Predicting Bond Fund Returns

Dale L. Domian, PhD, CFA, CFP
Professor of Finance
School of Administrative Studies
York University
Toronto, Ontario, Canada
ddomian@yorku.ca

William Reichenstein, PhD, CFA
Powers Chair in Investment Management
Hankamer School of Business
Baylor University
Waco, Texas, USA
bill_reichenstein@baylor.edu

Presented at the 2010 Academy of Financial Services Annual Meeting
Denver, Colorado

Predicting Bond Fund Returns

There is disagreement among prior studies as to whether bond fund returns can be predicted. Reichenstein (1999) suggests that these differences in conclusions are likely due to differences in the samples. When the sample contains funds across investment styles, multiple studies conclude that bond fund returns are not predictable. But Reichenstein concludes that when the sample contains funds with similar investment styles, expense ratios can partially predict bond funds' relative returns.

For example, suppose a sample contains high-grade funds with short- to long-term durations. In periods when interest rates rise, short-duration funds will tend to be the better performers. In periods when interest rates fall, long-duration funds will tend to be the better performers. If interest rates either rise in two consecutive periods or fall in two consecutive periods, then funds with above-average returns in the first period are likely to produce above-average returns in the second period. However, when either interest rates rise and then fall or fall and then rise in the two periods, then funds with above-average returns in the first period will generally produce below-average returns in the second. Since returns to bearing interest rate risk are not consistent across periods, relative bond returns should not be predictable in samples that contain funds with substantially different levels of duration risk. Similarly, since returns to bearing credit risk are not consistent across periods, relative bond fund returns should not be predictable in samples that contain funds with substantially different levels of credit risk.

However, in samples that contain taxable bond funds with similar duration risk and credit risk, Reichenstein (1999) concludes that funds with persistently low expense ratios should consistently produce good five-year net returns compared to those on higher-cost funds. Moreover, it is the investor's job to select the investment style. Once a bond style has been

selected, he or she should select a low-cost fund. According to Reichenstein, this strategy should produce persistently strong relative returns.

This study examines the ability of expense ratios and prior five-year returns to predict 1995-1999, 2000-2004, and 2005-2009 net returns on domestic taxable bond funds within the same Morningstar 3x3 fixed-income Style Box. The results are mixed. First, we conclude that funds in the high-grade short-term and medium-grade short-term Style Boxes contained numerous funds with substantially different credit risks and weighted average maturities. That is, these samples do not contain similar funds, so expense ratios and prior returns should not predict differences in their returns. Second, among the other samples, the lowest-cost fund and funds with annual expense ratios below 0.3% usually produced above average returns for the next five-year period. Selection criteria based on the funds' prior five-year returns showed some (but less) success in predicting future five-year bond fund returns in these average-return tests. Third, regression-based tests generally provided some support for the ability of expense ratios to predict returns, but this evidence was not as consistently strong as the support from the average-returns tests. Fourth, and perhaps most importantly, we relate our findings to research from Deng, McCann, and O'Neal (2010, 2011) that convincingly demonstrates problems with the Morningstar star-rating system. In particular, some funds have been gaming the Morningstar system by taking on substantially higher credit risk than their Morningstar Categories would suggest. These are the problems that led us to downplay the results from the high-grade short-term and medium-grade short-term samples. Finally, we show that expense ratios predict a smaller fraction of bond fund returns for 2005-2009 than in prior five-year periods. We attribute this result to the credit crisis. For this period, the variance of bond funds' returns was larger and that additional variability was likely attributable to differences in funds' holding by sector, e.g.,

mortgage, Treasury, etc. If and when markets return to normal, we predict that, once again, expense ratios will be able to predict a relatively large portion of the variance of returns among taxable bond fund with similar investment styles.

The next section describes Morningstar's fixed-income Style Box. We then review the literature on the predictability of bond funds' returns. Next, we describe how we formed our samples. We then present evidence on the abilities of expense ratios and prior returns to predict results for 1995-1999, 2000-2004, and 2005-2009 for samples that contain funds within the same Morningstar 3x3 Style Box. Next, we present results from prior studies that suggest that fund managers have been gaming Morningstar's ratings system. In particular, some fund managers self-reported adjusted durations that implied lower risk than really existed and some fund managers selected portfolios with substantially more credit risk than their average credit rating would suggest. We present evidence that is consistent with these studies' conclusions. The final section presents a summary and conclusions.

Morningstar Style Box

Morningstar separates fixed-income funds into a 3x3 style grid based on the portfolio's average credit rating (a.k.a., Quality) and either duration or self-reported adjusted duration as of a recent date. Exhibit 1 illustrates the Style Box. The numbers denote the final number of funds in each sample.

Duration measures a fund's price sensitivity to changes in interest rates. It is related to maturity. Adjusted duration is designed to reflect a fund's duration after adjusting for hedging strategies and the use of floating rate securities. Each bond fund can self-report its adjusted duration and, if it does, this metric is used to locate the fund within a Style Box. The short,

intermediate, and long portfolios have durations of three years of less, 3.1 to 6.0 years, and more than six years. Duration is designed to measure interest rate risk. Later, we present evidence from a recent study that indicates that Morningstar's system has classified some funds as Short-term or Ultrashort bond funds despite having longer-term average maturities.

