

Don't Take Their Word For It: The Misclassification of Bond Mutual Funds*

HUAIZHI CHEN

University of Notre Dame

LAUREN COHEN

Harvard Business School and NBER

UMIT G. GURUN

University of Texas at Dallas

This draft: April 5, 2020

* *Acknowledgements:* We would like to thank Robert Battalio, Nick Bollen, Robert Burn, Geoffrey Booth, John Campbell, Bruce Carlin, Tom Chang, Christine Cuny, Alex Dontoh, Mark Egan, Ilan Guttman, Yael Hochberg, Samuel Hartzmark, Alan Isenberg, Robert Jackson, Christian Julliard, Oguzhan Karakas, Craig Lewis, Dong Lou, Tim Loughran, Chris Malloy, Bill McDonald, Rabih Moussawi, Bugra Ozel, Jeff Pontiff, Joshua Ronen, Nick Roussanov, Stephen Ryan, David Solomon, Tarik Umar, Ingrid Werner, Bob Whaley, Paul Wildermuth, Paul Zarowin and seminar participants at Drexel University, the London School of Economics, New York University, the University of Notre Dame, Rice University, Vanderbilt University, the 2020 Conference on the Experimental and Behavioral Aspects of Financial Markets, the 2020 Consortium for Asset Management Conference, and the 2020 Review of Asset Pricing Studies Winter Conference for helpful comments and suggestions. We also thank James Ng for providing valuable research assistance. We are grateful for funding from the National Science Foundation, SciSIP 1535813. Please send correspondence to Huaizhi Chen (University of Notre Dame, 238 Mendoza College of Business, Notre Dame, IN 46556, phone: 574-631-3385, email: hchen11@nd.edu), Lauren Cohen (Harvard Business School, Baker Library 279, Soldiers Field, Boston, MA 02163, phone: 617-495- 3888, email: lcohen@hbs.edu), Umit G. Gurun (University of Texas at Dallas, 800 W Campbell Rd, SM 41, Richardson, TX 75080, phone: 972-8835917, email: umit.gurun@utdallas.edu).

Don't Take Their Word For It: The Misclassification of Bond Mutual Funds

ABSTRACT

We provide evidence that bond fund managers misclassify their holdings, and that these misclassifications have a real and significant impact on investor capital flows. In particular, many funds report more investment grade assets than are actually held in their portfolios to important information intermediaries, making these funds appear significantly less risky. This results in pervasive misclassification across the universe of US fixed income mutual funds. The problem is widespread - resulting in up to 31.4% of funds being misclassified with safer profiles, when compared against their true, publicly reported holdings. “Misclassified funds” – i.e., those that hold risky bonds, but claim to hold safer bonds – appear to on-average outperform the low-risk funds in their peer groups. Within category groups, “Misclassified funds” moreover receive higher Morningstar Ratings (significantly more Morningstar Stars) and higher investor flows due to this perceived on-average outperformance. However, when we correctly classify them based on their actual risk, these funds are mediocre performers. These Misclassified funds also significantly underperform precisely when junk-bonds crash in returns. Misreporting is stronger following several quarters of large negative returns.

JEL Classification: G12, G11, G4, G24

Key words: Misclassification, Rating, Fund flow, Fixed income, Mutual Fund.

I. Introduction

Information acquisition is costly. However, the exact cost of collecting any piece of information depends on timing, location, a person's private information set, etc. This is in addition to the idiosyncratic characteristics and complexities of the information signal and of the asset itself. External agents – both public and private - have emerged to fill this role and reduce the cost of information acquisition. However, the value of these agents depends on how much additional information provision is needed. To this end, delegated portfolio management is the predominant way in which investors are being exposed to both equity and fixed income assets. With over 16 trillion dollars invested, the US mutual fund market, for instance, is made up of over 5,000 delegated funds and growing. While the SEC has mandated disclosure of many aspects of mutual fund pricing and attributes, different asset classes are better (and worse) served by this current disclosure level. Investors have thus turned to private information intermediaries to help fill these gaps.

In this paper, we show that for one of the largest markets in the world, US fixed income debt securities, this had led to large information gaps that have been filled by strategic-response information provision by funds. In particular, we show that the reliance on (and by) the information intermediary has resulted in systematic misreporting by funds. This misreporting has been persistent, widespread, and appears strategic – casting misreporting funds in a significantly more positive position than is actually the case. Moreover, the misreporting has a real impact on investor behavior and mutual fund success.

Specifically, we focus on the fixed income mutual fund market. The entirety of the fixed income market is similarly sized to equities (e.g., 40 trillion dollars compared with

30 trillion dollars in equity assets worldwide). However, bonds are both fundamentally different as an asset cash-flow claim, along with having different dynamics in delegated portfolios. While equity funds hold predominantly the same security type (e.g., the common stock of IBM, Apple, Tesla, etc.), each of a fixed income funds' issues differ in yield, duration, covenants, etc. – even across issues of the same underlying firm - making them more bespoke and unique. Moreover, the average active equity fund holds roughly 100 positions, while the average active fixed income fund holds over 600 issues. For example, in Figure 1, we include an excerpt from the AZL Enhanced Bond Index Fund's N-Q Schedule of Investments from September 30, 2018.¹ The fund held over 700 issues, including 7 different bonds of McDonald's Corp – each with differing yields, durations, and callable features. Thus, while the SEC mandates equivalent disclosure of portfolio constituents for equity and bond mutual funds, this data is more complex in both processing and aggregating to fund-level measures for fixed income.

This has led information intermediaries to bridge this gap, providing a level of aggregation and summary on the general riskiness, duration, etc. of fixed income funds upon which investors rely. We focus on the largest of such intermediaries that provides data on categorization and riskiness at the fund level – Morningstar, Inc. In particular, we compare fund profiles provided by the intermediary (Morningstar) to investors against the funds' *actual* portfolio holdings. We find significant misclassification across the universe of all bond funds. This results on up to 31.4% of all funds in recent years, and is pervasive across the funds being reported as overly safe by Morningstar.

¹The full filing, including all eleven pages of holdings, is available here: <https://www.sec.gov/Archives/edgar/data/1091439/000119312518338086/d615188dnq.htm>.

How do these misclassifications occur? Morningstar “rates” each fixed income mutual fund into style boxes based their assessment of credit quality and interest rate sensitivity. For instance, a bond portfolio could be designated as a high credit quality fund with limited interest rate sensitivity. In addition, Morningstar places each fund into a category such as “Multisector Bond,” or “Intermediate Core Bond.” Within each of these fund categories, through a fund’s realized returns and volatility Morningstar then ranks and gives an aggregate rating in the form of “Morningstar Stars.”²

These Morningstar Star summaries of mutual funds have been shown throughout the literature to have a strong and significant impact on investor flow from both retail and institutional investors (Nanda, Wang, and Zheng (2004), Del Guercio and Tkac (2008), Evans and Sun (2018), Reuter and Zitzewitz (2015), Ben-David et al. (2019)).³ In addition, the data releases provided by Morningstar are used ubiquitously throughout the industry.

The central problem that we show empirically, however, is that Morningstar itself has become overly reliant on summary metrics, leading to significant misclassification across the fund universe. In particular, Morningstar requires data provision from each fund it rates (and categorizes) on the breakdown of the bonds the fund holds by risk rating classification. Specifically, what percentage of the fund’s current holdings are in AAA bonds, AA bonds, BBB bonds, etc. One might think that Morningstar uses these

² The ratings methodology and proprietary adjustments and assumptions (e.g., tax burden) are described here: https://www.morningstar.com/content/dam/marketing/shared/research/methodology/771945_Morningstar_Rating_for_Funds_Methodology.pdf, but to a first-order approximation, the rating is determined by their risk and net return categorization (with high expenses detracting from net returns), within official Morningstar Category (included in Appendix D).

³ Investors also respond to other attention grabbing and easy to process external ranking signals, such as Wall Street Journal (Kaniel and Parham, 2017) and sustainability rankings (Hartzmark and Sussman, 2018).

self-reported “Summary Report,” data sent to it by funds to augment the detailed holdings it acquires from the SEC filings on the fund’s holdings. However, Morningstar makes credit risk-summaries solely based on this self-reported data.

Now this would be no issue if funds were accurately passing on a realistic view of the fund’s actual holdings to Morningstar. Unfortunately, we show that this is not the case. We provide robust and systematic evidence that funds on average report significantly safer portfolios than they actually (verifiably) hold. In particular, funds report holding significantly higher percentages of AAA bonds, AA bonds, and all investment grade issues than they actually do. For some funds, this discrepancy is egregious – demonstrably with large holdings of non-investment grade bonds, despite being rated AAA portfolios. Due to this misreporting, funds are then misclassified by Morningstar into safer categories than they otherwise should be.

We define “Misclassified Funds” in a straightforward way: namely as those funds that are classified into a different category than they should be if their *actual* holdings were used as opposed to the self-reported Summary Report percentages that are used to classify them. We show that misclassification is widespread, and continues through present-day, rising up to 31.4% of high and medium credit quality funds in 2018. Moreover, as mentioned above misclassifications are overwhelmingly one-sided: very few misstatements push funds toward a higher risk category – while the vast majority of misstatements push to a “safer” risk category.

So, what are the characteristics of these “Misclassified Funds?” First, Misclassified Funds have higher average risk - and accompanying yields on their holdings - than its category peers. This is not completely surprising, as again Misclassified Funds are holding riskier bonds than the correctly classified peers in their risk category. Importantly, this

translates into significantly higher returns earned on-average by these Misclassified Funds relative to peer funds. They earn 3.04 basis points ($t=3.47$) per month more, implying a 16% higher return than peers.

