Maturity-matched bond fund performance

Markus Natter, Martin Rohleder, Marco Wilkens

Angaben zur Veröffentlichung / Publication details:

Natter, Markus, Martin Rohleder, and Marco Wilkens. 2021. "Maturity-matched bond fund performance." Financial Analysts Journal 77 (2): 83–96. https://doi.org/10.1080/0015198X.2020.1865695.

Maturity-Matched Bond Fund Performance

Markus Natter, Martin Rohleder, and Marco Wilkens

Markus Natter is a portfolio manager for Provinzial Asset Management GmbH, Münster, Germany. Martin Rohleder is a habilitated doctor at the University of Augsburg Faculty of Business Administration and Economics, Augsburg, Germany. Marco Wilkens is a full professor at the University of Augsburg Faculty of Business Administration and Economics, Augsburg, Germany.

Performance regressions lever expected benchmark returns linearly to the risk exposures of the fund. The interest rate (IR) risk premium, however, usually follows a decreasingly upward-sloping yield curve, characterizing the nonlinearity between expected return and IR risk exposure—for example, maturity or duration. If the exposures of the fund and the benchmark differ, this discrepancy causes alpha to deviate from the active bond selection performance it is supposed to measure. Performance ratings and investor flows are affected by this *alpha deviation*. Our simple remedy is to individually match funds and benchmarks using their durations. Beta and R^2 are candidates for alternative matchings.

Disclosure: The authors report no conflicts of interest.

Fund returns are affected by three different kinds of investment decisions. Decisions based on predicting securities' idiosyncratic deviations from expected returns generate active security selection performance. The decision to harvest systematic risk premiums by choosing long-term risk exposures generates passive style returns.[1] Decisions to temporarily deviate from long-term risk exposures generate active market-timing performance. Therefore, it is very important to measure especially the success of active and passive decisions separately and correctly to keep their return contributions clearly distinguishable.

We concentrate on active bond selection performance,[2] which is usually measured as the constant (α, or alpha) of a linear time-series regression (e.g., Equation 1) explaining the fund's excess returns (*er*) with those of a broadly diversified bond or Treasury market index (e.g., Cornell and Green 1991; Blake, Elton, and Gruber 1993; Elton, Gruber, and Blake 1995) plus the returns of additional risk and style factors:

$$er_{i,t} = \alpha_i + \beta_i \, er_{Broad,t} + \varepsilon_{i,t}. \tag{1}$$

The role of the broad Treasury index is to represent the interest rate (IR) risk premium, $E(er_{Broad,t})$.[3] The regression's slope coefficient (β, or beta) measures the fund's long-term exposure to IR risk. Consequently, the fund's model-implied expected return from harvesting the IR risk premium—that is, its passive style return—is $\beta_i E(er_{Broad,t})$.[4] For this

We thank Rainer Baule, Henk Berkman, Stephen Brown (the executive editor), Ilja Dergunov, Oliver Entrop, Iraj Fooladi, Raquel Gaspar, Daniel Giamouridis (co-editor), Alexander Hillert, Christian Koziol, David Maslar, Steffen Mayer, Fabio Moneta, Sebastian Müller, Andreas Neuhierl, Nial O'Sullivan, Andreas Rathgeber, Heidi Raubenheimer, CFA, Ryan Riordan, Hendrik Scholz, Jules van Binsbergen, Stijn van Nieuwerburgh, Andreas Walter, Florian Weigert, David Yermack, and the participants at the 2017 European Retail Investment Conference in Stuttgart, the 2017 Meeting of the German Association of Business Research at the University of St. Gallen, the 2017 FMA European Conference at ISEG Lisbon, the 2017 World Finance Conference at the University of Cagliari, the 2017 Meeting of the German Finance Association at the University of Ulm, and the 2019 FMA International Annual Meeting in New Orleans for very helpful comments and suggestions. We are responsible for all remaining errors.

