaluation model that incorporates the factors proposed by Huij and Derwall (2008) and a fund-specific benchmark to analyse the performance of US fixed income funds. Using the full sample, and accounting for the possibility of false discoveries we find fund management companies extract most of any abnormal performance produced by their fund managers. Our sub-sample analysis indicates that after the Global Financial Crisis (GFC) there was a substantial increase in the number of bond funds with: both positive gross-of-fee alpha and positive net-of-fee alpha performance; and also a reduction in funds with negative-alpha performance. This result indicates that many US bond fund managers anticipated the Quantitative Easing that followed the GFC, positioned their bond funds against their benchmarks accordingly and added value to their portfolios for their investors.

JEL: G11; G12

Keywords: Mutual Fund Bond Performance; False Discovery Rates

---

1. Introduction

The performance of equity mutual funds has been investigated extensively over the last few decades. Many papers document equity mutual fund alphas where excess returns are conditioned on both formal models such as the CAPM (Sharpe (1964)) and also on factor models that derive their validity from more empirical considerations, most notably the Fama-French three factor model (Fama and French 1992, 1993), its momentum-based enhancement (Cahart 1997) and its more recent incarnation, the Fama-French five factor model (Fama and French 2015). Other authors investigating equity mutual funds have focussed on: persistence in mutual fund performance (Carhart 1997); the impact of manager characteristics such as experience (see for example Porter and Trifts (1998), gender (see for example Bliss and Porter (2002) and education (see for example Gottesman and Morey 2006); and the impact of factors such as the location of the fund manager (see for example Oten and Bams 2007) and the "family status" of a fund (see for example Kempf and Ruenzi 2008). Researchers have investigated many other aspects of the performance of equity mutual funds using both US and non-US data as they have sought to establish the value that the fund management industry provides.

By contrast, far fewer papers have focussed on the performance of fixed income (bond) mutual funds. There may be a number of reasons for this. First, arguably, the focus of attention on equity fund returns reflects the higher, historic equity allocations in investor portfolios, particularly in the 1980 and 1990s during the "*cult of equity*". By contrast bond investment was seen to be less relevant and perhaps less interesting. Second, there has been a general lack of agreement in the academic literature about the appropriate factor model to use for fixed-income funds, although neither the CAPM nor the APT are limited to explaining equity returns. However, in the lead up to, and in the wake of the Global Financial Crisis (GFC), financial innovation made more fixed-income asset classes available to investors. Some of these asset classes are complex, as pre-crisis investors in bank subordinated debt tranches will attest when they experienced equity-style losses. Further, the sums of money invested in fixed-income mutual funds is quite substantial, partly because of increased returns after the fall in global interest rates which has only just begun to reverse, but also because some investors have lost faith in equity markets after two major bear markets in the space of less than ten years. Although equity markets recovered in both cases, for anyone drawing an income from a predominantly equity-focused investment portfolio over these periods, portfolio values would have suffered permanent impairment as the perils of sequence risk became all too apparent. More conventional bond markets (that is, excluding investments in, for example, CDOs and subordinated bank debt) have given investors a smoother investment experience, which has in turn led to higher investor allocations into bond mutual funds. Of the $46trn invested in regulated open-end funds, equity mutual funds comprise 43% of this total while bonds and money market funds comprise 35% (2019 Investment Company Factbook, page 11[1]).

This paper contributes to the literature by providing an up-to-date assessment of the performance of over 1,000 US bond mutual funds over the period from January 1998 to May 2018 using three alternative factor models. Our preferred model, which combines factors used by Huij and Derwall (2008) and a fund-specific benchmark, and represents another contribution to the literature, performs particularly well. In calculating the key parameters from the factor models we are careful to account for the non-normality of mutual fund returns by using bootstrap methods (Kosowski et al. 2006, Fama and French 2010, Busse et al. 2010). Finally, we apply the false discovery rate (FDR) methodology to correct for estimated alpha-performance that may in part be due to luck, when analysing performance across many funds (Barras et al 2010). Accounting for false discoveries in overall fund performance may give

[1] https://www.ici.org/research/stats/factbook

different inferences from the standard approach – which normally involves a count of the number of funds which have statistically significant, non-zero abnormal performance (negative or positive), but takes no account of possible false positives.

Our key results are as follows. A four-factor model comprising a fund's (Morningstar) designated benchmark index, a broad market bond index, a high-yield bond index and an index of mortgage-backed bonds captures around 80% of the variation in sectoral bond-fund returns. Over the whole sample period, after accounting for false discoveries, we find that around 25% of funds have positive gross-of-fee alphas but only around 5% of funds have positive net-of-fee alphas (at a 2.5% significance level). This result implies that fund management companies extract most of any abnormal performance produced by their fund managers. Turning to negative performance, very few bond-funds (3%) have truly negative gross-of-fee alphas but this rises to 30% of funds when using net-of-fee alphas, so again, fees lead to a substantial number of underperformers.

Results for bond funds, before and after the Global Financial Crisis (GFC), are clearly of interest given the Quantitative Easing (QE) programmes instigated by many of the world's most important Central Banks. After the GFC we document a substantial increase in the number of bond funds with both: positive gross-of-fee alpha and positive net-of-fee alpha performance; and also a reduction in funds with negative-alpha performance. This result indicates that many US bond fund managers anticipated the QE that followed the GFC, positioned their bond funds against their benchmarks accordingly and added value to their portfolios for their investors.

The remainder of this paper is organized as follows. In section 2 we briefly discuss the related literature; in section 3 we present our models, methodology and data; while our empirical results and conclusions are in sections 4 and 5, respectively.

2. Related Literature

Most previous studies of net-of-fee mutual bond fund returns conclude that the funds do not generate positive alpha. Cornell and Green (1991) study the performance of US "low grade" bond funds defined as those bond funds that have at least two thirds of their holdings in bonds rated BAA or lower by Moody's, or BBB or lower by Standard and Poor's. Their aim was to use this data to investigate the claim of Drexel Burnham Lambert that the risk in holding "junk" bonds was more than offset by higher, risk-adjusted returns. However, to do so they estimate a model of bond mutual fund returns using contemporaneous and lagged values of both the level of the US T-Bill rate and the return on the S&P500. Although the focus of their paper was not alpha, close inspection of their results show that the low-grade bond funds did produce positive alphas. However, it would clearly be possible to argue that the model specification did not capture enough of the sources of risk and return that are typically faced by bond fund managers, such as the slope of the yield curve and by a variety of different credit premia.