Morningstar measures a fund's credit risk based on its average credit rating. This metric is designed to measure the default or credit risk of the fund's portfolio. It is the average bond rating as assigned by Standard & Poor's or Moody's. Morningstar classifies a fund as high, medium, or low grade if the average credit rating is AAA or AA, A or BBB, or BB or lower, respectively. Morningstar (and other firms) calculates the average credit rating based on a linear scoring system. It gives a score of 2 through 8 for bonds rated, respectively, AAA, AA, A, BBB, BB, B, and below B. Thus a fund with all A-rated bonds and another fund with 50% AAA and 50% BB-rated bonds will each receive an average credit rating of A. Later, we present evidence from a recent study that indicates that linear rating systems misrepresent bond funds' true credit risk. In addition, that study concludes that fund managers have gamed the Morningstar rating system by assuming more risk than other funds within its Morningstar Category.

Literature Review on Predictability of Bond Funds' Returns

Some studies including those by Blake, Elton and Gruber (1993) and Philpot et al. (1998) suggest that net returns among taxable bond funds are not predictable. Funds with good Sharpe ratios in the first period are, on average, no more likely to produce good Sharpe ratios in the second period, where "good" may be defined as above average, top quartile, or by another metric.

As discussed in the introduction, Reichenstein (1999) suggests that these results were likely due to these studies' use of samples that contain bond funds across investment styles. Reichenstein (1999) presents a different and what he believes is a more pragmatic approach to testing the predictability of bond funds' returns. As Ellis (1998) and others conclude, it is the investor's job to select an investment style. He may decide to invest in a high-grade intermediate-term bond fund or perhaps a high-grade long-term fund. Once that decision is made, Reichenstein finds a strong negative relationship between 1994-1998 net returns and expense ratios for taxable funds within the same Morningstar 3x3 fixed-income Style Box.

The idea is that funds within the same 3x3 Style Box should contain *similar* levels of interest rate and credit risk. Consider two High Intermediate funds at opposite ends of that Style Box. One has an average credit rating of AAA and average duration of 3.1 years, while the other has an average credit rating of AA and duration of 6 years. These differences are meaningful for short-run returns, but less important in the long run. If interest rates rise sharply one year the shorter duration bond will likely produce the higher return that year. Moreover, for that year, return differences due to differences in duration may swamp return differences due to differences in expense ratios. Similarly, for one year, return differences due to differences in credit risk may exceed return differences due to expense ratios. However, differences in five-year average returns due to differences in duration are likely to be small because longer duration produces lower returns in some year and higher returns in others. The same story applies to differences in five-year average returns due to differences in credit risk. In contrast, a fund with a persistently lower expense ratio is likely to produce a higher five-year net return. Using 1994-1998 net returns, Reichenstein (1999) finds that differences in expense ratios explain a small portion of differences in one-year net returns but a significant portion of differences in five-year net returns

for taxable bond funds within the same Morningstar 3x3 Style Box. In short, for samples containing similar style funds, he predicts that expenses should be able to predict relative returns among taxable bond funds.

Consistent with this prediction, Blake, Elton and Gruber (1993) and Detzler (1999) conclude that, everything else the same, higher expenses lower a fund's net returns. Indeed, since expenses are subtracted from gross fund returns, everything else the same, higher expenses should lead to lower net returns. Therefore, among funds with similar investment styles, we should expect that expense ratios have some ability to predict bond fund returns.

Separately, recent research suggests that prior returns have the ability to predict relative returns. Huij and Derwall (2008) examine the performance persistence of taxable bond funds for 1990-2003. They "find strong evidence of performance persistence in bond funds." Bond funds with good performance in one year tend to perform well in the next year. They write, "Regardless of whether we measure performance by alphas or by the Sharpe ratio, both the probability that a winning fund continues to be a winner and the probability that a loser remains a loser is about 60 percent."

However, they also find that the "vast majority of deciles [of bonds funds] earn negative and statistically significant risk-adjusted returns." That is, they underperform their benchmark portfolios' returns. Only the top decile of funds produces positive alphas, and those alphas were not statistically significant. They conclude that "the difference in return between the funds is largely due to the strong and statistically significant underperformance of the lowest deciles."

Huij and Derwall recommend that "sophisticated investors, with sizable financial capacity, such as institutional investors," form a portfolio consisting of a subset of the top decile of funds based on Modern Portfolio Theory regressions. These portfolios must be adjusted

frequently based on updated regressions. They realize that this strategy is not feasible for private investors. They conclude, "Given our finding that the vast majority of bond funds earn negative and statistically significant risk-adjusted returns, we hasten to comment that index funds are feasible, low-cost alternatives for unsophisticated investors." In short, their study does not suggest that individual investors or investment advisors working with private investors can select bond funds that will outperform index funds.

The Samples

Table 1 presents the number of funds in the final sample for each similar style sample. Data come from the January 2010 version of Morningstar Principia. The high-grade short-term bond funds (henceforth, High Short) consists of all funds that meet the following criteria: Morningstar Fixed-Income Style Box = High-Short; Morningstar Category = Short Government, Ultrashort Bond, or Short-Term Bond; % U.S. Stocks and % non-U.S. Stocks < 2% and % Other assets < 3%; and Special Criteria = distinct portfolios only. In addition, when studying predictability of, respectively, 2000-2004 and 2005-2009 returns, the fund had to have historic returns for the prior five-year period and historic expense ratios for 1999 and 2004. The prior five-year return is needed to test the performance persistence tests and the expense ratio is needed to test the ability of expense ratios to predict bond fund returns.