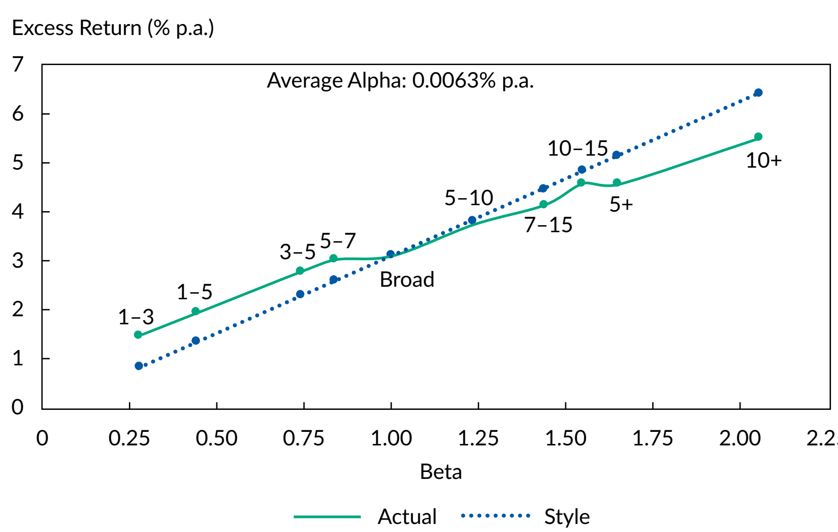
separation into active bond selection performance and passive style return to work, the most important assumption of the linear regression approach is that the relation of IR risk exposure and expected return is, in fact, linear. However, it is well known that the relation between IR risk exposure and expected return is nonlinear and is best explained by a decreasingly upward-sloping yield curve (e.g., Nelson and Siegel 1987; Litterman and Scheinkman 1991).

To demonstrate the problem this discrepancy between the model assumption and the real world behavior of expected returns entails for the correct measurement of active bond selection performance, we first analyzed passive US Treasury bond total return indexes with different maturity ranges.[5] By definition, passive Treasury indexes have zero active alpha and are free of default risk. IR risk exposure is their sole source of expected return, and any nonzero alpha must be the result of an incorrect model-implied style return. We regressed the excess return time series of nine different maturity-range-specific indexes on the time series of a broad index using Equation 1. Then, we compared their actual expected excess returns, $E(er_{i,t})$, with their expected style returns, $\beta_i E(er_{Broad,t})$. If the linearity assumption holds, the two expected returns should be identical; that is, all indexes should have zero alpha.

Figure 1 plots the various expected returns as functions of the indexes' betas.[6] Both expected excess returns increase with beta (maturity range), consistent with higher risk premiums for higher IR risk exposure. For all indexes with betas less than 1, however, the style returns are systematically lower than the actual returns.[7] This outcome leads to positive differences between actual and style returns—that is, positive alphas. For all indexes with betas greater than 1, the style returns are systematically higher than the actual returns, leading to negative alphas. The average alpha of the indexes is near zero (0.0063% p.a.), but individual index alphas are nonzero.[8]

For the performance measurement of active bond funds, these findings mean that conclusions about the average active alpha depend on the maturity or duration of the benchmark. More importantly, using an index that produces the correct average alpha is insufficient because it still creates incorrect alphas for those individual funds that do not have a duration close to that of the benchmark. This *alpha deviation* arises because a part of the passive average style return is transferred into active alpha by linearly levering the expected benchmark return along a nonlinear yield curve. The sign of the alpha deviation depends systematically on the IR risk exposure

Figure 1. Expected Returns of US Treasury Indexes



Notes: This figure plots the actual mean excess returns over the one-month T-bill (in percentage per annum) of various maturity-range-specific passive US Treasury bond total return indexes against their betas from linear regressions using Equation 1 and applying the broad index as the common benchmark. The figure plots style excess returns—that is, beta multiplied by the actual mean excess returns of the broad index. The maturity ranges of the indexes are provided above or below the expected excess returns (e.g., "1–3" indicates that the index holds Treasury bonds with maturities between one and three years). Average alpha is the equal-weighted average of the differences between actual and style excess returns over the 10 indexes. (Individual index alphas are shown in Table A2.)

being less than or greater than that of the common benchmark—that is, the maturity mismatch.

As a simple and practical remedy to the alpha deviation problem, we propose a new maturity-matched performance measure (MM alpha) that considers an individual benchmark Treasury index for each fund, matched by the funds' and the indexes' reported durations.[9] The expected beta of the fund toward the matched index is 1. Thereby, that benchmark index considers just the IR risk premium suitable for the particular duration of the fund, mitigating the alpha deviation in measured active bond selection performance.