Blake et al (1993) estimate both single and multi-index models of bond mutual fund returns using factors that arguably better capture the sources of performance available to bond fund managers. They liken their single index model to the "market model" typically used to evaluate the performance of equity funds, but also imply that the single index essentially represents risk and return that could be achieved by a passive bond fund, that is, by simply holding the components of the index, in their index weights. The single index is the Lehman Brothers US government/corporate bond index. This factor (or index) is very broad indeed and captures, by definition, the broad trends in the market. However, it may produce misleading results for a manager that, for example, only invests in US Treasuries, or only in corporate bonds. In recognition of this, Elton et al (1993) augment the single index model with indices that capture the opportunities to add value that arise from bond maturity and credit spreads between categories. They add the Lehman Brothers intermediate and long-term corporate indices to capture the maturity opportunities and the Lehman Brothers mortgaged backed securities index and the Blume/Keim high yield index to capture credit spread opportunities. They find that estimated net-of-fee alphas are indistinguishable from zero. Indeed, underperformance is found to equal the fees charged by the funds, indicating that bond gross-alpha fund performance does not exceed that of the fund benchmark

Elton et al (1995) develop a model based upon the Arbitrage Pricing Theory (APT). In keeping with the requirements of APT, their model employs forecasts (prepared by economists and investment professionals) to measure unexpected changes in the fundamental economic influences that affect returns. The model comprises factors that capture three broad sources of risk and return. First, the excess return on the US stock market and the excess return on a broad index of bonds capture broader financial market risks. Second, Elton et al include a measure of default risk, a measure of term risk and a measure of mortgage credit risk. This set of factors capture the risk-return opportunities available to bond fund managers. Finally, two factors that appear in Chen, Roll and Ross's (1986) APT model of equity returns are included in the model. These are unexpected changes in inflation and unexpected changes in economic growth. Using this equilibrium model approach Elton et al find negative and statistically significant, net-of-fee alphas in all categories of bond funds examined.

Detzler (1999) examines the performance of a small set of *global* fixed-income mutual funds, and employs a wide set of factors. As well as estimating single index models as Elton et al (1993) do, Detzler estimates three multi-index models incorporating foreign country bond indices and concludes that the sample of global bond funds did not outperform a broad-based US bond index. This suggests that expenses might have outweighed any diversification benefits that might have accrued to the fund over the sample period. A less generous interpretation of the results is that the managers detracted value with their exchange rate-related positions.

Ayadi and Kryzanowski (2011) use bootstrapping techniques to investigate the performance of Canadian mutual bond funds and find, amongst other things, that the performance of the best performing funds is due to "good luck" rather than to skill, and that "bad luck" could explain the performance of the worst performing funds (gross of fees).

A number of authors have investigated the persistence of bond mutual fund performance. Evidence of persistence is weak in earlier sample periods: Philpot et al (1998) and Philpot et al (2000) find short term persistence (over one year) for high yield, global and convertible funds based on contingency table tests, using Sharpe ratios. However, this finding is driven by funds ranked in the middle and lower end of the cross-sectional distribution of Sharpe ratios. Furthermore, the authors find no evidence of persistence over longer, five-year periods. A more recent study by Huij and Derwall (2008) however, does provide evidence of positive performance persistence in US bond mutual funds.

Huij and Derwall (2008) estimate alphas for a large sample of US bond mutual funds. Their multi-factor model comprises the excess returns on a broad investment grade corporate bond index, the excess return on an index of High Yield bonds, and the excess return on an index of Mortgage-backed securities. They argue that estimated alphas correctly measure abnormal returns that are not due passive fund management. Recursive portfolio formation tests indicate that decile net-alphas are generally monotonically decreasing from top to bottom deciles. While the difference between top and bottom decile alphas is a significant 3% p.a., the estimated alphas are generally negative. This indicates that while there is performance persistence, any *positive* persistence is not sufficient to produce returns that would reward investors for choosing an active bond fund manager.

In a recent paper, Clare et al (2019) estimate alphas on a large sample of US bond mutual funds using a single index model and a parsimonious Blake et al (1993) model. Clare et al's model comprises a broad index of bond returns, a measure of the US Treasury yield spread (ie. US ten-year Treasury yield minus one-year US Treasury yield) and a measure of credit conditions (ie. Baa-rated corporate bond yield minus the Aaa-rated bond yield). However, rather than only calculating alphas where the dependent variable is the return on the bond fund minus a proxy for the risk free rate, the authors also estimate alphas where the dependent variable is the return on bond fund-j minus the return generated by bond-j's self-declared benchmark. Calculating excess returns in this way acknowledges that in practice fund manager performance is judged by both employers and by investors against self-declared benchmarks. The authors find evidence in support of long-run abnormal performance (alpha) in the top 10% of funds, but they do not find evidence that positive performance persists.

3. Factor Models, Methodology and Data

3.1 Factor Models

Clearly, an important element in performance measurement is the factor model used. We assess individual US bond fund performance for both gross-of-fees and net-of-fees returns, using Morningstar designated-benchmarks. However, noting that there may be a 'tilt' away from the designated-benchmarks (Sensoy 2009) we augment this approach with other relevant factors. More specifically, we use three models to calculate alphas. We then use these models and the FDR methodology to investigate the proportion of US bond mutual funds that produce performance that is driven by skill, as opposed to being the result of luck – good or bad.

We estimate bond fund alphas using two, single-factor models and also an enhanced version of the three-factor model used by Huij and Derwall (2008). The first single-factor model ("market model") uses a broad index of US bond returns, incorporating both corporate bonds and Treasuries, analogous to the use of a broad index of equities when calculating alphas for equity mutual funds. The model is shown in expression (1):

$$[1] \qquad (R_i - rf)_t = \alpha + \beta_1 (R_m - rf)_t + \varepsilon_{it}$$

where R_i is the monthly return on the mutual fund i for month t and r_{ft} is the risk-free rate (i.e. yield on US 30-day T-Bills); $(R_m - r_f)$ is the excess monthly return on a broad bond market index.