The fixed-income Style Box reflects the portfolio's investment style as of a recent balance sheet date, while the Category presents its average position over the prior three years. These dual criteria are designed to eliminate funds with style drift. The asset composition criteria--e.g., % U.S. Stocks, % non-U.S. Stocks, and % Other assets--was designed to eliminate balanced funds and other funds that were not true bond funds. The distinct-portfolios-only

criterion excludes all but one share class of a given mutual fund. These samples contain survivorship bias, as discussed in the next section.

The medium-grade short-term bond sample (henceforth, Medium Short) has the same criteria as High Short except it has Fixed-Income Style Box = Medium-Short.

The high-grade intermediate-term sample (henceforth, High Intermediate) has the same criteria except the Fixed-Income Style Box = High-Intermediate and the Category must be Intermediate Government or Interm-Term Bond. The Medium Intermediate sample has Fixed-Income Style Box = Medium-Intermediate.

The High Long sample has the same criteria except the Fixed-Income Style Box = High-Long and the Category must be Long Government or Long-Term Bond. There were too few medium-grade, long-term funds to provide a meaningful sample. Domian and Reichenstein (2008) examine low-grade bond funds returns using Sharpe's return-based style analysis. The analyses suggest that junk bond funds should be viewed as hybrid funds consisting of stocks and bonds. Moreover, the effective stock and bond weights vary considerably among these funds. Consequently, we did not consider low-grade bond funds.[1]

Predictability of Bond Funds' Returns

This section presents predictability tests for three periods, 1995-1999, 2000-2004, and 2005-2009. There are two sets of tests. The first is average-returns tests, while the second is regression-based tests.

1. Occasionally, a fund has a missing 1999 expense ratio, in which case we used the expense ratio for the next earlier year, usually 1998. We excluded the American Century Target Maturity 2010 fund because this fund holds essentially only 2010 Treasury strips. This fund had duration between 10 and 15 years for 1995-1999 and between 5 and 10 years for 2000-2004. It has not been a consistent short-term bond fund.

Exhibit 2 summarizes the average-returns tests. The first panel summarizes the ability of expense ratios to predict 2000-2004 and 2005-2009 returns. The first row of the first panel should be interpreted as follows. The average return among High Short funds for 2000-2004 was 4.78%. The lowest quartile of funds (henceforth, Low Quartile) based on 1999 audited expense ratios produced an average return of 4.99%. The lowest-cost three funds (henceforth, Low 3) earned an average return of 5.45%. Funds with an expense ratio below 0.3% (henceforth, <0.3%) earned 5.14%, on average. The low-cost fund (henceforth, Low 1) earned 4.87%.

Across 2000-2004 and 2005-2009 returns, the average returns on Low Quartile funds exceeded the average return on all funds seven of 10 times. The average return on Low 3 funds also beat its benchmark nine of 10 times. The more select criteria proved best for these periods. The funds with expense ratios below 0.3% produced higher average returns than the average fund in nine of 10 samples and in the other sample its return trailed the average return by a trivial 0.06%. The results for Low 1 are similar.[2] The Low 1 fund's return beat the average return on all funds in nine of 10 samples and in the other sample its average return trailed the average return by a trivial 0.06%. In summary, for all four criteria—Low Quartile, Low 3, <0.3%, and Low 1—expense ratios showed the ability to beat the average returns, but the two most select criteria, <0.3% and Low 1, proved most successful. This is important because, in practice, an individual investor is likely to select one fund per investment style.

The second panel summarizes the ability of prior returns to predict 2000-2004 and 2005-2009 returns. Both the average return on High Quartile funds and the return on High 1 (the fund with the previous highest return) beat the average return on all funds in seven of 10 samples. The average returns on High 3 funds beat the average return on all funds in eight of 10 samples.

[2] For some samples there was only one fund with an expense ratio below 0.3%, so the <0.3% and Low 1 results are identical.

Although these results suggest some ability to predict future returns, their predictive abilities are below the abilities of the Low 1 or <0.3% expense ratio criteria. For example, the High 1 fund from 1995-1999 for the Medium Short sample underperformed this style's average return in 2000-2004 by an economically significant 1.21%, and the High 1 fund for 2000-2004 for the Medium Intermediate sample underperformed this style's average return in 2005-2009 by 2.18%. Although relying on prior returns based on High Quartile, High 3, and High 1 usually produces higher future returns than the style's average return, prior returns could not match the overall success of relying on expense ratios to select funds, especially the <0.3% and Low 1 criteria.

As noted above, the samples contain survivorship bias. That is, the January 2010 Morningstar disk does not show returns for funds that no longer existed as of that date. These non-surviving funds likely had poor returns. So, the average returns on the surviving funds likely exaggerate the average returns of all investors. Prior studies have not found survivorship bias to be a significant problem for bond funds. To the degree it exists, this bias would tend to support the hypothesis that prior winners are likely to repeat and to work against the hypothesis that low-cost funds are likely to produce consistent winners. As discussed earlier and examined in detail later, some bond funds were substantially riskier than other funds in the same 3x3 Style Box. To the degree that fund managers deliberately increased their risk, it is likely that they did so in part to offset higher expense ratios. If that higher risk paid off in consecutive five-year periods then these bond funds would remain in the sample and they would suggest the predictive ability of prior returns. That is top-performing funds in the first period were also top performers in the second. In this case, the top performance might be due to the higher risks that paid off in consecutive five-year periods. If a fund's higher risk paid off in the first five-year period but not the second then it may be eliminated to hide its poor returns. If they had remained in the sample,

they would have suggested that prior returns cannot predict future returns. In short, survivorship bias would likely favor the predictive ability of prior returns. But we find that expense ratios have more predictive ability than prior returns. So, this conclusion would be in spite of the potential survivorship bias, and not because of it.