In order to estimate what portion of this seeming return outperformance of Misclassified Funds comes from skill versus what comes from the unfair comparison to safer funds, we turn to the funds' *actual* holdings reported in their quarterly filings to the SEC. We use these actual holdings to calculate the correct risk category that the fund should be classified into were it to have truthfully reported the percentage of holdings in each risk category. When we re-run the same performance regression specification, but using proper peer-comparisons, we find that Misclassified Funds no longer exhibit any outperformance. In point estimate they even *underperform* by 0.558 basis points per month ($t=0.65$). Thus, it appears that 100% of the apparent outperformance of Misclassified Funds is coming from being misclassified to a less risky comparison group of funds than they should be.

However, the Misclassified Funds still reap significant real benefits from this incorrectly ascribed outperformance. Even after controlling for Morningstar category and risk classification, Morningstar rewards these Misclassified Funds with significantly more Morningstar Stars. In particular, these Misclassified Funds receive an additional 0.38 stars ($t=5.97$), or a 12.3% increase in the number of stars. Armed with higher returns relative to (incorrect) peers and higher Morningstar Ratings, Misclassified Funds then are able to charge significantly higher expenses. In particular, they charge expense ratios that are 11.4 basis points higher than peers ($t=6.35$).

So what are the drivers of misclassifications? Morningstar has posited that it is due nearly entirely to their classification formula's dealing with non-rated bonds.⁴ We show in the Appendix, however, that even kicking out all funds that have any non-rated bonds, all of the results and dynamics remain large and significant (in fact larger in point-estimate in some cases). Looking more closely at the characteristics and behaviors of the non-rated bonds themselves, and the Misclassified Funds that hold them, we find: i.) that the yields of non-rated bonds look incredibly similar to junk bond yields (and very little like the higher rated bonds that they are proposed to be by fund managers, and at which Morningstar takes their word); and ii.) that the Misclassified funds that hold these non-rated bonds curiously underperform precisely when the junk bond market crashes, along with experiencing their greatest fund outperformance when the junk bond market surges (even though they are supposedly holding predominantly highly rated, safe securities).

Lastly, we estimate to what extent misclassification impacts investor behavior. Namely, we examine whether Misclassified Funds – even with higher fees – might attract more investor flows, presumably due to the favorable comparison benefits of being misclassified. We run this in two ways. First, we simply test whether Misclassified Funds receive higher inflows; they do – significantly higher inflows. Next, we run a two stage least squares procedure. In the first stage, we estimate – controlling for other fund, category, and time effects – the impact of being a Misclassified Fund on the number of Morningstar Stars that a fund receives. We then take this estimate of *just* the extra portion of Morningstar Stars a Misclassified Fund gets from being misclassified, and take

⁴ In Section V, we detail our ongoing conversations regarding these large Misclassifications. We have been in contact with Morningstar since we first began the project. Included are their proposed causes of the discrepancies, along with our replies, and evidence on their proposed causes.

this piece of their Stars – Misclassified Stars - to see if it has an impact on investor flows. We find that it has a significantly positive impact. In particular, a one Misclassified Star increase raises the probability of positive flows by almost 17.1% ($t=5.16$).

The remainder of the paper proceeds as follows. Section II provides background for our study, while Section III describes the data, and methodology that Morningstar uses to classify funds into categories. Section IV then presents our main results on the misreporting of funds, and misclassification of these funds by Morningstar based on these faulty reports. Section IV also documents the return implications, along with the real benefits for funds in terms of expenses, Morningstar Stars, and investor flows. Section V then explores non-rated securities, and more of the dynamics of the holdings and behavior of Misclassified Funds, along with discussing Morningstar’s response and proposed causes. Section VI concludes.

II. Background

Our results primarily contribute to three lines of literature. First, our evidence is related to studies on the implications of accuracy and completeness of data sources. Along these lines, Ljungqvist, Malloy, and Marston (2009) show that I/B/E/S analyst stock recommendations have various changes across vintages and these changes (alterations of recommendations, additions and deletions of records, and removal of analyst names) are non-random and likely to affect profitability of trading signals, e.g. profitability of consensus recommendation, among others. Other examples include Rosenberg and Houglet (1974), Bennin (1980), Shumway (1997), Canina et al. (1998), Shumway and Warther (1999), and Elton, Gruber, and Blake (2001). The asset management literature also documents biases in reporting. In the hedge fund setting, Bollen and Poole (2009,

2012) exploit a discontinuity at 0% for reported returns by fund managers (i.e., investors view 0% as a natural benchmark for evaluating hedge fund performance) and document a discontinuous jump in capital flows to hedge funds around this zero-return cut-off. There is also recent work that shows the mutual funds also exhibit considerable variation in their month-end valuations of identical corporate bonds (Cici, Gibson and Merrick, 2011). Similar biases have been shown for valuation of private companies by mutual funds (Agarwal, et al. 2019). Likewise, Choi, Kronland and Oh (2018) show that zero returns are prevalent in fixed income funds and that zero-return reporting is essentially driven high illiquidity of fund holdings.

Second, our finding on the association between misclassification and performance is related to studies on deviations from stated investment policies by equity funds. For example, Wermers (2012), Budiono and Martens (2009) and Swinkels and Tjong-A-Tjoe (2007) show that equity mutual funds that drift from the stated investment objective do better than counterparts. Brown, Harlow and Zhang (2009) and Chan, Chen, and Lakonishok (2002) show that funds that exhibit discipline in following a consistent investment mandate outperform less consistent funds. More recently, Bams, Otten, and Ramezanifer (2017) study performance and characteristics of funds that deviate from stated objectives in the prospectuses. In the equity space, Sensoy (2009) shows that a fraction of size and value/growth benchmark indices disclosed in the prospectuses of U.S. equity mutual funds do not match the fund s actual style.

Lastly, our study contributes to the literatures on style investment. Barberis and Shleifer (2003) argue that investors tend to group assets into a small number of categories, causing correlated capital flows and correlated asset price movements. Vijh (1994) and Barberis, Shleifer, and Wurgler (2005) provide examples using S&P 500 Index membership

changes. Other examples in the empirical literature include Froot and Dabora (1999), Cooper, Gulen, and Rau (2005), Boyer (2011), and Kruger, Landier, and Thesmar (2012), who find that mutual fund styles, industries, and countries all appear to be categories that have a substantial impact on investor behavior (and asset price movements). Our work complements these studies by showing that investors categorize bond funds along the credit risk dimension as provided by the mutual fund industry’s primary data source, Morningstar.

III. Data

In this section, we describe in detail the three major databases used in this paper. Specifically, we combine (1) the Morningstar Direct database of mutual funds and their characteristics, (2) the Morningstar database of Open-Ended Mutual Fund Holdings, and (3) our assembled collection of credit rating histories to document the substantial gap between the reported and the true portfolio compositions in fixed income funds.

III.1 The Morningstar Direct Database

Morningstar Direct contains our collection of fixed income mutual funds. These are the U.S. domiciled, dollar denominated, mutual funds that belong to the “U.S. Fixed Income” global category. We filter out the U.S. government, agency, and municipal bond funds using lagged Morningstar sub-categories. The full collection is 2,029 unique fixed income mutual funds from Q1 2003 to Q2 2018. After applying filters to maintain that 1) more than 85% of each portfolio’s total holdings are observable; 2) the long side of each portfolio is no greater than 115% of its total value; 3) the TNA of each fund is over \$10 million dollars in value and 4) each fund has no more than 35% in holdings on which we

have no ratings information, we have 675 unique funds. Information on these funds also come from Morningstar Direct. This data service contains detailed characteristics that originate both from the regulatory open-ended mutual fund filings and from direct fund surveys.

A key element of our study is the self-reported asset compositions from mutual fund companies. Figure 2 displays the survey used by Morningstar to collect this information from managers. Since the first quarter of 2017, Morningstar began calculating percent asset compositions directly from holdings, but as March 2020, still use the self-reported, surveyed compositions to place fixed income funds in style categories. Notably, we also obtain historical returns, share-level investor flow, and fixed income fund styles from this dataset. For a full list of variables used in this study, refer to Appendix A.

III.2. Open-Ended Mutual Fund Holdings

Our open-ended mutual fund holdings come directly from Morningstar. This service provides us with linkages of portfolio holdings to the Morningstar Direct funds. The fixed income portfolio positions are identified by FundID, Security Name, CUSIP, and Portfolio Date. Along with the identity of these positions, we use portfolio weight, long/short profile, and asset type from this data. We focus on positions that are listed as “Bond” broad-types, and we exclude assets that are listed as swaps, futures, or options.

III.3. Credit Rating Histories

Our analysis centers on the presentation of credit risk in reports heavily used by investors, therefore we collect credit rating histories from a large variety of data sources in order to achieve comprehensive coverage. Due to Dodd-Frank, credit rating agencies

are required to post their rating histories within a year of each ratings announcement as XBRL releases. These releases enable us to achieve coverage by Standard & Poor's, Moody's, and Fitch of all CUSIP-linked securities after June 2012. In addition to these three main NRSROs, we also have coverage of Ambest, DBRS, Egan-Jones, Kroll, and Morningstar credit rating services covering all of the designated US domicile NRSROs during our sample period. We obtain credit ratings for pre-June 2012 from the Capital IQ and the Mergent FISD databases. Capital IQ contains credit rating histories from Standard & Poor's for all of our sample history. In addition, Mergent FISD provides coverage of credit ratings from Moody's, Standard & Poor's, and Fitch on corporates, supranational, agency, and treasury bonds. Table 1 Panel A lists these data sources, the rating agencies reported in these sources, and the time span of their respective coverage. Panel B and Panel C tabulates the actual (as calculated using our credit rating histories) and the reported percentage holding compositions of fixed income mutual funds in the various credit rating categories from Q1 2003 to the end of each respective samples.