The MM alpha combines the widely used and very flexible returns-based regression approach with the holdings-based idea of matching the durations of bonds and benchmarks (e.g., Cici and Gibson 2012; Moneta 2015; Palhares and Richardson 2020). It has the advantage that it requires much less data. A further advantage is that the nonlinearity between IR risk exposure and expected return does not need to be explicitly modeled (e.g., Nelson and Siegel 1987; Litterman and Scheinkman 1991). A possible disadvantage, however, is that not all bond funds report their durations, though this obstacle can be expected to diminish over time with funds' reporting becoming more transparent. Moreover, we tested several practical alternative criteria for matching funds and benchmarks to consider a suitable IR risk premium and found that beta works well for government bond funds and R^2 works well for corporate bond funds.

We applied the MM alpha to samples of government and corporate bond funds and found that the average deviation between the alpha using a broad Treasury index as the benchmark for all funds (broad alpha) and the MM alpha is positive for our sample period, meaning that the usual measure overestimates the average bond selection performance of the funds in our sample. On the single-fund level, many funds have a significantly positive alpha deviation and many others have a significantly negative alpha deviation, leading to a large average absolute alpha deviation. Moreover, we examined the practical implications of the alpha deviation in active bond selection performance and demonstrated that the popular Morningstar rating is more sensitive to the broad alpha than to the MM alpha, especially for corporate bond funds. A similar conclusion holds for investor flows, which, overall, are more sensitive to the broad alpha than to the MM alpha, especially for institutional government bond funds and retail and investment-grade corporate bond funds.

Fund Data and Summary Statistics

We obtained bond mutual fund data from CRSP and Morningstar. We selected active mutual funds reporting, on average, more than 50% portfolio weight in government or corporate bonds and excluded all funds with an investment objective other than US domestic government or corporate bonds. Monthly returns net of expenses (net returns),[10] monthly total net assets (TNA), turnover, total expenses, age, and institutional and retail classification are from CRSP. We unsmoothed monthly net returns with a three-step procedure following Couts, Gonçalves, and Rossi (2020).[11] Monthly effective durations and style and objective categories are from Morningstar. We aggregated share classes at the fund level and excluded all funds with either incomplete information or less than one year of monthly returns after surpassing TNA of US$5 million (following Fama and French 2010). The final samples contain 127 (291) active US domestic government (corporate) bond funds for the period from 1990 to 2014.

Table 1 reports summary statistics on fund characteristics separately for government bond funds (Panel A) and corporate bond funds (Panel B). Mean net excess returns of government (corporate) bond funds are 2.08% p.a. (3.68% p.a.), which is greater than (less than) the average excess return of the broad Treasury index in Table A2 in Appendix A (3.13% p.a.). The average effective duration of government (corporate) bond funds is 4.12 (4.09) years, which is 1.3 years shorter than the average duration of the broad index (5.44 years) reported in Table A2. Other average fund characteristics are in line with previous research (e.g., Huij and Derwall 2008).

Performance Model

To consider IR risk, we used a linear multifactor regression model:

$$er_{i,t} = \alpha_i + \beta_{i,Treasury}\, er_{Treasury,t} + \beta_{i,Def}\mathrm{Def}_t + \beta_{i,Option}\,\mathrm{Option}_t + \beta_{i,mkt}\, er_{mkt,t} + \varepsilon_{i,t}, \quad (2)$$

where $er_{i,t}$ is the excess return of fund i at time t; $er_{Treasury,t}$ is the excess return of a Treasury index; Def_t is a zero-investment default risk factor, constructed as the return of a high-yield index minus the return of a Treasury index with a matching duration to minimize the factor's IR risk exposure; and Option_t is a zero-investment prepayment risk factor, constructed as the return of a mortgage-backed