In summary, the results generally suggest that both expense ratio and prior returns have been useful in predicting five-year bond returns. Moreover, the more select expense ratio criteria proved best. For 2000-2004 and 2005-2009, the <0.3% and Low 1 expense ratio criteria would have allowed an individual investor to essentially match or beat the bond style's average return in every sample.

Table 3 summarizes the regression-based predictability tests. In the top half, expense ratios are used to predict, respectively, 1995-1999, 2000-2004, and 2005-2009 returns. The 1999 audited expense ratio was used to predict 1995-1999 and 2000-2004 net returns, while the 2004 expense ratio was used to predict 2005-2009 returns. Although it would have been preferable to use 1994 expense ratios in 1995-1999 regressions, some funds began in 1995 or did not report expense ratios back to 1994.

In the first regression under 1995-1999, the coefficient on the expense ratio is -0.56%, which indicates that, on average, net returns decrease by 0.56% for each 1% increase in expense ratio. The standard error of the coefficient is 0.15. The next column presents the results of the tests of the null hypothesis that the expense ratio's coefficient is zero. The "5% sig" indicates that the expense ratio's coefficient is significantly different from zero at the 95% confidence level, which implies that expenses have some ability to predict five-year returns. A "Yes" indicates that the null hypothesis is supported at the 95% confidence level. The following column, labeled H: Coef = -1, presents the results of the tests of the null hypothesis that the

expense ratio's coefficient is -1. A -1 coefficient implies that expenses are a dead-weight loss. That is, net returns fall one for one with an increase in the expense ratio. If bond markets are perfectly efficient then the expense ratio should be -1. In Reichenstein (1999) and Domian and Reichenstein (2002), regressions usually support the hypothesis that expenses are a dead-weight loss in, respectively, taxable and tax-exempt bond fund returns. A "5% sig" indicates that the expense coefficient is significantly different from -1 at the 95% confidence level. A "Yes" indicates that the null hypothesis cannot be rejected at the 95% confidence level. The R^2 of 0.17 indicates that 17% of the variance of returns can be explained by the variance of expense ratio.

We draw the following conclusions from the regression tests. First, we recommend that the tests for High Short and Medium Short samples be ignored. Deng, McCann, and O'Neal (2010, 2011) convincingly demonstrate that there were problems with these Morningstar samples. In particular, several funds in Morningstar's Ultrashort and Short-term Categories had average maturities that should have placed them in intermediate or long duration Categories. In addition, some of these funds had substantially more credit risk than their average credit ratings would suggest. Due to these problems, we believe these results should be thrown out.

Second, the regression tests present more stringent tests for the predictive ability of expense ratios and prior returns. Of the remaining nine regressions on expense ratios (after throwing out High Short and Medium Short), the coefficient is always negative, but it is only significantly different from zero in four regressions. Perhaps more important, all nine expense ratio coefficients support the null hypothesis that the expense coefficient is -1. Therefore, these results imply that, on average, higher expenses reduce net returns. Of the remaining six regressions on prior returns, the coefficient is usually positive but only significantly different

from zero in one regression. These results provide weak support for the claim that prior returns are useful in predicting future returns.

Third, both expense ratios and prior returns had some ability to predict returns on High Intermediate and High Long funds for periods before 2005-2009, but their predictive contents, if any, were substantially lower in 2005-2009. Consider first at the results for High Intermediate samples, which is the largest style of funds. Results from Tables 2 and 3 support the following story. Expenses lower net returns one-for-one for these funds. Moreover, it is well known that expense ratios are relatively stable across funds; low-cost funds tend to remain low-cost funds and high-cost funds tend to remain high-cost funds. Therefore, by every criteria in Table 2—e.g., Low Quartile through Low 1 and High Quartile through High 1—expense ratios and prior returns had some ability to predict future returns for both 2000-2004 and 2005-2009. The regression results indicate that expense ratios' predictive ability was statistically significant in both periods, while prior returns' predictive ability was only statistically significant in 2000-2004. Due to the general stability of expense ratios, low-cost funds in 1995-1999 tended to remain low-cost funds in 2000-2004. Therefore, high-return funds in 1995-1999 tended to have low expense ratios and they continued to produce higher returns in 2000-2004. Although expense ratios demonstrated some ability to predict 2005-2009 returns in the regression tests, prior returns could not predict 2005-2009 returns in regression tests.

This same story receives considerable support for the High Long sample. At least for these two investment styles, the results suggest that individuals can garner above average returns within an investment style by selecting one or more low-cost funds. In Table 2, all low-cost and prior returns criteria provided higher-than-average returns for 2000-2004 and 2005-2009. Furthermore, for all time periods the expense coefficient is negative and the prior returns

coefficient is positive. However, only the expense coefficient approaches or attains statistical significance at the 5% level. We suspect that the small sample size prevented the economically significant negative expense ratio coefficients in 2000-2004 and 2005-2009 regressions from being statistically significant.