IV. Main Results

IV.1. Diagnostics Analysis

We start our analysis by examining histograms of fund reported percentage of holdings minus the calculated percentage holdings in various bond credit rating categories (Figure 3) between Q1 2017 and Q2 2018. The start of this diagnostic sample is dictated by the time that Morningstar began calculating the percent holdings of assets in each credit risk category per each fixed income fund. Ideally, if Morningstar and the bond funds in its database kept the same reporting standards in credit ratings, the fund reported percent should be almost same as the calculated percent holdings. Therefore, these

histograms should report a sharp spike around zero (e.g., no discrepancies), and exhibit no significant variation. This simple diagnostic shows that, on the contrary, there is a wide dispersion of discrepancies between the records of asset compositions. Most notably, for assets above investment grade (above BBB), the percentage of assets reported by funds is markedly higher than the percentage of assets calculated by Morningstar. When we check the same gap for below investment grade and especially in unrated assets, we see an opposite pattern; i.e., the percentage of assets reported by funds is significantly lower than the percentage of assets calculated by Morningstar.

IV.2. Implications of Composition Disagreement - Misclassification

In this subsection, we examine at the major implication of the difference between reported and actual holding implied composition of fund portfolios: namely misclassification of these funds. Figure 4 plots this main result graphically. More specifically, we plot the credit risk distribution of fund-quarter observations between first quarter of 2017 and the end of the second quarter of 2018. The dashed lines represent breaks in the fixed income fund style-box. AAA and AA credit quality funds are high credit quality; A and BBB credit quality funds are medium credit quality; and BB and B are low credit quality as deemed by Morningstar.

The first (blue) bar depicts the distribution of the Morningstar Assigned Credit Risk Category of the fixed income fund. In other words, the blue bar is what mutual fund investors observe if they use Morningstar as a data provider. The second (orange) bar then depicts the same category distribution, however calculated using the fund's self-reported percentage of holdings in the various credit risk categories (from Figure 2). Specifically, using Morningstar's published methodology, this credit risk categorization is

calculated as a function of a nonlinear score assigned to each category by Morningstar (see Appendix B) multiplied by the fund's self-reported percentage of holdings in AAA assets, AA assets, etc. Finally, the third (gray) bar is calculated using the fund's *actual* holdings and their ratings (multiplied by the same scores assigned to each rating type as in the orange bar).

If Morningstar relied on the actual holdings compositions of the funds themselves, the blue bar should track with the gray bar. If, instead, it simply “takes the funds’ word for it,” - simply multiplying the appropriate risk score times the self-reported percentages by the funds - the blue bar would track more closely the orange bar. From Figure 4, the blue bar tracks almost exactly the orange bar. As a result of this, many fixed income mutual funds that would have fallen into a higher credit risk bucket, are classified into safer categories.

More closely comparing these three distributions indicates that using fund self-reported credit risk composition has widely skewed the fund-level credit categorization in favor of lower perceived credit risk. For example, almost half of funds that are marked as A should not be in this category if the fund-level credit rating was assigned based on the actual holdings-implied, rather than self-reported, compositions. Likewise, half of the AAA rated funds should have received a riskier categorization according to the actual calculated holdings. Collectively, the evidence in this subsection suggests that when a fund reports high levels of investment grade assets, it will get classified as an investment grade fund regardless of its actual holdings.

IV.3. Misclassification in Detail

In this section, we explain how systematic patterns of over/under reporting vary with respect to various assumptions regarding (1) how we select our sample and (2) how we match credit ratings to securities. We discuss the baseline analysis in detail and also provide a set of scenarios in Appendix C that numerates the degree of the misclassification ratio in each scenario.

We combine the credit rating history on each fixed income asset in every bond fund portfolio in order to calculate the actual percentage of assets held in each credit risk category. In other words, we match the bond positions of mutual fund portfolios to their respective ratings to calculate their average credit risk classification. These are positions that are listed as “Bond” broad-types in the Morningstar Holdings database. In our baseline analysis, we exclude assets that are listed as swaps, futures, or options, i.e. we don’t assign these assets as a specific rated type or as unrated. When multiple credit rating agencies rate a single asset, we aggregate using the Bloomberg/Barclays method as prescribed by Morningstar’s own methodology document. According to this method, if a security is rated by only one agency, then that rating used as the composite. If a security is rated by two agencies, then the more conservative rating is used. If all three rating agencies are present, then the median rating is assigned. Additionally, government backed securities such as Agency Pass-thru’s, Agency CMO’s, and Agency ARMs are automatically designated as AAA-rated assets. We also search for treasuries and potentially missed government backed securities by searching keywords such as “FNMA”, “U.S. Treasuries”, “REFCORP”, etc. – assigning them each AAA-rating. We then use these holdings calculated compositions to calculate the implied average credit risk. According to this method, roughly 24.1% of bond funds receive counter-factual credit risk categorizations that are riskier than their official credit risk categorizations in the post

2016 sample. In Appendix C, we list the potential assumptions one can make and its corresponding misclassified bond ratio.

In Table 2, we tabulate the time series of Fund Quarter observations in each Morningstar Credit Quality Category using the longest time series we can obtain (2003-2018). The last column, misclassified, is the number of observations that have riskier counter-factual ratings than their official ratings. These numbers suggest that number of misclassified funds increased dramatically over the years but most notably post-August 2010 - the year Morningstar changed the way it calculated average credit risk. Prior to this date, the average credit risk was a simple weighted average of the underlying linear bond scores, in which a AAA bond had a score of 2, AA had a score of 3, and so on. After August 2010, nonlinear penalty scores are assigned to each category. We reproduce the weighting scheme in accordance to Morningstar's published methodology in Appendix B. The result of this change in methodology (as seen in Appendix B) was a much higher relative penalty placed on lower-rated bonds vs. higher-rated bonds. This resulted in a much more composition dependent categorization of fixed income funds (given the drastic ratings penalty-spreads), and an increased number of misclassified funds as according to the Morningstar Fund Style classification as shown in the Table 2. For our main regression analysis, we focus on the sample of funds that are misclassified from Q3 2010, on which Morningstar began its new bond credit risk classification system, to Q2 2018.

IV.4. Fund Performance and Misclassification

A natural follow-up question is whether these misclassified funds are, in fact, different than their risk-category peers, given that they hold a larger percentage of lower

credit-quality assets than their risk category peers (and lower credit-quality assets than their classifications suggest they should be). We explore both the risk and return characteristics of these misclassified funds vs. their correctly classified peers in this section.

In Table 3, we first regress the yield metrics of a fund on our metric of misclassification. Specifically, we define a *Misclassified* dummy variable which takes a value of one if the Morningstar credit quality (High or Medium) is higher than the counterfactual (true) credit quality calculated using the actual underlying holdings, and zero otherwise. We use three different types of yield metrics. In the first column, we use yields reported to Morningstar by the funds themselves. These yields are voluntarily reported. In the second column, we use the yields calculated by Morningstar. The sample size in this second column is limited because calculated holding yields were only available after 2017. In the third column, we use twelve-month yield which combines total interest, coupon, and dividend payments. We also include a credit score variable (the reported compositions score that is used to classify fund credit risks) – with increasing values signifying greater credit risk; and the duration of the bonds (as reported by the funds) as a control variable to capture the interest rate risk of the bond portfolio. In addition, we include a (Time x Morningstar Category) fixed effect to control for common variation in returns and risk due to category-time specific variation (Appendix D lists the official Morningstar Categories). In Columns 1-3, we also include a (Time x Morningstar Reported *Risk Style*) fixed effect to our specification which absorbs the mean yield of each funds corresponding Morningstar fund calculated risk classification in the given year. Doing so allows us to address the concern that a group of funds in a particular year systematically misclassify their riskiness and that misclassified dummy essentially captures this fund

style related reporting choice. We cluster the standard errors by time and fund to address the time series cross-sectional and individual variation in risk.

From Table 3, all three yield columns point to the same empirical dynamic. Namely, that there is a strong relation between misclassification and yields: Misclassified funds have significantly higher yields. The annualized reported yield to maturity is 27.7 basis points higher ($t = 5.49$), whereas the calculated yield from the holdings (second column) and the payout yield are 23.7 and 19.0 basis points higher, respectively, for misclassified funds over their official peers.

In Columns 4-6, we then explore how these misclassified funds would compare were we to compare them against their *correctly* classified risk peers. In particular, for each fund, we use its underlying holdings to calculate its *Correct Fund Risk Style* – note that for already correctly classified funds, this will be the same as Columns 1-3, and only will now be changed, and correctly reflect the risk of the underlying holdings, for misclassified funds.

Columns 4-6 of Table 3 then conduct the identical tests as Columns 1-3, but replace the Time x *Morningstar Reported Risk Style* fixed effect with Time x *Correct Fund Risk Style* fixed effect. From Columns 4-6, the *Misclassified* dummy variable drops in magnitude to near zero and is statistically insignificant. What this means is that when you properly account for the *true* risk of these underlying funds' holdings (based on their actual holdings, as opposed to what they self-report to Morningstar, and that Morningstar classifies risk classification based-upon), they have identical yields to their correct peer funds.

Next, we examine the performance of these misclassified funds vs. their correctly risk-classified peer funds. In Table 4, we regress actual fund returns on the *Misclassified*

dummy, along with the same controls and fixed effects from Table 3. In the Columns 1-2, we include Time x *Morningstar Reported Risk Style* fixed effects as we do in the previous table. From these columns, misclassified funds significantly outperform their Risk Style and Morningstar Fund Category peers, controlling for other determinants of returns. In particular, Column 2 implies that these funds outperform by 3.04 basis points per month ($t=3.42$), which represents a 16% higher return than peers.