The second, single-factor model uses each fund's (Morningstar) declared benchmark. Performance evaluation based on a publicly available benchmark-adjusted alpha has two key advantages. First, fund managers normally face a number of constraints, in terms of the exposures that they can assume within the fund, for example, to bonds with different credit ratings and different maturities. These constraints are normally reflected in the fund benchmark, allowing the fund manager and their managers to control risks effectively and to undertake meaningful performance attribution analysis. Using a common, catch-all benchmark could therefore result in very misleading conclusions about a fund manager's performance (see Clarke et al 2002, Kothari and Warner (2001) and Angelidis et al (2013)). Second, for the case of equity benchmarks, Cremers et al (2012) show that the benchmarks themselves can produce a non-zero alpha when compared against a broader index like the S&P 500 Composite index. If a manager "benchmark hugs" a benchmark with a non-zero alpha then this could lead to the conclusion that the manager has produced alpha when compared against the catch-all, broader index. This would clearly be an inappropriate conclusion to draw. Therefore, when we use a fund's (self-declared) benchmark-adjusted return the calculated alphas are unbiased in this regard. Few studies have focussed on benchmark-adjusted returns for bond funds. Our second model is specified as follows:

$$[2] \qquad (R_i - rf)_t = \alpha + \beta_2 (R_{bi} - rf)_t + \varepsilon_{it}$$

where $(R_{bi} - r_f)$ is the excess monthly return on the designated-benchmark for bond fund-i.

Our third model uses the Huij and Derwall (2008) three-factors, but in addition we incorporate each fund's (Morningstar) designated-benchmark as a fourth factor. This has the advantage over an approach that would otherwise impose a unitary coefficient on the fund's designated-benchmark (which is implicit if the dependent variable is expressed as the return on fund-j, minus the return on fund-j's designated-benchmark). Our approach does not impose the unitary coefficient constraint. This 4-factor model we denote as "4FHD". Our version of the Huij and Derwall (2008) model comprises three indices: a broad bond market index; a high yield bond index; and a mortgage bond index which we then augment with each fund's designated-benchmark, to give a four-factor Huij and Derwall (4FHD) model, which is specified as follows:

$$[3]\ (R_i - rf)_t = \alpha + \beta_1(R_m - rf)_t + \beta_2(R_{bi} - rf)_t + \beta_3(R_{HY} - rf)_t + \beta_4(R_{Mort} - rf)_t + \varepsilon_{it}$$

where R_{HY} *and* R_{Mort} are the monthly returns on an index of high-yield bonds and mortgage-backed bonds respectively.

3.2 Methodology: False Discovery Rate and Bootstrapping

In preliminary work using the factor models described above, we found that a high proportion of residuals are non-normal. Given that our focus here is on the statistical significance of alpha, we address this issue by calculating p-values using a bootstrap procedure. If the factor model residuals are not normally distributed then the bootstrap p-values, may give different inferences from the standard (parametric) approach.

The false discovery approach asks the question: '*What proportion of statistically significant funds are false discoveries?*'. Storey (2002) and Barras, Scaillet and Wermers (2010) provide a detailed account of the FDR methodology, so it is only briefly summarized below. The null hypothesis that fund-i has no skill in security selection (alpha) and the alternative of either positive or negative performance is:

$$H_0 : \alpha_i = 0 \qquad\qquad H_A : \alpha_i > 0 \text{ or } \alpha_i < 0$$

A 'significant' fund is one for which the p-value for the test statistic (e.g. bootstrapped Newey-West t-statistic on alpha) is less than or equal to some threshold $\gamma / 2$ ($0 < \gamma \leq 1$). At a given significance level γ the probability that a zero-alpha fund exhibits 'good luck' is $\gamma / 2$. If the proportion of truly zero-alpha funds in the population of M-funds is π_0 then the expected proportion of false positives (or 'lucky' funds) is:

[4] $$E(F_\gamma^+) = \pi_0\,(\gamma/2)$$

If $E(S_\gamma^+)$ is the expected proportion of statistically significant positive-alpha funds, then the expected proportion of truly skilled funds (at a significance level γ) is:

[5] $$E(T_\gamma^+) = E(S_\gamma^+) - E(F_\gamma^+) = E(S_\gamma^+) - \pi_0(\gamma/2)$$

The expected FDR amongst the statistically significant positive-alpha funds is:

[6] $$FDR_\gamma^+ = \frac{E(F_\gamma^+)}{E(S_\gamma^+)} = \frac{\pi_0(\gamma/2)}{E(S_\gamma^+)}$$

The observed number of significant funds S_γ^+ provides an estimate of $E(S_\gamma^+)$. To provide an estimate of π_0, (i.e. the proportion of truly null funds in the population of M-funds), we use the result that truly alternative features have p-values clustered around zero, whereas truly null p-values are uniformly distributed, U(0,1). To estimate $\hat{\pi}_0(\lambda)$ we can simply choose a value λ for which the histogram of p-values from all M-funds becomes flat and use:

[7] $$\hat{\pi}_0(\lambda) = = \frac{W(\lambda)}{M(1-\lambda)} = \frac{\#\{p_i > \lambda\}}{M(1-\lambda)}$$

where $W(\lambda)/M$ is the area of the histogram of p-values to the right of the chosen value of λ (on the x-axis of the histogram). An alternative estimate of π_0 is to choose λ to minimize the mean-square error $E\{\pi_0(\lambda) - \pi_0\}^2$ (Storey 2002, Barras et al. 2010)[2] .

[2] Barras et al. (2010) use a Monte Carlo study to show that the estimators outlined above are accurate, are not sensitive either to the method used to estimate π_0 or to the chosen significance level γ. The estimators are also robust to the typical cross-sectional dependence in fund residuals (which tend to be low in monthly data). However, Andrikogiannopoulou and Papakonstantinou (2019) using simulation, show that for US equity mutual fund returns, which have a low signal-to-noise ratio, relatively limited observations per fund and possible cross-sectional correlation across funds, estimates of the false discovery rate may be biased and gives estimates of zero-alphas (non-zero-alphas) that are upward (downward) biased.