As discussed before, there are differences in the results from the average-returns tests and the regression tests. In general, the former suggest both expense ratios and prior returns have been useful in predicting returns, while the regression results generally are less supportive of their predictive abilities. Methodologically, neither set of tests is clearly better than the other. The average-return test relies on the average returns of a small subset of funds. In contrast, the regression results depend on all funds' returns. At first glance, this comparison appears to favor the regression tests. However, anyone applying an expense ratio or prior returns filter would likely select the best or one of the best funds based on that criterion. As such, the average-returns results may be more relevant.

For reasons mentioned before and examined later in more depth, we essentially throw out the results for High Short and Medium Short samples. Also, the average credit ratings likely understate the true credit risk of many Medium Intermediate funds. Therefore, it is not surprising that the evidence of predictive abilities in Tables 2 and 3 for Medium Intermediate funds in especially weak. In short, the errors in placing funds in their proper Morningstar Categories as demonstrated by Deng, McCann, and O'Neal (2010) were likely to be especially pronounced for High Short, Medium Short, and Medium Intermediate samples.

Given the financial crisis, the conclusion that net returns were less predictable in 2005-2009 than in earlier periods should not be surprising. If and when we return to more stable

markets, we suspect expense ratios and, to a lesser degree, prior returns, will prove more successful in predicting future returns.

Gaming Morningstar's Style Ratings System

As discussed earlier, Morningstar's rating system is designed to rate funds from one to five stars compared to other funds within the same Morningstar Category. Each Category is intended to contain funds with similar duration risk and credit risks. Duration is measured by either the portfolio's duration or by a fund's self-reported adjusted duration, while credit risk is measured by the average credit rating as calculated by Morningstar. Two recent studies by Deng, McCann, and O'Neal (2010, 2011) reveal shortcomings in Morningstar's ratings system. In particular, fund managers can take on additional credit risk without affecting their Category. These studies show that numerous bond funds in the High Short and Medium Short samples contained substantially more credit risk than other funds in their Category. We discuss shortcomings in the self-reported adjusted duration and average credit rating. By gaming the system, a manager can improve his chances of receiving a coveted four- or five-star rating. Since Morningstar's star ratings strongly influence cash flows into and out of mutual funds, fund managers have a strong incentive to try to game the Morningstar system (see Mamudi, 2010).

Deng, McCann, and O'Neal (2011) reveal a problem of relying on bond funds' self-reported adjusted durations when placing funds in a Category. Several funds in the Ultrashort and Short-term Categories held portfolios with weighted average maturities associated with longer-term bond funds, but still reported low adjusted durations. They may hedge the risk of a rise in Treasury bill rates with interest rate swaps or by holding securities with yields that float based on Treasury bill rates. The problem with this classification system was revealed in the

credit crisis. Many funds in these Categories experienced substantial losses in 2008 as the default risk premium and liquidity risk premium on their bond portfolios increased. Think of an individual bond's yield as consisting of a Treasury yield plus a default risk premium plus a liquidity risk premium. Funds holding bonds with longer weighted average maturities suffered losses that were statistically related to their weighted average maturities; the hedging strategies did not hedge for the increase in default and liquidity risk premiums.

Exhibit 4 comes from Deng, McCann, and O'Neal (2011). It illustrates this Morningstar classification problem. It presents 2008 net returns and weighed average maturities on Ultrashort bond funds. The 2008 returns were negatively associated with the funds' weighted average maturities. For example, the Evergreen Ultrashort Opportunities fund lost about 40% in 2008 despite being in listed as an Ultrashort fund. Its losses could be explained by its weighted average maturity exceeding 25 years. Exhibit 4 suggests that many funds in Morningstar's Ultrashort Category should not have been in the High Short or Medium Short style-box samples. Deng, McCann, and O'Neal also present a similar graph for Short-term bond funds. That graph supports the same story: Many Short-term funds were misplaced in the High Short and Medium Short samples. The self-reporting bias associated with adjusted durations should primarily affect funds in the High Short and Medium Short samples.

In their other study, Deng, McCann, and O'Neal (2010) reveal the problem with any linear rating system such as the one employed by Morningstar and others. As shown in Panel B of Exhibit 5, which comes from Deng, McCann, and O'Neal (2010), Morningstar gives ratings of 2 through 8 to bonds rated AAA, AA, A, BBB, BB, B, and below B. Based on a weighted average of these linear ratings and the fund's holdings across credit rating, Putnam Income Fund receives an average rating of 3.04 as indicated in Panel B. This translates to AA average credit

rating. Based on Moody estimates in Panel C, AA bonds have a 0.2% probability of default within five years. So, investors are led to believe that this Putnam fund has the same credit risk as a bond fund with only AA bonds. Panel C calculates the probability of default on the average bond in the Putnam fund as 3.96%, where this calculation is based on a weighted average of this fund's bond holdings and probabilities of default by credit rating. A 3.96% probability of default corresponds to the credit risk of a bond fund containing only bonds with ratings between BB and BBB. Based on this weighted average probability of default of 3.96%, this fund has almost 20 times the 0.2% default risk on a fund that contains only AA bonds. Moreover, the Putnam fund's true credit risk is comparable to that of a bond fund containing only BB and BBB bonds, while its reported average credit rating suggests its true credit risk is comparable to a bond fund containing only AA bonds. Clearly, the average credit rating misestimates the true credit risk of this Putnam bond fund.