In Columns 3-4, we then replace this *Morningstar Reported Risk Style* fixed effect with Time x *Correct Fund Risk Style* fixed effect. The idea is to estimate the percentage of this seeming return outperformance of Misclassified Funds that comes from skill versus what percentage comes from the unfair comparison to safer funds. From Columns 3-4, once we compare Misclassified funds against their correctly classified peers, they exhibit no outperformance. In fact, in point estimate, from Column 4, once compared against their correct risk peers, Misclassified funds actually slightly underperform in point estimate by 0.558 basis points per month ($t=0.65$), though insignificantly so. The sum of the results in Table 4 suggests that Misclassified funds appear to outperform, but that 100% of that outperformance comes from being compared against an incorrect (overly safe) set of category peers.

IV.5. Incentives to Misclassify

In our next analysis, we test whether misclassified funds obtain various benefits from being classified in less-riskier groups of funds. From Table 4, Misclassified funds do appear to generate outperformance to their incorrectly classified risk peers (which disappears when comparing against the correct risk-peer funds). The first benefit we

explore in this section is the awarding of Morningstar Stars by the Morningstar, Inc. itself. As referenced above, Morningstar uses their Star rating system to reward funds for “true outperformance” in their designated Morningstar Category (which are listed in Appendix D). These Morningstar Stars have been shown by a vast literature to have a strong relationship to investor fund flows (for instance, Del Guercio and Tkac (2008), Evans and Sun (2018), Reuter and Zitzewitz (2015), Ben-David et al. (2019)), and by revealed preference are used by many fund companies as an explicit part of their marketing strategy.

We explore this relationship by regressing various Morningstar rating metrics on the *Misclassified dummy*, the reported credit rating score, reported duration, average expense ratio, Time x Morningstar Reported Risk Style fixed effects, and importantly the Time x Morningstar Category fixed effect (as this is the peer group against which Morningstar asserts to make its risk and net return comparison). Because the ratings and expenses are reported at the share class level, the fund level Morningstar Ratings and the Average Expense ratio are calculated as the value weighted average of their respective share-class level values.

The results are reported in Table 5. Table 5 shows that there are economically large increase in Morningstar Stars awarded to *Misclassified* funds. *Misclassified* funds receive 0.17 ($t=3.77$) to 0.38 ($t=5.97$) more Morningstar Stars compared to their peer funds. This level of higher rating corresponds to 18% to 41% of a standard deviation in Morningstar Stars ratings, or up to a 12.3% increase in the number of stars.

In Table 6, we then investigate whether misclassified funds are able to charge higher expense ratios than their peers. Perhaps intuitively, we explore whether *Misclassified* funds charge higher expenses to their investors because their “reported” (but

not actual) performance is better and relatedly that they are able to be rewarded higher Morningstar Star ratings.

Prior research has explored in depth whether equity mutual funds are able to consistently earn positive risk-adjusted returns, and if so, whether funds are able to charge, in equilibrium, higher fees for this outperformance.⁵ The line of argument often suggests that there be a positive relation between before-fee risk-adjusted expected returns and fees. On the other hand, Gil-Bazo and Ruiz-Verdu (2009) argue funds often engage in strategic fee-setting in the presence of investors with different degrees of sensitivity to performance and this could lead to an ambiguous – or even negative - relation between fund performance and fee.

Table 6 contains the results exploring fees of *Misclassified* funds. From Column 3 in Table 6, we find that, on average, the misclassified funds have 7.6 basis point higher ($t = 4.17$) average annual expenses than funds within the same style-category, which implies they are able to charge 10.8% higher fees than peers.⁶

In Table 7, we then investigate the fund flows to *Misclassified* funds. There are several reasons why misclassification might be related to bond fund flows. First, Barberis and Shleifer (2003) argue that investors tend to group assets into a small number of categories, causing correlated capital flows and correlated asset price movements. If an

⁵ See, for example, Brown and Goetzmann (1995); Carhart (1997); Daniel et al. (1997); Wermers (2000); Cohen, Coval, and Pastor (2005); Kacperczyk, Sialm, and Zheng (2005); Kosowski et al. (2006).

⁶ Past research in the equity space has investigated whether funds alter their investment style and whether funds with characteristics are more likely to deviate from stated objectives in their mandate due to various reasons including fund manager incentives. In particular, DiBartolomeo and Witkowski (1997) show that younger mutual funds are particularly prone to misclassification and Frijns et al. (2013) show that funds which switch across fund objectives aggressively tend to have higher expense ratios. Along these lines, Huang, Sialm and Zhang (2011) argue that funds with higher expense ratios experience more severe performance consequences when they alter risk.

asset ends up being in the wrong classification category then it may receive a disproportionately higher (or lower) investment than its correct bucket – especially if it has a favorable ranking attribute within that category (e.g., reported returns). Several papers in the literature show the power of style investment in explaining asset flows. Froot and Dabora (1999), Cooper, Gulen, and Rau (2005), Boyer (2011), and Kruger, Landier, and Thesmar (2012), find that mutual fund styles, industries, and countries all appear to be categories that have a substantial impact on investor behavior (and asset price movements).

We test for the relationship between Misclassification and flows in two ways. First, we simply test whether *Misclassified* Funds receive higher inflows; they do – significantly higher inflows. This is shown in Column 1 of Table 7. However, given that Misclassification is also related to other attributes which drive flows (e.g., Morningstar Stars), it is difficult to interpret what magnitude of the flows are coming from the *Misclassification* itself. Next, we run a two stage least squares procedure. In the first stage, we estimate – controlling for other fund, category, and time effects – the impact of being a *Misclassified* Fund on the number of Morningstar Stars that a fund receives (run in Table 5). We then take this estimate of *just* the extra portion of Morningstar Stars a *Misclassified* Fund gets from being misclassified, and take this piece of their Stars – *Misclassified Stars* - to see if it has an impact on investor flows. We find that it has a significantly positive impact. In particular, Column 2 of Table 7 implies that a one *Misclassified Star* increase raises the probability of positive flows by almost 17.1% ($t=5.16$).

From investor behavior with respect to these *Misclassified* funds vs. other funds, we turn to examining the characteristics that correlate with a fund’s decision to misclassify

their self-reported holdings. We define two variables to capture the change in fund reporting behavior over time. The first variable, *Starts*, takes a value of one if a previously correctly classified fund starts misclassifying its holdings. In addition to this variable, we define another indicator variable, *Stops*, which takes a value of one if a previously misclassified fund starts correctly classifying its holdings. The negative coefficient for 3-year return, -0.125 ($t = 2.74$), indicates that funds with worse performance, in comparison to other funds, are more likely to start misclassifying their holdings (Table 8, Column 1). Likewise, the corresponding positive coefficient in the second column, 0.180 ($t = 2.11$), suggests funds with better past returns are more likely to *Stop* misclassification. Collectively the results in Table 8 suggest that self-reported *Misclassification* does vary with recent past fund performance.

V. Misclassified Funds Return Dynamics, Non-Rated Securities, and Morningstar's Response

We have been in contact with Morningstar since the beginning of the project. We were first referred to technical support teams with whom we checked each step of our process and the self-reported surveys that fund managers fill out, along with Morningstar's scoring process, to ensure that we had each step correct. Then, following the first posting of a draft of our work, Morningstar released an official organizational response shown in Appendix E. In Appendix F, we include our reply to Morningstar's initial comments. Morningstar then responded with a second response contained in Appendix G, along with our reply to these comments in Appendix H.

Essentially, Morningstar posited two points in their initial response. First, that the star analysis in particular was mis-specified due to not comparing within Morningstar

Official Fund Category (Appendix D). As seen in the current draft, all specifications include official *Morningstar Category* fixed effects. From these tests, comparing within categories, all of our results are strong and significant. Which is to say: *Misclassified* funds receive significantly more stars than peer-group funds within an official Morningstar Category. Second, Morningstar posited that the discrepancies are due nearly entirely to their classification formula’s dealing with non-rated bonds. We show in Appendix F, however, that even kicking out all funds that have *any* non-rated bonds, all of the results and dynamics remain large and significant (in fact larger in point-estimate in some cases).

We then look more closely at the characteristics and behaviors of the non-rated bonds themselves, and the Misclassified Funds that hold them. First, we look at the non-rated bonds themselves in Table 9. From Table 9, the yields of non-rated bonds look incredibly similar to junk bond yields, and very little like the higher rated bonds that they are proposed to be by fund managers, and at which Morningstar takes their word.

Second, in Table 10, we examine the performance of *Misclassified* funds around times of junk bond crashes, and junk bond outperformance. If these classified into “safer” categories by Morningstar truly did hold the safe, high credit-quality bond issues they claimed – and represented by Morningstar in their relatively safe risk classifications of the funds – the funds should not be sensitive to the movement of junk bonds. However, this is not what is seen in Table 10. Table 10 shows that *Misclassified* funds’ over- and under-performance relative to their peer funds relates strongly to junk bond returns (captured by the return on a junk bond index – JNK). *Misclassified* funds significantly underperform precisely when the junk bond market crashes, along with experiencing their greatest fund outperformance when the junk bond market surges (even though they are supposedly holding chiefly highly rated, safe securities).

Morningstar’s second reply (Appendix G) then shifts focus to more technical points, stating: “*To that end, we were able to largely reproduce the authors’ multivariate analysis of the binary “misclassified” dummy variable they defined and various ratings metrics.*” In Appendix H, we explore the points and claims from this response in further detail in the data, unfortunately not finding strong support.

VI. Conclusion

Investors rely on external information intermediaries to lower their cost of information acquisition. While *prima facie* this brings up no issues, if the information that the intermediary is passing on is biased, these biases propagate throughout markets and can cause real distortions in investor behavior and market outcomes. We document precisely this in the market for fixed income mutual funds. In particular, we show that investors’ reliance on Morningstar has resulted in significant investment based on verifiably biased reports by fund managers that Morningstar simply passes on as truth.