We use a bootstrap approach to calculate p-values for our estimated t-statistics because of the non-normality in regression residuals (Politis and Romano (1994); Kosowski et al (2006)). The estimated factor model of returns is: $r_{i,t} = \hat{\alpha}_i + \hat{\beta}_i ' X_t + \hat{e}_{i,t}$ for i = 1, 2, … M funds, where T_i = number of observations on fund-i, $r_{i,t}$ = excess return on fund-i, X_t = vector of risk factors, $\hat{e}_{i,t}$ are the residuals and $\hat{t}_i$ is the (Newey-West) t-statistic for alpha. We draw a random sample (with replacement) of length T_i from the residuals $\hat{e}_{i,t}$ and use these bootstrap residuals $\tilde{e}_{i,t}$ to generate an excess return series $\tilde{r}_{i,t} = 0 + \hat{\beta}_i ' X_t + \tilde{e}_{i,t}$ under the null hypothesis $\alpha_i = 0$. Using $\tilde{r}_{i,t}$ the performance model is estimated and the resulting t-statistic for the alpha-performance measure, $t_i^{b=1}$ is obtained. This is repeated B=10,000 times and for a two-sided, equal-tailed test the bootstrap p-value for the alpha of fund-i is:

$$[8] \qquad p_i = 2.\min[B^{-1}\sum_{b=1}^{B} I(t_i^b > \hat{t}_i), \quad B^{-1}\sum_{b=1}^{B} I(t_i^b < \hat{t}_i)]$$

where $I(.)$ is a (1,0) indicator variable. The above 'basic bootstrap' uses residual-only resampling, under the null hypothesis of no outperformance.[3] A similar procedure is used for other hypothesis tests.

3.4 Data

We apply the bootstrap methodology described in Section 3.3 using monthly fund (total) return data (gross and net of fees), on 1,254 fixed income mutual funds from Morningstar over the period from January 1998 to May 2018. The data set includes both surviving and non-surviving

[3] Alternative bootstrapping procedures such as simultaneously bootstrapping the residuals and the independent variables, or allowing for serial correlation (block bootstrap) or contemporaneous bootstrap across all (existing) funds at time t, produced qualitatively similar results, hence we only report results for the 'residuals only' bootstrap.

funds. To be consistent and to avoid duplication, we collect this monthly data on only the oldest share class of each fund, thereby ensuring that each fund in our database is unique. All of the funds in our sample are categorised by Morningstar as being part of the: Broad, Corporate, Government, Government/Corporate, Municipal, Securitized, Inflation-Protected and Government/Inflation categories. The Broad category is the largest group represented in our database, comprising 621 funds in total. The next three largest categories are Government/Corporate Funds, Government Funds and Corporate Funds comprising 241, 195 and 71 funds in the database, respectively. Overall, these four categories make up almost 90% of the funds in our sample. One of the issues that we explore in this paper is the appropriateness of the designated fund benchmark in capturing the returns generated by the funds. If the designated-benchmarks are indeed meaningful representations of the portfolios constructed by the managers, then we should probably use these benchmarks as a means of identifying skilful from lucky fund managers. Each fund's designated benchmark was also collected from Morningstar.

Table 1 presents descriptive statistics for our sample of bond funds. The statistics in Panel A of Table 1 are based on both total gross and net monthly fund returns in excess of the risk-free rate. The excess gross returns indicate that, compared to the return on T-bills, all fund groups have, on average, outperformed cash over sample period. The corporate bond category produces the highest (excess), average monthly returns over this period of 0.36% per month, although the Govt/Corp sector produced the highest monthly Sharpe ratio of 0.28. The net-of-fee excess returns presented on the right-hand side of Panel A demonstrate the impact of fees. The difference between the gross and net-of-fee average monthly returns for all funds is 0.06% per month (approx. 0.72% p.a.). The equivalent figure for the category with the highest fees, the Corporate category, is 0.69% pa.

Panel B of Table 1 presents statistics analogous to those presented in Panel A, but the excess return is now defined as the fund's return in excess of its designated-benchmark. As would be expected, the average gross excess fund returns over the designated-benchmark (Panel B) are lower than gross fund returns over the T-Bill rate (see Panel A). The average gross returns over the designated benchmark for all categories of bond funds is approximately 1 basis point per month. It is highest for the Govt/Corp category at 2.87 basis points per month, and is negative for the Corporate sector at -3.5 basis points per month. On average, all categories of bond fund produce net-of-fee benchmark-adjusted monthly returns that are negative. For all funds this is -5.06 basis points per month (the right hand side of Panel B). Of course the benchmarks do not incorporate fees, but if equivalent passive vehicles (ETFs, or tracker funds) are available to investors at less than 5.06 bps per month, then on average investors would be better off with passive investment vehicles[4]. However, Table 1 presents average performance figures; one of our goals in this paper is to identify whether there are some funds in this fund universe that produce excess performance that is due to skill, after correcting for non-normality and for any false positives.

4. Results

4.1 Factor Model Results

In Table 2 we present summary results generated by gross (before deduction of fees) and net (after fee) alphas, using the three different factor models described in expressions (1), (2) and (3), over the full sample January 1998 to May 2018. Column 3 in Table 2 shows the proportion of funds where we can dismiss the hypothesis that the residuals of the regressions are normal,

[4]At the time of writing we find that both iShares and Vanguard ETFs, based upon the Barclays Capital US Aggregate Bond Index, are available to investors at an *annual* holding fee of 5 basis points.

based upon the Bera-Jacque test. Because of this, inferences regarding the statistical significance of alpha are based upon the bootstrap procedure described in section 3.2 above.

Panel A in Table 2 presents summary results relating to gross-alphas. Average gross-alphas for all three models are positive. The market model (where the factor is the excess return on the Barclays Capital US Aggregate Bond Index) produces an average gross-alpha of nearly 9 basis points per month, with an average $\overline{R}^2$ of around 54%, while the designated-benchmark model produces an average gross-alpha of just over 3 basis points per month with an average $\overline{R}^2$ of 73%. The difference in the two sets of results should not be surprising. The manager is (or should be) focussed on their benchmark, since this is how their investors will evaluate their performance and how the fund manager's management team will also evaluate their performance.