The problem with any linear rating system is that the probability of default rises much faster than linearly as the credit rating decreases. For example, the probability of default increases trivially by 0.08% as the credit rating decreases from AAA to AA, but it increases 8.64% and 19.76% as the credit rating decreases from, respectively, BBB to BB and from BB to B. Any linear rating system will understate the probability of default on the average bond of a fund that holds bonds across a wide breadth of credit ratings.

This explains the bias with any linear rating system. As noted by Deng, McCann, and O'Neal (2010), fund managers can game the Morningstar rating system by increasing their credit risk without having to report the higher risk. They can do that by spreading out the holdings across ratings and by focusing their holding within a letter category in the riskiest third of the ratings within that letter grade. For example, instead of holding only A-rated bonds, a fund

could hold 50% AAA-rated and 50% BB-rated bonds. An example of the second strategy is to hold bonds rated A- or A3 instead of bonds with ratings of A+ or A or A1 or A2. This bias associated with average credit rating could affect funds in any 3x3 Style Box sample.

Exhibit 6 summarizes the results of regressions of High Short, Medium Short, and Medium Intermediate funds' 2005-2009 net returns on a measure of credit risk and, respectively, 2004 audited expense ratios or 2000-2004 returns. Credit risk is calculated as $0.12*w_{AAA} + 0.2*w_{AA} + 0.5*w_{A} + 2.08*w_{BBB} + 10.72*w_{BB} + 30.48*w_{B} + 59.72*w_{belowB} + 20*w_{unrated,}$ where each w denotes the weight by rating and $w_{unrated}$ is the weight on unrated bonds. This formula includes the 5-year default risk percentages shown in Exhibit 5. The Moody's study did not reveal the probability of defaults on unrated bonds. We assigned it a probability of 20%, which is between that of double-B and single-B bonds. The credit risk coefficient is statistically significant at the 95% or better confidence level in the High Short and Medium Short regressions. In these regressions, credit risk is more closely associated with 2005-2009 returns than either prior returns or expense ratios. As noted earlier, Deng, McCann, and O'Neal (2011) conclude that 2008 returns on Ultrashort and Short-term bond funds were negatively associated with the funds' weighted average maturities. They calculated this variable based on the funds' balance sheets in late 2008. Although we did not make this time-consuming calculation, their study clearly shows its significance. Based on Deng, McCann, and O'Neal's studies and the evidence in Exhibit 6, we conclude that the High Short and Medium Short samples contain numerous funds that should not be in these samples. Therefore, we downplay our earlier results from these samples.

The credit risk coefficients in the Medium Intermediate samples are positive but not significant at the 95% confidence level. We suspect that the average credit ratings likely

understate the credit risk of some funds in this sample. Since these samples probably do not contain funds with similar credit risk, it is not surprising that the expense ratios, although negative, failed to achieve statistical significance at the 95% confidence level in these regressions.

Summary and Conclusions

This study examined the ability of expense ratios and prior five-year returns to predict 1995-1999, 2000-2004 and 2005-2009 net returns on domestic taxable bond funds within the same Morningstar 3x3 fixed-income Style Box. We reached several conclusions. First, the High Short (i.e., high-grade short-term) and Medium Short samples contain funds with substantially different styles. In particular, numerous funds self report low adjusted durations, which places them in these samples. However, as demonstrated conclusively by Deng, McCann, and O'Neal (2010, 2011), many of these funds were inappropriately placed in short-term Categories despite having weighted average maturities that should have placed them in intermediate-term or long-term Categories. In addition, many funds had substantially higher credit risks than their average credit ratings suggested. In short, these samples do not contain funds with similar duration and credit risk. Therefore, we tend to ignore predictability tests on these samples.

Second, based on average-return tests, we conclude that expense ratios, and to a lesser extent, prior returns can predict five-year taxable bond fund returns. Investors who followed either of two criteria—select the lowest-cost fund or all funds with annual expense ratios below 0.3%—would have produced at least average returns (and usually above average) for the next five years across every other sample for both 2000-2004 and 2005-2009.

Third, we examined regression-based tests. They generally provided some support for the ability of expense ratios to predict returns, but this evidence was not as consistently strong as the support from the average-returns tests.

Finally, we conclude that expense ratios predicted a smaller portion of returns for 2005-2009 than in prior five-year periods. For this period, the variance of returns within any given sample was much larger. Differences in returns were more closely associated with differences in holding in mortgages and Treasuries. However, for the High Intermediate and High Long samples, the evidence implies that there was still a one-to-one negative relationship between expense ratios and net returns. So, expenses still hurt net returns. However, for this period associated with the credit crisis, the variance of returns within a sample was also influenced by differences in funds' holding by sector—e.g., mortgage or Treasury holdings especially in 2008. If and when markets return to normal, we predict that, once again, expense ratios will be able to predict a relative large portion of the variance of returns among taxable bond funds with similar investment styles.

References

Blake, Christopher R., Edwin J. Elton, and Martin J. Gruber. 1993. "The Performance of Bond Mutual Funds." *Journal of Business*, Vol. 66, No. 3, 371-403.