We provide the first systematic study that compares fund reported asset profiles provided by Morningstar against their *actual* portfolio holdings, and show evidence of significant misclassification across the universe of all bond funds. A large portion of bond funds are not passing on a realistic view of the fund’s actual holdings to Morningstar and Morningstar creates its risk classifications, and even fund ratings, based on this self-reported data. Up to 31.4% of all funds in recent years, are reported as overly safe by Morningstar. This misreporting has been not only persistent and widespread, but also appears to be strategic. We show that misclassified funds have higher average risk - and accompanying yields on their holdings - than their category peers. We also show evidence suggesting the misreporting has real impacts on investor behavior and mutual fund success.

Misclassified funds reap significant real benefits from this incorrectly ascribed outperformance in terms of being able to charge higher fees, receiving “extra” undeserved Morningstar Stars within peer groups, and ultimately receiving higher flows from investors.

We exploit a novel setting in which investors reliance on external information intermediaries can lead to predictable patterns in fund ratings and capital flows, and in which we can ex-post verify the veracity of the information conveyed. We believe that our study is a first step to think about a market design in which information intermediaries have more aligned incentives to better process and deliver the information they gather from market constituents. Future research should explore alternate monitoring and verification mechanisms for the increasingly complex information aggregation in modern financial markets, along with ways that investors can engage as important partners in information collection and price-setting.

References

- Agarwal, V., Barber, B.M., Cheng, S., Hameed, A. and Yasuda, A., 2019. Private company valuations by mutual funds. Available at SSRN 3066449.
- Barberis, N., Shleifer, A., 2003. Style investing, *Journal of Financial Economics* 68, 161–199.
- Brown, Stephen J., and William N. Goetzmann, 1997. Mutual fund styles, *Journal of Financial Economics* 43, no. 3, 373-399.
- Brown, Keith C., Harlow, W. Van and Zhang, Hanjiang, 2009. Staying the course: The role of investment style consistency in the performance of mutual funds. Available at SSRN 1364737.
- Bams, Dennis, Otten, Roger, and Ramezanifar, Ehsan, 2017. Investment style misclassification and mutual fund performance. In 28th Australasian Finance and Banking Conference.
- Ben-David, Itzhak and Li, Jiacui and Rossi, Andrea and Song, Yang, 2019. What do mutual fund investors really care about?, Fisher College of Business Working Paper No. 2019-03-005.
- Bennin, Robert, 1980. Error rates in CRSP and COMPUSTAT: A second look, *Journal of Finance* 35, 1267–1271.
- Boyer, Brian H., 2011. Style-related comovement: Fundamentals or labels?, *Journal of Finance* 66, 307-332.
- Bollen, Nicholas and Pool, Veronica, 2009. Do hedge fund managers misreport returns? Evidence from the pooled distribution, *Journal of Finance*, 2257-2288.
- Bollen Nicholas and Pool, Veronica, 2012. Suspicious patterns in hedge fund returns and risk of fraud, *Review of Financial Studies* 25, 2673-2702.
- Budiono, Diana and Martens, Martens, 2009. Mutual fund style timing skills and alpha. Available at SSRN 1341740.
- Canina, Linda, Roni Michaely, Richard Thaler, and Kent Womack, 1998. Caveat compounder: A warning about using the daily CRSP equal-weighted index to compute long-run excess returns, *Journal of Finance* 53, 403–416.
- Carhart, Mark M, 1997. On persistence in mutual fund performance. *The Journal of Finance* 52, no. 1, 57-82.
- Chan, Louis K., Chen, Hsiu-Lang, and Lakonishok, Joseph, 2002. On mutual fund investment styles. *The Review of Financial Studies* 15, pp.1407-1437.

- Choi, Jaweon, Kronlund, Mathias and Oh, Ji Y.J., 2018. Sitting Bucks: Zero Returns in Fixed Income Funds, Working Paper.
- Cici, Gjergji, Gibson, Scott and Merrick Jr, John J., 2011. Missing the marks? Dispersion in corporate bond valuations across mutual funds. *Journal of Financial Economics*, 101(1), pp.206-226
- Cohen, Randolph B., Coval, Joshua D. and Pástor, Lubos, 2005. Judging fund managers by the company they keep. *The Journal of Finance*, 60(3), pp.1057-1096. Vancouver
- Cooper, Michael, Gulen, Huseyin, Rau, Raghavendra, 2005. Changing names with style: Mutual fund name changes and their effects on fund flows, *Journal of Finance* 60, 2825–2858.
- Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997. Measuring mutual fund performance with characteristics-based benchmarks, *Journal of Finance* 52, 1035–1058.
- Del Guercio, Diane, and Paula A. Tkac, 2008. Star power: The effect of Morningstar ratings on mutual fund flow, *Journal of Financial and Quantitative Analysis* 43, 907-936.
- Di Bartolomeo, Dan, and Erik Witkowski, 1997. Mutual fund misclassification: Evidence based on style analysis, *Financial Analysts Journal* 53, no. 5, 32-43.
- Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake, 2001. A first look at the accuracy of the CRSP Mutual Fund Database and a comparison of the CRSP and Morningstar Mutual Fund Databases, *Journal of Finance* 56, 2415–2430.
- Evans, Richard B., and Yang Sun, 2018. Models or stars: The role of asset pricing models and heuristics in investor risk adjustment, Working paper, University of Virginia.
- Frijns, Bart, Aaron B. Gilbert, and Remco CJ Zwinkels, 2013. On the Style-based Feedback Trading of Mutual Fund Managers. Available at SSRN 2114094.
- Froot, Kenneth, Dabora, Emil, 1999. How are stock prices affected by the location of trade? *Journal of Financial Economics* 53, 189–216.
- Gil-Bazo, Javier and Pablo Ruiz-Verdu, 2009. The relation between price and performance in the mutual fund industry, *The Journal of Finance* 64, no. 5, 2153-2183.
- Hartzmark, Samuel M., and Abigail Sussman, 2018. Do investors value sustainability? A natural experiment examining ranking and fund flows, Working paper, University of Chicago.
- Huang, Jennifer, Clemens Sialm, and Hanjiang Zhang, 2011. Risk shifting and mutual fund performance. *Review of Financial Studies* 24, no. 8, 2575-2616.

- Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng, 2008. Unobserved actions of mutual funds, *Review of Financial Studies* 21, no. 6, 2379-2416.
- Kaniel, Ron, and Robert Parham, 2017. WSJ category kings: The impact of media attention on consumer and mutual fund investment decisions, *Journal of Financial Economics* 123, 337-356.
- Kruger, P., Landier, A., and Thesmar, D., 2012. Categorization bias in the stock market. Unpublished working paper. University of Geneva, Toulouse School of Economics, and HEC.
- Kosowski, Robert, Timmermann, Allan, Wermers, Russ and White, Hal, 2006. Can mutual fund “stars” really pick stocks? New evidence from a bootstrap analysis. *The Journal of Finance*, 61(6), pp.2551-2595.
- Ljungqvist, Alexander, Malloy, Christopher and Marston, F., 2009. Rewriting history. *The Journal of Finance*, 64(4), pp.1935-1960.
- Nanda, Vikram, Wang, Z. Jay, and Zheng Lu., 2004. Family values and star phenomenon: strategies for mutual fund families, 17(3), pp.667-698.
- Reuter, Jonathan, and Eric Zitzewitz, 2015. How much does size erode mutual fund performance? A regression discontinuity approach, Working paper, Boston College.
- Rosenberg, Barr, and Michel Houglet, 1974. Error rates in CRSP and Compustat data bases and their implications, *Journal of Finance* 29, 1303–1310.
- Sensoy, B.A., 2009. Performance evaluation and self-designated benchmark indexes in the mutual fund industry. *Journal of Financial Economics*, 92(1), pp.25-39.
- Shumway, Tyler, 1997. The delisting bias in CRSP data, *Journal of Finance* 52, 327–340.
- Shumway, Tyler, and Vincent A. Warther, 1999. The delisting bias in CRSP’s NASDAQ data and its implications for interpretation of the size effect, *Journal of Finance* 54, 2361–2379.
- Swinkels, Laurens, and Liam Tjong-A-Tjoe, 2007. Can mutual funds time investment styles?, *Journal of Asset Management* 8, no. 2, 123-132.
- Vijh, Anand, 1994. S&P 500 trading strategies and stock betas, *Review of Financial Studies* 7, 215–251.
- Wermers, Russ, 2012. Matter of style: The causes and consequences of style drift in institutional portfolios. Available at SSRN 2024259.