Table 1. Summary Statistics

	Mean	Std. Dev.	5th Percentile	Median	95th Percentile
A. Government bond funds (127 funds)					
Net excess return (% p.a.)	2.08	16.74	–22.05	1.81	25.26
Effective duration (years)	4.12	2.70	1.00	4.07	7.68
Total net assets (US$ m)	598.32	2,028.76	9.50	127.90	2,156.50
Age (years)	8.88	5.91	1.08	7.83	20.17
Total expense ratio (% p.a.)	0.70	0.37	0.26	0.68	1.13
Turnover ratio (% p.a.)	153.80	235.95	13.00	90.00	480.00
Government bond holdings (%)	71.63	16.10	52.03	67.89	99.37
Corporate bond holdings (%)	15.81	12.24	0.77	11.67	37.39
Cash holdings (%)	2.38	8.09	–6.07	2.86	12.48
Institutional fund (%)	26.14	43.94	0.00	0.00	100.00
Retail fund (%)	40.35	49.06	0.00	0.00	100.00
High-yield fund (%)	0.00	0.00	0.00	0.00	0.00
B. Corporate bond funds (291 funds)					
Net excess return (% p.a.)	3.68	31.04	–41.02	4.19	45.36
Effective duration (years)	4.09	1.83	1.40	4.14	6.77
Total net assets (US$ m)	560.98	1,591.04	10.20	133.50	2,317.70
Age (years)	8.30	5.86	0.83	7.17	19.42
Total expense ratio (% p.a.)	0.92	0.38	0.34	0.89	1.61
Turnover ratio (% p.a.)	111.65	135.92	19.00	72.00	353.00
Government bond holdings (%)	10.85	12.05	0.00	5.50	33.40
Corporate bond holdings (%)	73.83	14.74	51.26	74.53	94.41
Cash holdings (%)	5.30	6.10	0.33	4.22	16.30
Institutional fund (%)	28.64	45.21	0.00	0.00	100.00
Retail fund (%)	50.11	50.00	0.00	100.00	100.00
High-yield fund (%)	38.41	48.64	0.00	0.00	100.00

Notes: This table shows pooled summary statistics for active US domestic government bond funds (Panel A) and active US domestic corporate bond funds (Panel B) in the sample period from January 1990 to December 2014. "Net excess return" means after deduction of the fund's total expense ratio and in excess of the risk-free rate of return (one-month T-bill).

security index minus the return of a Treasury index with a matching duration. Finally, we included the excess return of a broadly diversified stock index to capture equity-related risk. The model is based on the "index-4" model of Elton et al. (1995), which has been used regularly in previous bond fund and bond performance studies (e.g., Huij and Derwall 2008; Bessembinder, Kahle, Maxwell, and Xu 2009; Amihud and Goyenko 2013).[12,13]

In the course of the empirical analysis, we used several different Treasury indexes to demonstrate the alpha deviation problem and as a reference for our proposed MM alpha. For the MM alpha, we individually matched funds and Treasury indexes by the minimum difference between the average duration of the fund and the average duration of the respective index during the fund's life span within our sample period.

Bond Fund Performance

Panel A of **Table 2** shows equal-weighted mean net alpha estimates summarized over all government

Table 2. Fund Performance

Treasury Index	Alpha	*t*-Statistic	Beta	R^2
A. Government bond funds				
1–3	–1.9355**	–13.47	2.440	0.712
1–5	–1.7838**	–12.43	1.737	0.760
3–5	–1.4983**	–10.69	1.088	0.773
5–7	–1.4635**	–10.61	0.991	0.778
Broad	–1.1565**	–8.46	0.883	0.763
5–10	–1.0149**	–7.98	0.699	0.771
7–15	–0.8356**	–6.81	0.599	0.751
10–15	–0.7973**	–6.64	0.541	0.721
5+	–0.7374**	–5.93	0.511	0.720
10+	–0.4280**	–3.86	0.367	0.643
Maturity matched	–1.3873**	–10.92	1.061	0.792
Alpha deviation (vs. broad)	0.2307**	5.02		
Absolute alpha deviation	0.4608**	15.84		
B. Corporate bond funds				
1–3	–2.6849**	–6.27	2.975	0.746
1–5	–2.4891**	–6.19	1.974	0.774
3–5	–2.1847**	–5.53	1.201	0.782
5–7	–2.1270**	–5.30	1.053	0.780
Broad	–1.9996**	–4.59	0.945	0.773
5–10	–1.7094**	–4.10	0.721	0.772
7–15	–1.5926**	–3.69	0.605	0.761
10–15	–1.6155**	–3.74	0.529	0.744
5+	–1.6234**	–3.77	0.532	0.751
10+	–1.3687**	–3.44	0.351	0.707
Maturity matched	–2.2165**	–5.53	1.343	0.785
Alpha deviation (vs. broad)	0.2169**	3.38		
Absolute alpha deviation	0.8102**	18.09		

Notes: This table shows the equal-weighted average net-of-fee alphas, alpha deviations, and absolute alpha deviations of active US domestic bond funds from January 1990 to December 2014, where performance is measured using Equation 2 with different Treasury indexes as benchmarks. Panel A reports government bond funds, and Panel B reports corporate bond funds. Alphas and alpha deviations are in percentage per annum. Alpha deviation is the difference between the broad alpha and the MM alpha. Statistical significance of alpha deviations and absolute alpha deviations is based on paired *t*-tests. "Maturity matched" matches the fund and benchmark index on the basis of their average durations during the fund's life span in our sample period.