Deng, Geng, Craig McCann and Edward O'Neal. 2010. "What Does a Mutual Fund's Average Credit Quality Tell Investors?" *The Journal of Investing*, forthcoming.

Deng, Geng, Craig McCann and Edward O'Neal. 2011. "What Does a Mutual Fund's Term Tell Investors?" *The Journal of Investing*, forthcoming.

Detzler, M. L. 1999. "The Performance of Global Bond Mutual Funds." *Journal of Banking and Finance*, Vol. 23, 1195-1217.

Domian, Dale L., and William Reichenstein. 2002. "Predicting Municipal Bond Fund Returns." *Journal of Investing*, Fall, 53-65.

Domian, Dale L., and William Reichenstein. 2008. "Returns-Based Style Analysis of High-Yield Bonds." *Journal of Fixed Income*, Vol. 17, No. 4, Spring, 72-87.

Ellis, Charles D. 1998. *Winning the Loser's Game*, 3rd ed. New York: McGraw-Hill.

Huij, Joop and Jeroen Derwall. 2008. "'Hot Hands' in Bond Funds." *Journal of Banking and Finance*, Vol. 32, 559-572.

Mamudi, Sam. 2010. "Investors Have Stars in Their Eyes." *Wall Street Journal*, June 1, C7.

Philpot, James, Douglas Hearth, James N. Rimbey, and Craig T. Schulman. 1998. "Active Management, Fund Size, and Bond Mutual Fund Returns." *The Financial Review*, Vol. 33, 115-125.

Reichenstein, William. 1999. "Bond Fund Returns and Expenses: A Study of Bond Market Efficiency." *The Journal of Investing*, Winter, 8-16.

Exhibit 1. Morningstar Style Boxes

	Duration			
Short	Intermediate	Long		
92 and 75	142 and 104	8 and 8	High	Quality
15 and 13	28 and 20		Med	
			Low	

The numbers in each box reflect sample sizes for, respectively, tests on 2005-2009 returns, and tests on either 1995-1999 or 2000-2004 returns.

Exhibit 2. Average-Based Predictability Tests

Expense Ratios					
	2000-2004 Returns				
	Average	Low Quartile	Low 3	<0.3%	Low 1
High Short	4.78	4.99	5.45	5.14	4.87
High Inter	6.97	7.49	7.30	7.68	7.16
High Long	9.65	10.20	10.44	10.20	10.52
Med Short	5.15	5.32	5.32	5.55	5.55
Med Inter	7.28	6.99	7.62	8.27	8.27
	2005-2009 Returns				
High Short	3.47	3.32	4.09	4.10	4.18
High Inter	4.11	4.47	4.68	4.64	4.42
High Long	4.80	4.99	5.06	4.99	4.97
Med Short	2.37	-2.22	-3.99	4.29	4.29
Med Inter	4.60	5.16	5.65	4.54	4.54

Prior Returns				
	2000-2004 Returns			
	Average	High Quartile	High 3	High 1
High Short	4.78	5.18	5.54	6.52
High Inter	6.97	7.47	7.80	7.92
High Long	9.65	10.72	10.44	10.93
Med Short	5.15	4.90	4.90	3.94
Med Inter	7.28	7.06	7.26	6.95
	2005-2009 Returns			
High Short	3.47	4.13	4.56	4.77
High Inter	4.11	4.18	4.27	4.98
High Long	4.80	5.09	5.16	5.21
Med Short	2.37	4.63	4.56	6.41
Med Inter	4.60	4.09	4.74	2.42

The numbers in the table are average annual returns (in percent) over the specified five-year period. The upper panel shows funds selected by expense ratios prior to the period. The lower panel shows funds selected by returns over the preceding five years.

Exhibit 3. Regression-Based Predictability Tests

Expense Ratio Regressions					
1995-1999 Returns					
	Expense Ratio	Standard Error	H: Coef = 0	H: Coef = -1	R^2
High Short	-0.56	0.15	5% sig	5% sig	0.17
High Inter	-0.74	0.15	5% sig	Yes	0.19
High Long	-2.91	0.94	5% sig	Yes	0.61
Med Short	-0.27	0.60	Yes	Yes	0.02
Med Inter	-0.71	1.20	Yes	Yes	0.02
2000-2004 Returns					
High Short	-0.25	0.23	Yes	5% sig	0.02
High Inter	-0.96	0.17	5% sig	Yes	0.23
High Long	-1.90	1.24	Yes	Yes	0.28
Med Short	0.82	1.28	Yes	Yes	0.04
Med Inter	-0.42	0.72	Yes	Yes	0.02
2005-2009 Returns					
High Short	0.03	0.40	Yes	5% sig	0.00
High Inter	-0.83	0.40	5% sig	Yes	0.03
High Long	-1.04	0.69	Yes	Yes	0.28
Med Short	7.25	4.42	Yes	Yes	0.17
Med Inter	-0.88	0.88	Yes	Yes	0.04

Prior Return Regressions				
2000-2004 Returns				
	Return 95-99	Standard Error	H: Coef = 0	R^2
High Short	0.59	0.15	5% sig	0.17
High Inter	0.66	0.10	5% sig	0.56
High Long	0.49	0.34	Yes	0.50
Med Short	-0.59	0.62	Yes	0.08
Med Inter	-0.01	0.14	Yes	0.00
2005-2009 Returns				
	Return 00-04	Standard Error	H: Coef = 0	R^2
High Short	0.49	0.16	5% sig	0.10
High Inter	0.06	0.17	Yes	0.00
High Long	0.24	0.43	Yes	0.18
Med Short	1.79	1.13	Yes	0.16
Med Inter	-0.40	0.26	Yes	0.08

H: Coef = 0 and H: Coef = -1 denote the null hypotheses that the coefficient is, respectively, zero and -1. "5% sig" indicates that the null hypothesis can be rejected at 95% confidence level. "Yes" indicates that the null hypothesis cannot be rejected at 95% confidence level.