Appendix A: Variable Definitions

Variable Name	Definition	Data Source
Reported AAA %	% of holdings in AAA assets as reported by a fund	Morningstar Direct
Calculated AAA %	% of holdings in AAA assets as calculated by Morningstar	Morningstar Direct
AAA %	% of holdings in AAA assets as calculated by us from the portfolio	Constructed
...
Morningstar Reported Risk Style	A fixed fund could be categorized as any of the following: “High Limited”, “Medium Limited”, “Low Limited”, “High Moderate”, “Medium Moderate”, “Low Moderate”, “High Extensive”, “Medium Extensive”, and “Low Extensive”	Morningstar Direct
Correct Fund Risk Style	We re-evaluate a fund as either high, medium, or low credit quality using their holdings. This counterfactual fund style is the fund style as indicated by the re-evaluated credit quality.	Constructed
Misclassified Dummy	Dummy variable that indicates whether a fund is misclassified in their fund credit quality dimension. It is 1 if the official credit style (High or Medium) is higher than the counterfactual credit quality as indicated by holdings, and 0 otherwise.	Constructed
Morningstar Category	Morningstar’s categorization of mutual funds based on the types of assets owned. See Appendix D for the various categories included in the sample.	Morningstar Direct
Reported Credit Score	Default scores calculated using the reported % asset holdings using Morningstar’s credit risk model formula (Appendix B)	Constructed
Reported Duration	The reported effective duration of a portfolio	Morningstar Direct
Reported Yield	The reported yield to maturity of a portfolio (in % points)	Morningstar Direct
Calculated Yield	Morningstar calculated average yield to maturity of a portfolio (in % points)	Morningstar Direct
12-Month Yield	The total coupon and dividend payment from the past 12 months (in % points)	Morningstar Direct
Fund Return	The fund return is the value weighted average of the share class returns. Share class returns come from Morningstar Direct	Constructed
Morningstar Rating 3-yr	The fund level Morningstar Rating is the value weighted average of share level Morningstar Ratings	Constructed

Morningstar Rating General	The fund level Morningstar Rating is the value weighted average of share level Morningstar Ratings	Constructed
Average Expense	Average expense at the fund level is calculated by taking the value weighted average of the share-class level expense ratios	Constructed
Monthly Flow	Monthly fund level investor flows	Morningstar Direct
Flow	Quarterly fund level investor flow is the quarterly sum of monthly flow	Constructed
JNK	Monthly returns of SPDR Bloomberg Barclays High Yield Bond ETF	Bloomberg

Appendix B: Credit Risk of a Fund Portfolio

Morningstar defines a bond portfolio's average credit risk using a weighted average score using the credit rating of the underlying assets. Prior to August 2010, a bond asset's score is defined using the following table:

Bond Quality	AAA	AA	A	BBB	BB	B	Below B	Not Rated	Not Rated Muni
Score	2	3	4	5	6	7	8	7	6

The portfolio's average position size weighted score then defines its credit quality using the following breakpoints.

Portfolio Average Score	0 to 2.5	2.5 to 3.5	3.5 to 4.5	4.5 to 5.5	5.5 to 6.5	6.5 to 7.5	>7.5
Quality Rating	AAA	AA	A	BBB	BB	B	Below B
Fund Style Quality	High	High	Medium	Medium	Low	Low	Low

After August 2010, the scores are based on a relative default rate:

Bond Quality	AAA	AA	A	BBB	BB	B	Below B	Not Rated	Not Rated Muni
Score	0	0.56	2.22	5.00	17.78	49.44	100.00	49.44	17.78

The respective breakpoints for post August 2010 are then:

Portfolio Average Score	0 to 0.13889	0.13889 to 1.25000	1.25000 to 3.47223	3.47223 to 9.02778	9.02778 to 31.25000	31.25000 to 72.36112	≥ 72.36112
Quality Rating	AAA	AA	A	BBB	BB	B	Below B
Fund Style Quality	High	High	Medium	Medium	Low	Low	Low

Appendix C: Credit Rating Composites

The following table lists the assumption used for our analysis and the subsequent % misclassified bond funds in the High and Medium credit quality categories from Q1 2017 to Q2 2018 in our final sample.

Assumptions in Classifying Bond Funds	% Fund-Quarter Obs. Misclassified Between Q1 2017 and Q2 2018
1) Use Morningstar's Calculated % Composition for each bond fund portfolio.	31.4%
1) Calculate % composition directly using the Bloomberg/Barclays Method for all Bond Broad Type assets holdings.	31.0%
Baseline Analysis	24.1%
1) Ignore Bond Subtypes that are Warrants/Rights (Calls), Warrants/Rights (Puts), Futures, Option (Calls), Options (Puts), Interest Rate Swaps, Interest Rate Futures, Interest Rate Forwards, Credit Default Swaps, and Treasury Futures. 2) Calculate % composition directly using the Bloomberg/Barclays Method for all other Bond Broad Type assets holdings.	
1) Ignore Bond Subtypes that are Warrants/Rights (Calls), Warrants/Rights (Puts), Futures, Option (Calls), Options (Puts), Interest Rate Swaps, Interest Rate Futures, Interest Rate Forwards, Credit Default Swaps, and Treasury Futures. 2) Calculated % composition directly using the best rating from the Fitch, Moody's, and S&P for all other Bond Broad Types assets holdings.	20.6%
1) Ignore Bond Subtypes that are Warrants/Rights (Calls), Warrants/Rights (Puts), Futures, Option (Calls), Options (Puts), Interest Rate Swaps, Interest Rate Futures, Interest Rate Forwards, Credit Default Swaps, and Treasury Futures. 2) Calculated % composition using the best rating from the 8 US NRSROs (Fitch, Moody's, S&P, AMBest, DBRS, Egan-Jones, Kroll, and Morningstar) for all other Bond Broad Types assets holdings.	14.4%

Appendix D: Morningstar Categories

This table describes the fraction of fund observations in each Morningstar Category that are misclassified according to our baseline classification system between Q1 2017 to Q2 2018.

Morningstar Category	Fraction of Bond-Fund Observations Misclassified
US Fund Ultrashort Bond	41.89%
US Fund Inflation-Protected Bond	29.7%
US Fund Short-Term Bond	26.7%
US Fund Intermediate Core-Plus Bond	22.1%
US Fund Intermediate Core Bond	18.8%
US Fund Corporate Bond	8.90%

Appendix E:
[Link to Morningstar's Response](#)

Appendix F:
[Link to Our Reply](#)

Appendix G:
[Link to Morningstar's Second Response Comments](#)

Appendix H
[Link to Our Reply to these Comments](#)

<u>Principal Amount</u>		<u>Fair Value</u>
Corporate Bonds, continued		
Health Care Providers & Services, continued		
\$ 1,320,000	UnitedHealth Group, Inc., 4.75%, 7/15/45	\$ 1,412,659
220,000	WellPoint, Inc., 3.50%, 8/15/24, Callable 5/15/24 @ 100	215,543
		<u>13,229,679</u>
Hotels, Restaurants & Leisure (0.1%):		
10,000	McDonald's Corp., 2.75%, 12/9/20, Callable 11/9/20 @ 100	9,927
50,000	McDonald's Corp., 4.70%, 12/9/35, Callable 6/9/35 @ 100	51,974
310,000	McDonald's Corp., 3.70%, 2/15/42, MTN	271,124
120,000	McDonald's Corp., 4.60%, 5/26/45, Callable 11/26/44 @ 100	120,542
780,000	McDonald's Corp., 4.88%, 12/9/45, Callable 6/9/45 @ 100, MTN	815,713
800,000	McDonald's Corp., 4.45%, 3/1/47, Callable 9/1/46 @ 100, MTN^	785,569
160,000	McDonald's Corp., 4.45%, 9/1/48, Callable 3/1/48 @ 100, MTN^	156,882
315,000	Starbucks Corp., 3.80%, 8/15/25, Callable 6/15/25 @ 100	312,919
215,000	Starbucks Corp., 3.75%, 12/1/47, Callable 6/1/47 @ 100	187,160
120,000	Starbucks Corp., 4.50%, 11/15/48, Callable 5/15/48 @ 100	118,018
		<u>2,829,828</u>
Household Products (0.0%):		
740,000	Clorox Co. (The), 3.90%, 5/15/28, Callable 2/15/28 @ 100^	738,344
Industrial Conglomerates (0.3%):		
2,565,000	General Electric Capital Corp., Series A, 5.55%, 5/4/20, MTN	2,657,323
2,120,000	General Electric Co., Series D, 5.00%(US0003M+333bps), 12/31/49, Callable 1/21/21 @ 100^	2,065,410
164,000	Georgia-Pacific LLC, 5.40%, 11/1/20(a)	170,635
505,000	Georgia-Pacific LLC, 3.73%, 7/15/23, Callable 4/15/23 @ 100(a)	507,482
309,000	Georgia-Pacific LLC, 3.60%, 3/1/25, Callable 12/1/24 @ 100(a)	305,973
		<u>5,706,823</u>

Figure 1.

This figure contains an excerpt from the AZL Enhanced Bond Index Fund's September 30, 2018 N-Q Schedule of Investments held.

SURVEY AS OF DATE	UNIQUE IDENTIFIER	FUND NAME	CREDIT RATINGS RISK								
			AAA	AA	A	BBB	BB	B	BELOW B	NOT RATED	
(mm/dd/yyyy) or (yyyy-mm-dd) Must be a month-end date. All other dates will be rounded (rounded forward where dd =>16). Do not include formulas.	Char(75) This MUST be the same identifier used for submitting portfolio holdings. Please contact us if you do not have this information.	Char(75) Must be kept consistent per fund, per fund file. Can vary from delivery to delivery due to name changes.	Numeric(5.2) % of assets for this rating assigned by a Nationally Recognized Statistical Rating Organizations.	Numeric(5.2) % of assets for this rating assigned by a Nationally Recognized Statistical Rating Organizations.	Numeric(5.2) % of assets for this rating assigned by a Nationally Recognized Statistical Rating Organizations.	Numeric(5.2) % of assets for this rating assigned by a Nationally Recognized Statistical Rating Organizations.	Numeric(5.2) % of assets for this rating assigned by a Nationally Recognized Statistical Rating Organizations.	Numeric(5.2) % of assets for this rating assigned by a Nationally Recognized Statistical Rating Organizations.	Numeric(5.2) % of assets for this rating assigned by a Nationally Recognized Statistical Rating Organizations.	Numeric(5.2) % of assets for this rating assigned by a Nationally Recognized Statistical Rating Organizations.	Numeric(5.2) % of assets that do not have a rating.

Figure 2.

This figure contains a portion of the fixed income template sent by Morningstar to survey mutual funds in August 2019.