When we consider the net-of-fee alphas in Table 2, the market model produces an average alpha of just under 3 basis points per month, indicating that on average the active managers are adding value to investor portfolios. However, the designated-benchmark model tells a different story. The average net-alpha is almost *minus* 3 basis points per month. Given that the designated-benchmark model represents the practical hurdle that the managers set themselves, this result paints the asset management industry in a less positive light. The augmented four-factor Huij-Derwall model (4FHD) shows that using only the designated-benchmark model does not capture all of the sources of risk faced by investors in these bond funds. The average $\overline{R}^2$ of the 4FHD model is just over 84%, compared with around 74% for the one-factor, designated-benchmark model. For the 4FHD model the net-alpha is about -4.6 bps per month.

Table 2 also reports the number of funds in the sample that produce positive and positively significant alphas, as well as those that produce negative and negatively significant alphas. The gross-of-fee results (Panel A) show that there are a noteworthy number of funds that do produce positive and significant alphas (at a 2.5% significance level). For example, for the 4FHD model we find that 331 (26.4%) of the 1254 funds produce positive and statistically significant alphas. However, this number falls to 67 (5.3%) for positive net-alphas. At the other end of the scale, the 4FHD model shows that 37 (3%) of funds have negative and statistically significant gross-alphas (Panel A) but a much larger number, namely 369 (29.4%) funds, have negative net-of-fees alphas (Panel B).

In Table 3 we examine the gross and net-of-fees alpha-performance for the four bond fund sectors, as categorised by Morningstar, using the 4FHD model[5]. We find that more residuals are non-normal than is usually reported for equity mutual funds. On average more than 75% of bond funds (see Table 2) display non-normal residuals and for the corporate bond funds that number exceeds 81% (Table 3). To take this into account the results presented in Table 3 are based on bootstrapped p-values. Results based upon gross-of-fee returns are presented in Panel A. The Gov./Corp sector has the highest monthly gross alpha of 2 bps per month and also has the smallest negative net-alpha of *minus* 3.6 bps per month. For the Gov./Corp bond funds, 78 (32%) have positive significant gross-alphas and only 5 (2%) have statistically significant negative gross-alphas. Only 12 funds (5%) have statistically significant positive net-alphas, while 67 funds (28%) have statistically significant negative net-alphas. The largest category is the "Broad" category which comprises 621 funds. We find that 158 (25.4%) of the these bond funds produce positive and significant gross-alphas.

---

[5] We focus on the results from the 4FHD model as it has the highest $\overline{R}^2$. Qualitatively similar results are found for the other two, one-factor models, which are available on request.

In Panel B of Table 3 we see that average net-of-fee alphas are negative for all four sectors. s this figure falls to 32 (5.2%). For example the average net alpha is just under minus 5bps per month for the category of Broad bond funds, compared to an average of just over 1bp per month gross-of-fees. The number of Broad category bond funds producing a positive alpha falls from 418 gross of fees to 183 (29%); while the number producing significantly positive alphas falls from 158 gross of fees to 32 (5%) net of fees. The number of Broad category funds that produce negative alphas rises from 203 to 438 (70%) net of fees of which 109 (17%) produce statistically significant, negative alphas. This pattern is broadly repeated in the other three sectors.

4.2 False Discovery Rates

In the analysis discussed in section 4.1 of this paper we have counted the number of statistically significant positive and negative alpha funds using bootstrapped p-values – but some of these outcomes could be false discoveries. The FDR approach adjusts for the *proportion* of false discoveries amongst those funds which are found to be statistically significant, based on individual (bootstrapped) t-alpha statistics.

Table 4 shows the results of applying the FDR methodology to the 4FHD model on all 1,254 funds over the whole sample period January 1998 to May 2018 (for various significance levels). The histogram of p-values (using the minimum mean square error criterion) determines the optimal λ which then gives the proportion of null (i.e. zero) gross-alpha funds, $\pi_0 = 62.1\%$ and a similar number of null net-alpha funds, $\pi_0 = 53.9\%$. The column S^+ in Table 4 presents the number of statistically significant funds (at $\gamma = 1\%$, 2.5%, 5% and 10% significance levels). For example, (at a significance level of 2.5%) the number of estimated significant positive-alpha funds is $S^+ = 26.4\%$ (331 funds). The estimate of the false discovery rate $FDR^+ = \pi_0 (\gamma/2) / S^+ = 2.94\%$ is small, which implies that the proportion of the M=1,254 funds that are false discoveries is very small at $F^+ = 0.78\%$. Hence, the proportion of "truly" significant gross-alpha funds, $T^+ = S^+ - F^+ = 25.6\%$ (ie. 321 funds out of 1,254), is only slightly diminished after adjustment by the FDR.

Applying the above analysis to the net-of-fee alphas, shown in Panel B of Table 4, gives the proportion of null funds $\pi_0 = 53.88\%$ and a count of statistically significant net-alphas of $S^+ = 5.3\%$ (67 funds) – at a 2.5% significance level. The $FDR^+ = 12.6\%$ which provides a moderate downward adjustment to give the proportion of truly significant positive net-alphas of $T^+ = S^+ - F^+ \approx 5.34\% - 0.67\% = 4.67\%$ (i.e. 59 funds). Hence, correcting for false positives using the FDR does not change our earlier qualitative results with regard to the number of positive gross and net-alphas: $T^+ = 25.6\%$ of fund managers show positive gross-alpha performance but in most cases this positive manager performance is not passed on to investors, because only $T^+ = 4.67\%$ of funds have positive net-alphas.

Turning to the negative alpha funds and applying the FDR correction to negative gross-alpha funds (Table 4, Panel A) gives a count of $S^- = 2.95\%$ (37 funds) with the $FDR^- = 26\%$, resulting in the proportion of truly significant negative gross-alpha funds $T^- = 2.17\%$ (i.e. 27 funds) at the 2.5% level of significance. For negative net-alpha funds (Table 4, Panel B) the count of $S^- = 29.4\%$ (369 funds) is substantial and as the $FDR^- = 2.3\%$ is very small, this results in a substantial proportion of truly significant negative net-alpha funds of $T^- = 28.8\%$ (361 funds). Overall therefore, even after correction for false discoveries, the vast majority of bond fund managers do not produce negative gross-alphas, but negative net-alpha performance is much more prevalent at around 29% of funds.