Exhibit 4 (from Deng, McCann, and O'Neal, 2011): Ultra short term funds losses in 2008 resulted from the credit risk in long term bond holdings.

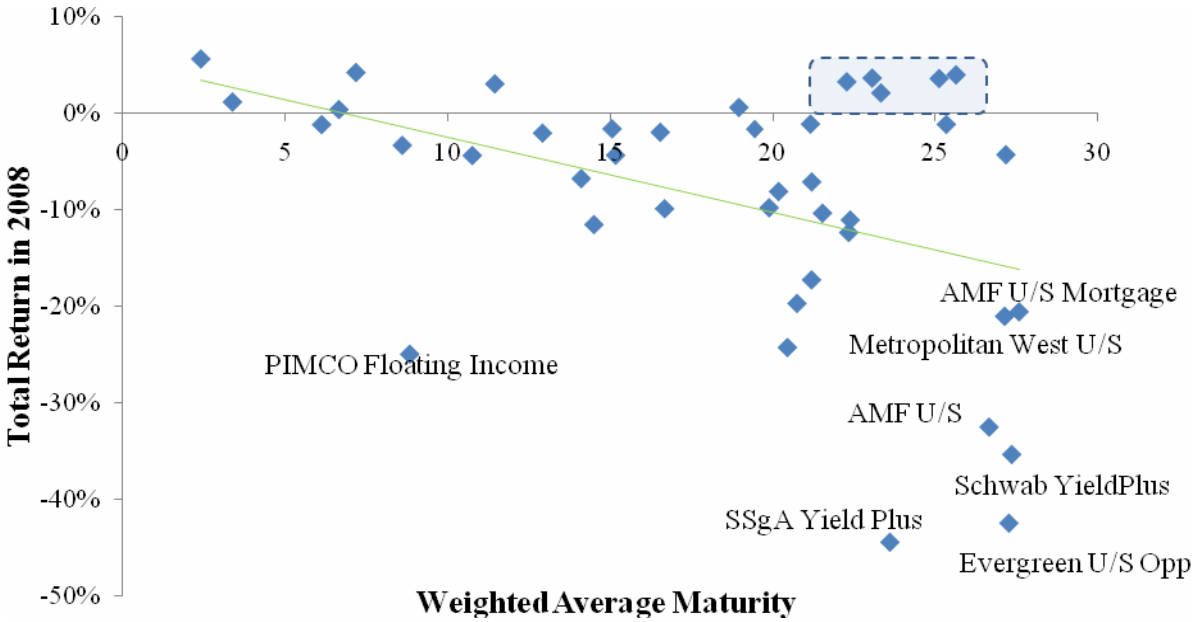

Exhibit 5 (from Deng, McCann, and O'Neal, 2010): Putnam Income Fund self-reports an "AA" Average Credit Quality. Morningstar also reports an Average Credit Quality of AA. Putnam Income Fund actually exposes investor to more credit risk than found in BBB rated bonds. Data as of September 9, 2009.

Rating Category	AAA	AA	A	BBB	BB	B	Below B	Unrated
Fund Holdings	66.5%	3.7%	9.4%	11.3%	1.8%	3.3%	4.0%	
A) Putnam Reported Credit Risk								
Linear Scoring	1	2	3	4	5	6	7	
Weighted Average		2.04: AA						
B) Morningstar Reporting								
Morningstar Scoring	2	3	4	5	6	7	8	
Weighted Average		3.04: AA						
C) Correct Credit Risk								
Moody's 5-Year Default Risk	0.12%	0.20%	0.50%	2.08%	10.72%	30.48%	59.72%	
Correct Credit Risk					3.96: BB to BBB			

Exhibit 6. Multiple Regression Predictability Tests

	2005-2009 Returns				
	Expense Ratio	Standard Error	Credit Risk	Standard Error	R^2
High Short	0.03	0.40			0.00
High Short	-0.01	0.33	-0.52*	0.08	0.32
Med Short	7.25	4.42			0.17
Med Short	6.46	3.89	-0.43*	0.19	0.41
Med Inter	-0.88	0.88			0.04
Med Inter	-0.82	0.85	0.13	0.08	0.13
	2000-2004 Returns	Standard Error	Credit Risk	Standard Error	R^2
High Short	0.50*	0.16			0.10
High Short	0.37*	0.13	-0.49*	0.08	0.37
Med Short	1.79	1.13			0.16
Med Short	1.17	1.10	-0.39	0.21	0.34
Med Inter	-0.40	0.26			0.08
Med Inter	-0.47	0.25	0.16	0.08	0.21

* Denotes coefficient is significantly different from zero at 95% confidence level.