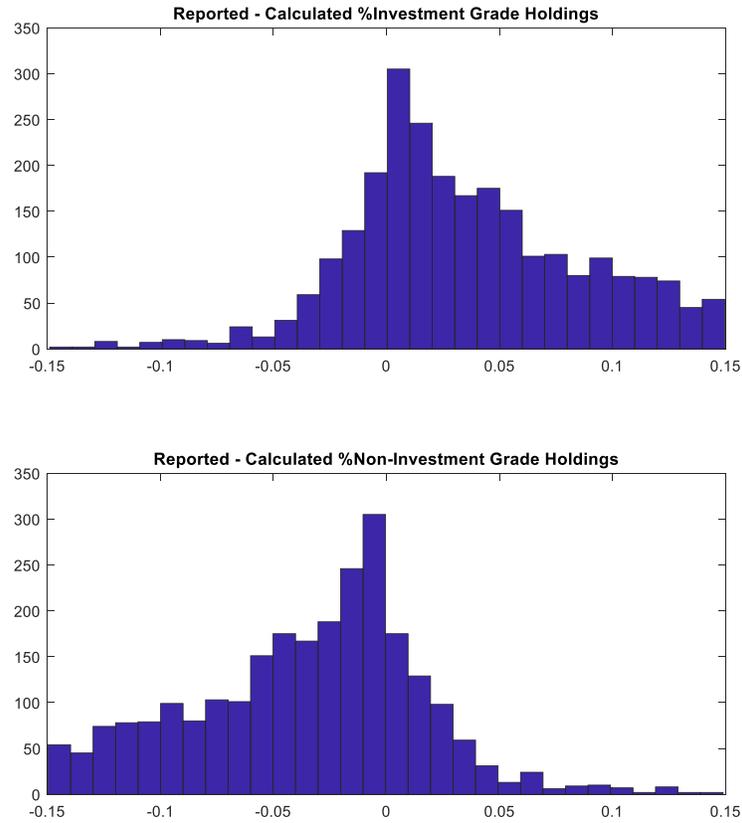


Figure 3.

This graph plots the histograms of fund reported % holdings minus the calculated % holdings in the various bond credit rating categories. The sample period begins in Q1 2017, when Morningstar began calculating % holdings of assets in each credit risk category per each fixed income fund, and ends in Q2 2018. Observations where fund reported % is exactly the same as the calculated % holdings are removed to aid readability.

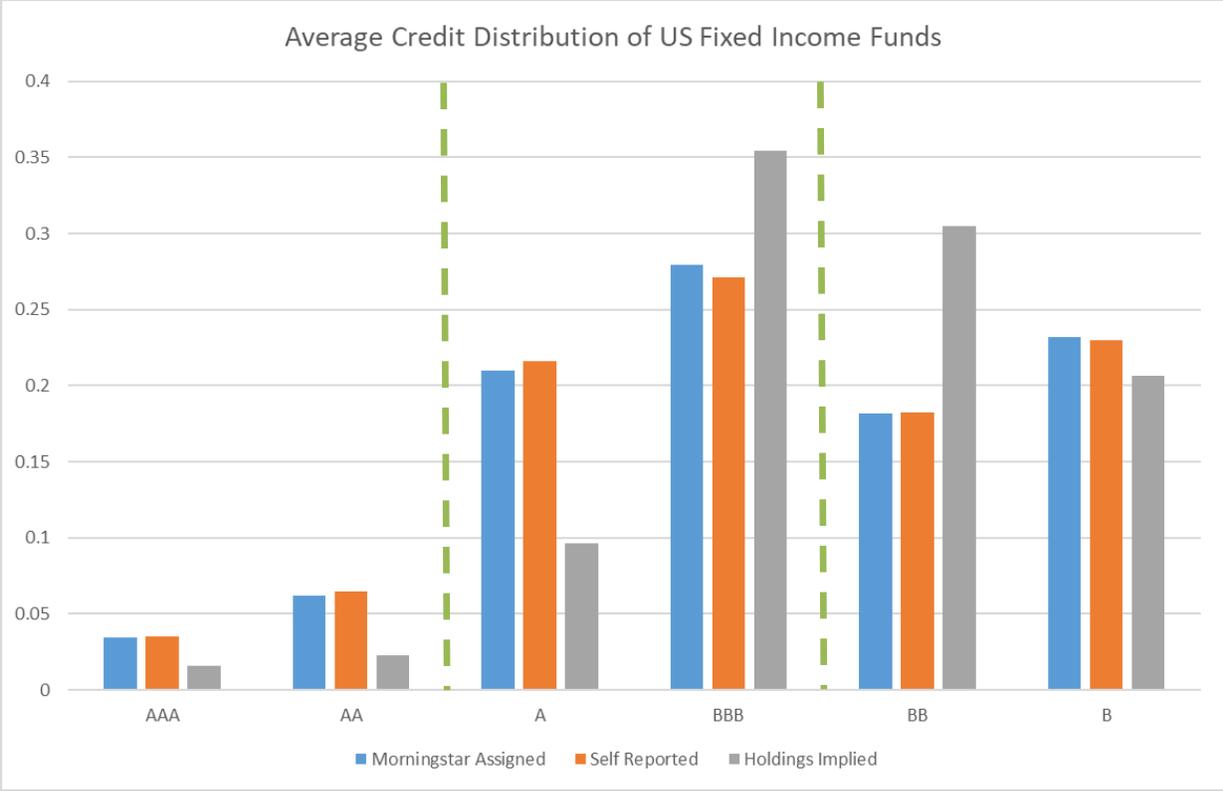


Figure 4.

This figure plots the credit risk distribution of fund-quarter observations between Q1 2017 and Q2 2018. The blue is the distribution of the official average credit quality category that Morningstar assigns to US Fixed Income funds. According to MS’s methodology, this official credit quality category is calculated using fund survey reported % holdings of assets in the various credit risk categories. In red, we replicate the official credit quality category using the fund survey-reported % holdings. The grey is the counter-factual credit risk category that would result if we had used MS calculated % holdings. The dashed lines represent breaks in the fixed income fund style-box. AAA and AA credit quality funds are high credit quality; A and BBB credit quality funds are medium credit quality; and BB and B are low credit quality as deemed by Morningstar.

Table 1.
Description of Data

We obtain credit ratings from three sources. Dodd-Frank requires all credit rating agencies to release their rating data history through XBRL filings with a one year delay. Capital IQ subscription contains the S&P rating history. Mergent FISD contains corporates, supranational, and agency/treasuries debts. Portfolio history is directly from Morningstar's collection of filings and surveys for each fund. The surveyed holdings % on individual fixed income funds comes from the Morningstar Direct database from Q1 2003 to Q2 2018.

Panel A. Sources of Credit Ratings:

Dates	Source	Coverage Description
Jun 2012 to Jun 2018	XBRL Filing	All NRSROs Rated Bonds
Jan 2003 to Jun 2018	Capital IQ	S&P Rating History
Jan 2003 to Jun 2018	Mergent FISD	S&P, Moody's, Fitch Ratings for Corporations and Treasuries

Panel B. Actual Holdings of US Fixed Income Funds from Q1 2003 to Q2 2018

	10th P	Median	90th P	Mean	Std.	N
AAA	0.00%	40.8%	81.4%	39.0%	31.2%	18,508
AA	0.00%	2.48%	9.15%	3.73%	4.92%	18,508
A	0.00%	7.97%	22.7%	9.58%	9.94%	18,508
BBB	0.326%	12.6%	35.8%	15.9%	15.9%	18,508
BB	0.00%	3.88%	28.2%	9.10%	11.6%	18,508
B	0.00%	1.52%	44.8%	11.4%	18.3%	18,508
Below B	0.00%	0.537%	18.1%	4.71%	8.08%	18,508
Unrated	0.0743%	4.12%	15.7%	6.50%	7.42%	18,508

Panel C. Surveyed Holdings of US Fixed Income Funds from Q1 2003 to Q2 2018

	10th P	Median	90th P	Mean	Std.	N
AAA	0.00%	41.1%	83.9%	40.1%	31.5%	18,508
AA	0.00%	3.56%	12.8%	5.51%	7.97%	18,508
A	0.00%	9.34%	25.6%	10.9%	10.7%	18,508
BBB	0.50%	12.5%	34.6%	15.7%	15.1%	18,508
BB	0.00%	4.20%	32.0%	10.3%	13.3%	18,508
B	0.00%	1.70%	46.0%	11.8%	18.6%	18,508
Below B	0.00%	0.39%	14.6%	3.99%	7.16%	18,508
Unrated	0.00%	0.32%	5.26%	1.67%	3.61%	18,508

Table 2.
Time Series

In this table, we report the time series of Fund-Quarter observations in each Morningstar Credit Quality Category. The last column is the number of funds that are misclassified into the high or med credit quality category. MS changed the way it calculated average credit quality in August 2010. Prior to August 2010, the average credit quality is a simple weighted average of the underlying linear bond scores, in which a AAA bond has a score of 2, AA has a score of 3, and so on. After August 2010, the credit risk variable attempts to describe a fund in terms of the returns and risks of a portfolio of rated bonds, and nonlinear scores are assigned to each category. The sample is from Q1 2003 to Q2 2018. We record the weighing scheme used after August 2010 in Appendix C.