4.3 Performance Before and After the 2008 Financial Crisis

When we look at the two sub-periods, before and after the 2008 financial crisis, the results for positive alpha-performance funds are very different. For the pre-financial crisis period (January 1998 – August 2008, Table 5 (Panel A)[6], the positive gross-alpha performance shows T^+ = 12.36% (122 funds with) truly significant alphas. In contrast, the post financial crisis period (Table 5, Panel A) reveals a much larger proportion T^+=36.5% (336) funds with truly significant positive gross-alphas. Looking at positive net-alpha funds (Table 5, panel B), these also show an increase between the two periods. In the pre-crisis period, there are only T^+ = 1.45% of funds with truly significant positive, net alphas, but this increases to 10.12% in the post-crisis period.

Turning now to negative gross-alpha performance in the pre- and post-crisis periods, we find T^- = 1.99% and T^- = 0% respectively (Table 5, Panel A), so there are very few funds which underperform their risk factors, on a gross-return basis, in either period. Negative fund performance on a net-alpha basis (Table 5, Panel B) is relatively high in both periods, but it does fall quite dramatically from T^- = 40.8% to T^- = 17% between the two periods. This is mainly due to a substantial fall in the proportion of statistically significant negative net-alpha funds, S^- from 41.2% to 17.8% while the FDR remains small and fairly constant in both periods (Table 5, Panel B).

Overall, even after adjustment for false discoveries (for either gross or net returns), the number of truly significant positive-alpha bond funds (T^+) increased and the number of negative-alpha funds (T^-) fell, in the post-financial crisis period. This may be due to mangers of bond funds

[6] We report results in the main text using a 2.5% significance level.

anticipating the fall in "official" interest rates immediately after the crisis and the repeated use of quantitative easing which reinforced downward pressure on rates.

5. Conclusions

In this paper we have developed a model for evaluating the performance of US bond fund managers, based upon the Huij-Derwall (2008) model which we enhance by incorporating each fund's, fund-specific benchmark. Arguably, the fund-specific benchmark should be the most important component in the evaluation of fund performance because: this is what the manager is trying to outperform; how their remuneration is determined; and how they are judged by both their employers and investors. Using this model and, for the purposes of comparison, a single factor model and the original Huij and Derwall model, we estimate the alphas generated by a set of 1,254 US mutual bond funds over the period from January 1998 to May 2018. However, being cognisant of the fact that residuals from the factor models may be very non-normal, we employ a bootstrap procedure from which we can draw more reliable inferences. Using bootstrapped p-values, we then further refine our results by applying a methodology that accounts for the possibility of "false discoveries", namely that some funds have a statistically significant alpha, which is due to luck rather than to skill.

After estimating robust p-values and accounting for any false discoveries, we find that around 25% of funds have positive gross-of-fee alphas but only around 5% of funds have positive net-of-fee alphas (at a 2.5% significance level). We also find that very few funds (3%) have truly negative gross-of-fee alphas, but that 30% of the sample have negative net-of-fee alphas 30% of funds when using net-of-fee alphas. These results should be of concern to the industry, to regulators and of course to investors. It implies that fund management companies extract most of any abnormal performance produced by the fund. So in answer to the question posed in the title of this paper – *how skilful are US fixed-income fund managers?* – the answer is that there is evidence of skill, but that the industry itself extracts most of the performance benefits of this skill.

We also conducted the same analysis for the pre- and post-GFC periods in our sample. Here the results are more encouraging. In the post-GFC period we find a substantial increase in the number of bond funds with both: positive gross-of-fee alpha and positive net-of-fee alpha performance; and also a reduction in funds with negative-alpha performance (even after accounting for false discoveries). This result indicates that many US bond fund managers anticipated the QE that followed the GFC, positioned their bond funds against their benchmarks accordingly and added value to their portfolios for their investors. However, the GFC was a very unique event, and we would not therefore be able to conclude that active US mutual bond fund industry can provide value for money for its investors, in more normal periods.

References

Andrikogiannopoulou, A. and Papakonstantinou, F. (2019). Reassessing false discoveries in mutual fund performance: skill, luck or lack of power? *Journal of Finance*, 74(5), 2667-87.

Angelidis, T., Giamouridis, D., Tessaromatis, N., 2013. Revisiting mutual fund performance evaluation. Journal of Banking and Finance 37(5), 1759–1776.

Ayadi, M.A., and Kryzanowski, L., 2011. Fixed-income fund performance: Role of luck and ability in tail membership, Journal of Empirical Finance 18, 379–392

Barras, L., Scaillet, O., Wermers, R., 2010. False discoveries in mutual fund performance: Measuring luck in estimated alphas. Journal of Finance 65(1), 179–216.

Blake, C., Elton, E., Gruber, M., 1993. The performance of bond mutual funds. The Journal of Business 66 (3), 371–403.

Blake, C., Elton, E., Gruber, M., 1995. Fundamental economic variables, expected returns, and bond fund performance. The Journal of Finance 50 (4), 1229–1256.

Bliss, R. and Potter, M., 2002. Mutual Fund Managers: Does Gender Matter? The Journal of Business and Economic Studies, Vol. (8), 1-17.

Jeffrey A. Busse, Amit Goyal, And Sunil Wahal, 2010. Performance and Persistence in Institutional Investment Management, The Journal of Finance, Vol. LXV, No. 2.

Carhart, M.M, On persistence in mutual fund performance, 1997. Journal of Finance, Volume 52, Issue 1, March 1, 57-82

Chen, N., R. Roll, and S. Ross, 1986. Economic forces and the stock market, Journal of Business 59, 383-404.

Clare, A., O'Sullivan N, Sherman M, Zhu S, 2019. The Performance of US mutual bond funds, International Review of Financial Analysis 61:1-8.

Clarke, R., De Silva, H., and Thorley, S., 2002. Portfolio constraints and the fundamental law of active management. Financial Analysts Journal 58(5), 48-66.