**Heteroskedasticity-consistent statistical significance at the 1% level.

bond funds and benchmarked separately against all 10 Treasury indexes and against the funds' maturity-matched benchmarks. The first interesting finding is that average alpha increases monotonically with the maturity range of the common benchmark index. Consistent with the index-based analysis in the introduction, this finding shows that the choice of benchmark affects conclusions regarding the active bond selection performance of the average bond fund.

The average MM alpha (–1.3873% p.a.) is less than the average broad alpha (–1.1565% p.a.), which is in line with the average fund duration being 1.3 years shorter than that of the broad index (see Tables 1 and A2). The resulting alpha deviation of 0.2307% p.a. is statistically significant and shows that using the broad index for all funds systematically overestimates the average active bond selection performance of the funds in our sample. Additionally, the result that the average absolute alpha deviation of 0.4608% p.a. is approximately twice the average alpha deviation demonstrates that the choice of benchmark is crucial for conclusions regarding the active bond selection performance of each individual fund. Regarding the other parameters, the average broad beta is less than 1 (0.883), reflective of the shorter average duration of the funds compared with the index. The average MM beta is close to 1 (1.061), and the highest R^2 is for the MM alpha (0.792).

Panel B reports similar findings for corporate bond funds. The MM alpha (–2.2165% p.a.) is less than the broad alpha (–1.9996% p.a.), indicating a general overestimation of the average active bond selection performance in our sample. The average alpha deviation is statistically significant at 0.2169% p.a., and the average absolute alpha deviation is 0.8102% p.a., indicating extreme positive and negative alpha deviations on the individual fund level. This finding shows that the alpha deviation is a problem not only in government bond funds, where IR risk is the most important determinant of expected return, but also in corporate bond funds, where other systematic risks, especially default risk, also play important roles.

Alternative Matching Criteria and Performance Models

The previous sections show the necessity of individually selecting an adequate IR risk benchmark for each fund. We advocate matching funds and benchmarks using their reported durations. However, if the funds do not report their durations, our proposed matching between index and fund is not applicable. In the future, with fund reporting becoming more and more transparent, we expect this problem to disappear. In the meantime, **Table 3** tests four alternative matching criteria and benchmark indexes for their ability to solve the alpha deviation problem.

As the first alternative, we looked at the funds' names because many of them include the funds' targeted maturity range, either explicitly or in the short-term, intermediate-term, and long-term denominations. We then matched the stated maturity ranges as closely as possible to the stated maturity ranges of the nine Treasury indexes. If the funds' names did not give any information, we used the broad Treasury index. The results are labeled "name match." For government bond funds, in Panel A, the average alpha is slightly greater than the MM alpha from Table 2, which leads to a relatively small average alpha deviation of 0.0843% p.a. Also, the average absolute alpha deviation is relatively small, at 0.2438% p.a., compared with the numbers for the broad alpha in Table 2. The funds' average beta is also closer to 1 relative to the broad alpha. The average R^2 is lower than that for the MM alpha. For corporate bond funds, in Panel B, the alpha deviation is in a range similar to that of the broad index (as shown in Table 2). The average absolute alpha deviation is only slightly smaller. In conclusion, the name match works well for government bond funds but not for corporate bond funds.

As the second alternative, we used the funds' self-reported primary benchmark indexes from Morningstar and labeled the results "primary index." If a fund did not provide such information, we used the broad Treasury index. For government bond funds, in Panel A, the primary index increases the average alpha deviation to 0.6364% p.a. and the average absolute alpha deviation to 0.7761% p.a. The proximity of these results shows that most government bond funds have a positive alpha deviation, which means that they report a primary benchmark with too long a duration. For corporate bond funds, in Panel B, the average alpha deviation becomes very small (–0.0505% p.a.), but the average absolute alpha deviation remains relatively large, at 0.5730% p.a. Many individual corporate bond funds thus still have a high positive or negative alpha deviation if measured against their self-reported primary benchmark index.