Year	High Credit Quality	Med Credit Quality	Low Credit Quality	# Misclassified
2003	251	412	321	7
2004	262	396	337	4
2005	255	364	282	4
2006	315	414	332	5
2007	322	516	422	7
2008	359	610	468	8
2009	246	698	548	9
2010	209	705	583	147
2011	189	765	658	307
2012	194	857	708	283
2013	191	887	824	297
2014	178	920	891	348
2015	181	1,056	1,022	321
2016	209	1,195	1,024	360
2017	225	1,215	993	370
2018	123	581	484	191

Table 3.
Yields and Misclassification

In this table, we regress various yield metrics on misclassified dummy and control variables. Misclassified dummy is 1 if the official credit quality (High or Medium) is higher than the counterfactual credit quality, and 0 otherwise. Funds voluntarily report their portfolio yields (1) and (4) to Morningstar. Morningstar began calculating the holding yields (2) and (5) in 2017. The 12-month total interest, coupon, and dividend payments constitute the 12-month yield (3) and (6). The sample period is Q3 2010 to Q2 2018. t-statistics are double-clustered by time and fund.

	(1) Reported Yield _t	(2) Calculated Yield _t	(3) 12-Month Yield _{t+11}	(4) Reported Yield _t	(5) Calculated Yield _t	(6) 12-Month Yield _{t+11}
Misclassified _{t-1}	0.277*** (5.494)	0.237*** (5.372)	0.190*** (3.344)	0.0106 (0.157)	0.0130 (0.273)	-0.0735 (-1.106)
Reported Credit Score _{t-1}	0.112*** (8.394)	0.0569*** (6.188)	0.0551*** (4.744)	0.0727*** (7.861)	0.0486*** (9.229)	0.0552*** (6.755)
Reported Duration _{t-1}	0.127*** (4.263)	0.0229** (3.083)	0.107*** (3.272)	0.138*** (4.820)	0.0359** (3.116)	0.110*** (3.637)
Time x Morningstar Reported Risk Style FE	Yes	Yes	Yes	No	No	No
Time x Correct Fund Risk Style FE	No	No	No	Yes	Yes	Yes
Time x Morningstar Category FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,402	1,303	7,127	7,957	1,542	8,800
Adjusted R-squared	0.673	0.816	0.587	0.736	0.873	0.607

Table 4.
Counterfactuals and Misclassification

In this table, we regress monthly fund returns on misclassified dummy and control variables. Misclassified dummy is 1 if the official credit quality (High or Medium) is higher than the counterfactual credit quality, and 0 otherwise. The sample period is Q3 2010 to Q2 2018. t-statistics are clustered quarterly.

	(1)	(2)	(3)	(4)
	Fund Return _t	Fund Return _t	Fund Return _t	Fund Return _t
Misclassified _{t-1}	3.579*** (2.951)	3.038*** (3.472)	-2.341** (-2.003)	-0.558 (-0.646)
Reported Credit Score _{t-1}		0.411** (2.419)		0.611** (2.259)
Reported Duration _{t-1}		1.522 (1.065)		1.468 (1.012)
Average Expense _{t-1}		-3.551*** (-3.393)		-3.392*** (-3.774)
Time x Morningstar Reported Risk Style FE	Yes	Yes	No	No
Time x Correct Fund Risk Style FE	No	No	Yes	Yes
Time x Morningstar Category FE	Yes	Yes	Yes	Yes
Observations	25,318	22,671	31,196	27,941
Adjusted R-squared	0.874	0.874	0.841	0.844

Table 5.
Morningstar Star Ratings and Misclassification

In this table, we regress Morningstar ratings on the misclassified dummy and controls. Since the ratings and expenses are reported at the share class level, the fund level Morningstar Ratings and the Average Expense ratio are calculated as the value weighted average of their respective share-class level values. The sample period is Q3 2010 to Q2 2018. t-statistics are double-clustered by time and fund.

	(1) Morningstar Rating 3 Yr _t	(2) Morningstar Rating 3 Yr _t	(3) Morningstar Rating Overall _t	(4) Morningstar Rating Overall _t
Misclassified _{t-1}	0.383*** (5.971)	0.170*** (3.774)	0.341*** (4.660)	0.182*** (3.218)
Reported Credit Score _{t-1}	0.0698*** (4.355)	0.0299** (2.553)	0.0588*** (3.090)	0.0289* (1.774)
Reported Duration _{t-1}	0.107*** (3.679)	-0.0277 (-1.138)	0.113*** (2.752)	0.0122 (0.386)
Average Expenses _{t-1}	-1.024*** (-6.915)	-0.755*** (-6.966)	-0.822*** (-5.045)	-0.622*** (-4.566)
3 Year Returns _{t-1}		15.22*** (8.036)		11.36*** (6.202)
Time x Morningstar Reported Risk Style FE	Yes	Yes	Yes	Yes
Time x Morningstar Category FE	Yes	Yes	Yes	Yes
Observations	7,391	7,391	7,391	7,391
Adjusted R-squared	0.211	0.541	0.170	0.373

Table 6.
Expense Ratios and Misclassification

In this table, we analyze whether misclassified funds are more expensive than usual. We regress average expense ratio on misclassified dummy and control variables. The average expense ratio is calculated at the fund level as the value weighted average of their respective share-class level values. The sample period is Q3 2010 to Q2 2018. t-statistics are double-clustered by time and fund.

	(1) Average Expense _t	(2) Average Expense _t	(3) Average Expense _t
Misclassified _{t-1}	0.114*** (6.356)	0.0765*** (4.186)	0.0760*** (4.172)
Reported Credit Score _{t-1}		0.0224*** (3.611)	0.0222*** (3.592)
Reported Duration _{t-1}			-0.00790 (-0.754)
Time x Morningstar Reported Risk Style FE	Yes	Yes	Yes
Time x Morningstar Category FE	Yes	Yes	Yes
Observations	8,373	7,586	7,586
Adjusted R-squared	0.125	0.153	0.154

Table 7.
Fund Flows and Misclassification

In this table, we regress the direction of investor flows into funds on their concurrent Morningstar Rating, as instrumented on lagged fund misclassifications. There are two specifications. The first column regresses flow indicator on misclassified dummy directly. The second column regresses the flow indicator on misclassified stars. The sample period is Q3 2010 to Q2 2018. t-statistics are clustered quarterly.

	(1) Flow _t >0	(2) Flow _t >0
Misclassified _{t-1}	0.0637*** (4.947)	
Misclassified Stars _t		0.171*** (5.155)
Reported Credit Score _{t-1}	0.00438 (1.198)	-0.00422 (-0.757)
Reported Duration _{t-1}	0.0191*** (3.998)	0.00201 (0.261)
Average Expenses _{t-1}	-0.238*** (-7.431)	-0.0685 (-1.409)
Time x Morningstar Reported Risk Style FE	Yes	Yes
Time x Morningstar Category FE	Yes	Yes
Observations	7,766	7,391
Adjusted R-squared	0.068	0.086

Table 8.
Characteristics of Misclassified Funds

In this table, we regress indicators of when a share class start and stop misclassifying their holdings on past returns and other characteristics. The sample period covers Q3 2010 to Q2 2018, when Morningstar used its most recent classification scheme. In column 1, *Starts* is an indicator representing when a previously correctly classified fund starts misclassifying. In column 2, *Stops* is an indicator for when a previously misclassified fund starts correctly classifying. In column 3, we regress *Starts* minus *Stops*. t-statistics are clustered quarterly.

	(1) Starts	(2) Stops	(3) Starts-Stops
3 Year Return	-0.125*** (-2.743)	0.180** (2.107)	-0.305*** (-3.151)
Time x Morningstar Reported Risk Style FE	Yes	Yes	Yes
Time x Morningstar Category FE	Yes	Yes	Yes
Observations	8,389	8,389	8,389
Adjusted R-squared	0.004	0.025	0.017

Table 9.
The Characteristics of Unrated Bonds held by Funds

This table summarizes the corporate bonds in the Mergent FISD database that were issued between 2010 and 2016. Each box describes the mean offering yield and the number of bonds in different ranges of offering maturities and credit qualities. A bond's credit rating at issuance is the Barclays/Bloomberg composite of Fitch, Moody's, and S&P's respective ratings that were available within 30 trading days of the offering date. N is the number of issue observations in each box.

	Issuing Maturity		
	0 to 3.5 Years	3.5 to 6 Years	6 to 10 Years
High Investment Grade (AA to AAA)	1.44% N=113	2.21% N=146	2.70% N=33
Medium Investment Grade (BBB to A)	1.75% N=483	2.80% N=1,110	3.75% N=370
Junk Grade Bonds (BB and Below)	5.14% N=43	8.08% N=563	7.69% N=1,655
Unrated Bonds	7.81% N=81	6.43% N=245	7.09% N=356

Table 10.
Misclassified Fund Performance around Junk Bond Crashes (and Outperformance)

In this table, we regress monthly fund returns on misclassified dummy and control variables. Misclassified dummy is 1 if the official credit quality (High or Medium) is higher than the counterfactual credit quality, and 0 otherwise. In the columns, we regress separately the sample months when JNK, the SPDR Bloomberg Barclays High Yield Bond ETF, had major negative returns (1), close to 0 returns (2), and substantial positive returns (3). The sample period is Q3 2010 to Q2 2018. t-statistics are clustered quarterly.

	(1) Fund Return _t JNK < -1%	(2) Fund Return _t -1% < JNK < 1%	(3) Fund Return _t JNK > 1%
Misclassified _{t-1}	-4.672*** (-3.117)	2.849** (2.270)	7.630*** (6.141)
Reported Credit Score _{t-1}	-1.421** (-2.270)	0.263 (0.828)	2.024*** (5.232)
Reported Duration _{t-1}	-4.784 (-1.246)	0.378 (0.187)	5.227** (2.327)
Average Expense _{t-1}	-9.683*** (-3.457)	-2.234* (-2.011)	-1.146 (-0.945)
Time x Morningstar Reported Risk Style FE	Yes	Yes	Yes
Time x Morningstar Category FE	Yes	Yes	Yes
Observations	4,522	9,972	8,177
Adjusted R-squared	0.855	0.879	0.820