Cornell, B. and K. Green, 1991. The Investment Performance of Low-Grade Bond Funds Journal of Finance, Vol. 46, No. 1 (Mar., 1991), pp. 29-48

Cremers, M., Petajisto, A., Zitzewtz, E., 2012. Should benchmark indices have alpha? Revisiting performance evaluation. Critical Finance Review 2, 1–48.

Detzler, M. 1999. The performance of global bond mutual Funds, Journal of Banking and Finance, 23, 1195-1217

Fama, E. F., French K, R., 1992. The cross-section of expected stock returns. Journal of Finance 47(2), 427-465.

Fama, E. F., French K, R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33(1), 3-56.

Fama, E. F., French K, R., 2010. Luck versus skill in the cross-section of alpha estimates. Journal of Finance (65)5 , 1915–1947.

Fama, Eugene F., and Kenneth R. French, 2015. A Five-Factor Asset Pricing Model. Journal of Financial Economics, vol. 116, no. 1 (April): 1–22.

Gottesman, A. and Morey M., 2006. Manager education and mutual fund performance, Journal of Empirical Finance, Volume 13, Issue 2, 145-182

Huij, J.J, and Derwall, J., 2008. Hot Hands in Bond Funds. Journal of Banking and Finance, 32(4), 559–572.

Kempf, A. and Ruenzi, S., 2008. Tournaments in mutual fund families. The Review of Financial Studies 21(2): 1013–1036.

Kosowski, R., Timmermann, A., Wermers, R., White, H., 2006. Can mutual fund 'stars' really pick stocks? New evidence from a bootstrap analysis. Journal of Finance (61) 6, 2551-2595.

Kothari, S. P., Warner, J. B., 2001. Evaluating mutual fund performance. Journal of Finance 56(5), 1985-2010.

Otten, R. and Bams, D., 2007. The Performance of Local versus Foreign Mutual Fund Managers, *European Financial Management*, Vol.13, No 4: 702-720.

Philpot, J., Hearth, D., Rimbey, J., and Schulman, C., 1998. Active management, fund size and bond mutual fund returns. The Financial Review, 33, 115-126.

Philpot, J., D. Hearth, J. Rimbey, 2000. Performance persistence and management skill in nonconventional bond mutual funds, Financial Services Review 9, 247-258

Politis, D.N. and J.P. Romano, 1994. The Stationary Bootstrap, Journal of the American Statistical Association, Vol. 89, 1303-1313.

Porter, G.E. and Trifts, J.W., 2014. The Career Paths of Mutual Fund Managers: The Role of Merit, Financial Analysts Journal, Volume 70 Issue 4.

Sensoy, B., 2009. Performance evaluation and self-designated benchmark indexes in the mutual fund industry, Journal of Financial Economics, 2009, vol. 92, Issue 1, 25-39

Sharpe , W.F., 1964. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk The Journal of Finance, Vol. 19, No. 3, 425-442.

Storey J. D., 2002. A Direct Approach to False Discovery Rates, Journal of Royal Statistical Society B, Vol. 64, pp. 497-498.

Table 1: Gross and Net Fund Returns, Summary Statistics (January 1998-May 2018).

This table shows summary statistics for the full sample of 1,254 funds, and for the four largest investment styles. Panel A presents gross and net monthly bond fund returns net of the risk-free rate (rf), while Panel B presents analogous statistics for bond fund returns in excess of each fund's (Morningstar) designated-benchmark return. "Average" is the average monthly percent excess return; SD is the average standard deviation of monthly returns; Sharpe is the average Sharpe Ratio of monthly fund excess returns; Min and Max are the average minimum and maximum percentage monthly fund returns.

Panel A: Gross and Net Returns over the risk-free rate

Category	# Funds	Gross Returns Average	SD	Sharpe	Min	Max	Net Returns Average	SD	Sharpe	Min	Max
All	1254	0.2468	0.2129	0.2425	-2.0367	0.9681	0.1861	0.2139	0.1757	-2.0831	0.9495
Broad	621	0.2634	0.2399	0.2248	-2.0367	0.9681	0.2011	0.2410	0.1689	-2.0831	0.9495
Govt/Corp	241	0.2224	0.1763	0.2810	-1.5610	0.8762	0.1652	0.1779	0.2004	-1.6212	0.8299
Govt	195	0.1772	0.1273	0.2492	-0.2997	0.8521	0.1168	0.1288	0.1481	-0.3310	0.8100
Corporate	71	0.3577	0.3147	0.2465	-1.0759	0.7980	0.2833	0.3205	0.2003	-1.1476	0.7434

Panel B: Gross and Net Returns over Designated-Benchmark Return

Category	# Funds	Gross Returns Average	SD	Sharpe	Min	Max	Net Returns Average	SD	Sharpe	Min	Max
All	1254	0.0101	0.1671	0.0663	-2.3141	0.8041	-0.0506	0.1711	-0.0877	-2.3861	0.6367
Broad	621	0.0023	0.1976	0.0511	-2.3141	0.8041	-0.0599	0.2005	-0.0884	-2.3861	0.6367
Govt/Corp	241	0.0287	0.1387	0.1057	-1.7174	0.5747	-0.0285	0.1412	-0.0631	-1.7775	0.4612
Govt	195	0.0241	0.0915	0.0871	-0.4151	0.6300	-0.0363	0.0984	-0.1011	-0.4867	0.5470
Corporate	71	-0.0350	0.2316	0.0209	-1.2326	0.2015	-0.1104	0.2400	-0.0909	-1.3436	0.1420

Table 2: Gross and Net Alphas for Alternative Factor Models (January 1998 - May 2018).

This table shows average results across all 1,254 funds for gross-alphas (Panel A) and net-alphas (Panel B) for the three factor models: the market model, the designated-benchmark model and the 4-Factor Huij-Derwall (4FHD) model. The average adjusted R-squared across all funds is presented in column 2. The proportion of funds with non-normal residuals (Bera-Jacque test) is shown in column 3. Average alphas are presented in column 4 as % per month (% per annum). The number (#) of funds with positive or negative alphas and those with significant alphas (.) are shown in columns 5 and 6 respectively. Statistically significant alphas use a bootstrapped null distribution at a 2.5% critical value (one-tail test).

	(2) Average Adjusted R-squared	**(3) Proportion funds with non-normal residuals**	**(4) Alpha %pm (%pa)**	**(5) # positive alpha (.) # significant**	**(6) # negative alpha (.) # significant**
Panel A : Gross-Alpha					
Market Model	54.2%	77.53%	0.0884 (1.0605)	962 (440)	293 (20)
Designated-Benchmark Model	73.5%	77.99%	0.0328 (0.3927)	954 (405)	300 (36)
4FHD Model	84.0%	76.32%	0.0149 (0.1787)	874 (331)	380 (37)
Panel B : Net-Alpha					
Market Model	54.2%	77.69%	0.0277 (0.3329)	687 (86)	568 (132)
Designated-Benchmark Model l	73.5%	77.99%	-0.0278 (-0.3332)	526 (109)	728 (259)
4FHD Model	84.0%	76.32%	-0.0457 (-0.5480)	334 (67)	920 (369)

Table 3: Bond Fund Styles, Gross and Net Alphas, 4FHD Model (January 1998 - May 2018).

This table shows average gross-alphas (Panel A) and net-alphas (Panel B) for the 4FHD model (only), for 4 alternative bond styles. Alpha is % per month (% per annum). The number (#) of funds with positive or negative alphas and those with significant alphas (..) are shown in columns 4 and 5 respectively. Statistically significant alphas use a bootstrapped null distribution at a 2.5% critical value (one-tail test). The proportion of funds with non-normal residuals (Bera-Jacque test) is shown in column 6 and the average adjusted R-squared is shown in column 7.

Bond Fund Styles	**(2) # of funds**	**(3) Alpha %pm (%pa)**	**(4) # positive alpha (.) # significant**	**(5) # negative alpha (.) # significant**	**(6) Proportion funds with non-normal residuals**	**(7) Average Adjusted R-squared**
Panel A : Gross Alpha						
Broad	621	0.0134 (0.1603)	418 (158)	203 (19)	73.59%	86.11%
Gov./Corp	241	0.0207 (0.2478)	180 (78)	61 (5)	81.74%	80.85%
Gov.	195	0.0104 (0.1252)	130 (48)	65 (7)	72.31%	78.53%
Corporate	71	0.0111 (0.1337)	51 (22)	20 (2)	81.69%	85.06%
Panel B : Net Alpha						
Broad	621	-0.0487 (-0.5847)	183 (32)	438 (109)	73.75%	86.11%
Gov./Corp	241	-0.0364 (-0.4363)	64 (12)	177 (67)	81.74%	80.83%
Gov.	195	-0.0499 (-0.5993)	31 (8)	164 (100)	72.31%	78.49%
Corporate	71	-0.0631 (-0.7573)	29 (7)	42 (11)	81.69%	85.06%

Table 4 : False Discovery Rate: 4FHD Model, (January 1998 to May 2018)

The augmented Huij Derwall Model (HD4F) includes the designated-benchmark index, an aggregate bond index, a high yield bond index and a mortgage bond index. The figures reported are percentages (of the total number of funds). S+ = proportion (%) of statistically significant positive-alpha funds (at a 1%, 2.5%, 5% and 10% significance levels) based on bootstrap p-values, FDR+=proportion (%) of statistically significant funds that are false discoveries, F+ = proportion of false positive alpha funds and in brackets we report the actual number of funds. 1254 funds are included in the analysis.

	Positive Alpha Funds				**Negative Alpha Funds**			
Panel A: Gross-Alpha π_0 = 0.6209								
Significance level	S^+	FDR^+	T^+	F^+	S^-	FDR^-	T^-	F^-
1%	20.02 (251)	1.55	19.71	0.31	1.52 (19)	20.49	1.20	0.31
2.5%	**26.40 (331)**	**2.94**	**25.62**	**0.78**	**2.95 (37)**	**26.30**	**2.17**	**0.78**
5%	31.18 (391)	4.98	29.63	1.55	5.34 (67)	29.05	3.79	1.55
10%	38.92 (488)	7.98	35.81	3.10	8.77 (110)	35.39	5.67	3.10
Panel B: Net-Alpha π_0 = 0.5388								
Significance level	S^+	FDR^+	T^+	F^+	S^-	FDR^-	T^-	F^-
1%	3.59 (45)	7.51	3.32	0.27	22.41 (281)	1.20	22.14	0.27
2.5%	**5.34 (67)**	**12.61**	**4.67**	**0.67**	**29.43 (369)**	**2.29**	**29.75**	**0.67**
5%	7.34 (92)	18.36	5.99	1.35	36.92 (463)	3.65	35.57	1.35
10%	10.13 (127)	26.60	7.43	2.69	44.82 (562)	6.01	42.12	2.69

Table 5: False Discovery Rate (4FHD Model), Pre- and Post-Crisis Periods, Gross and Net Alphas

The augmented Huij-Derwall Model (4FHD) includes the designated-benchmark index, an aggregate bond index, a high yield bond index and a mortgage bond index. The figures reported are percentages (of the total number of funds). S^+ = proportion (%) of statistically significant positive-alpha funds (at a 2.5% significance level) based on bootstrap p-values, FDR^+=proportion (%) of statistically significant funds that are false discoveries, F^+ = proportion of false positives and in brackets we report the actual number of funds. The pre-financial crises period is January 1998 – August 2008 and the post-financial crisis period is September 2008-May 2018.

		Positive Alpha Funds				Negative Alpha Funds			
	Number of funds	S^+	FDR^+	$\mathbf{T^+}$	F^+	S^-	FDR^-	$\mathbf{T^-}$	F^-
Panel A: Gross-Alpha									
Pre-Crisis π_0 = 0.7636	984	13.31 (131)	7.17	**12.36**	0.95	2.95 (29)	32.39	**1.99**	0.95
Post-Crisis π_0 = 0.5412	945	36.19 (342)	1.87	**36.51**	0.68	0.53 (5)	99	**1**	0.68
Panel B: Net -Alpha									
Pre-Crisis π_0 = 0.3846	984	1.93 (19)	24.90	**1.45**	0.48	41.26 (406)	1.17	**40.78**	0.48
Post-Crisis π_0 = 0.6259	945	10.90 (103)	7.18	**10.12**	0.78	17.78 (168)	4.40	**17.00**	0.78