The third alternative uses the logic that the funds' MM beta should be close to 1. Moreover, it does not require information from the fund, which could be missing. We calculated Equation 2 for each fund, proceeding through the list of maturity-range Treasury

Table 3. Alternative Matching Criteria

Treasury/Benchmark Index	Alpha	*t*-Statistic	Beta	R^2
A. Government bond funds				
Name match	–1.3029**	–10.00	1.049	0.789
Alpha deviation	0.0843**	2.73		
Absolute alpha deviation	0.2438**	10.47		
Primary index	–0.8343**	–13.20	0.928	0.860
Alpha deviation	0.6364**	4.93		
Absolute alpha deviation	0.7761**	6.41		
Beta match	–1.3029**	–11.74	0.960	0.794
Alpha deviation	0.0844**	2.79		
Absolute alpha deviation	0.1620**	5.86		
R^2 match	–1.2517**	–9.92	0.922	0.804
Alpha deviation	0.1355**	3.91		
Absolute alpha deviation	0.2547**	8.83		
B. Corporate bond funds				
Name match	–2.0054**	–4.62	1.214	0.781
Alpha deviation	0.2111**	3.41		
Absolute alpha deviation	0.6829**	14.00		
Primary index	–2.2104**	–4.72	1.744	0.806
Alpha deviation	–0.0505	–0.88		
Absolute alpha deviation	0.5730**	12.81		
Beta match	–1.9369**	–4.57	0.995	0.784
Alpha deviation	0.2796**	6.80		
Absolute alpha deviation	0.4126**	11.12		
R^2 match	–2.1674**	–5.10	1.196	0.796
Alpha deviation	0.0492	1.04		
Absolute alpha deviation	0.3828**	9.19		

Notes: This table shows the equal-weighted average net-of-fee alphas, alpha deviations, and absolute alpha deviations of active US domestic bond funds from January 1990 to December 2014, where alpha is measured using Equation 2 with different benchmark indexes. Panel A reports government bond funds, and Panel B reports corporate bond funds. Alphas and alpha deviations are reported in percentage per annum. Alpha deviation is the difference between the respective alpha and the MM alpha from Table 2. Statistical significance of alpha deviations and absolute alpha deviations is based on paired *t*-tests. "Name match" uses the maturity range provided in the fund's name. "Primary index" is the fund's self-reported primary benchmark from Morningstar. "Beta match" uses the logic that beta should be 1. "R^2 match" uses the logic that R^2 should be maximized.

**Heteroskedasticity-consistent statistical significance at the 1% level.

indexes, and chose the index with beta closest to 1 as the beta-match benchmark. For government bond funds, in Panel A, this procedure works very well; the average alpha deviation (0.0844% p.a.) and the average absolute alpha deviation (0.1620% p.a.) become very small. The average beta is close to 1, and the R^2 is high. However, for corporate bond funds, in Panel B, the beta match is less successful, with an average alpha deviation of 0.2796% p.a. and an average absolute alpha deviation of 0.4126% p.a.

The fourth alternative follows the observation in Table 2 that the MM alpha has the highest R^2. Therefore, we calculated Equation 2 for each fund, proceeding through the list of maturity-range Treasury indexes, and chose the index with the highest R^2 as the R^2-match benchmark. For government bond funds, in Panel A, the procedure works reasonably well, producing a relatively small average alpha deviation (0.1355% p.a.) and a relatively small average absolute alpha deviation (0.2547% p.a.). For corporate bond funds, in Panel B, the R^2 match produces the best results, with a very small average alpha deviation (0.0492% p.a.) and a relatively small average absolute alpha deviation (0.3828% p.a.)

Fund Ratings and the Alpha Deviation

An important practical question is whether popular and widely acknowledged performance metrics are influenced by the alpha deviation problem. Therefore, we analyzed whether the Morningstar rating is more sensitive to the maturity-unmatched broad alpha or to the MM alpha. Following the revealed preference approach of Barber, Huang, and Odean (2016), we ran a "horse race" between broad alpha and MM alpha in predicting the Morningstar rating separately for government and corporate bond funds. Equation 3 estimates the relation between rating and two-dimensional decile rank dummies based on the competing models:

$$Rating_{i,t+1} = \alpha + \sum_j \sum_k b_{Broad,MM} D_{Broad,MM,i,t} + \mathbf{c}\mathbf{X}_{i,t} + \pi_i + \mu_t + \varphi_s + \varepsilon_{i,t}. \quad (3)$$

The dependent variable ($Rating_{i,t+1}$) is the Morningstar rating of fund i in month t + 1. $D_{Broad,MM,i,t}$ is a dummy variable that takes on a value of 1 if fund i is in Decile j based on broad alpha and Decile k based on MM alpha in month t, and $b_{Broad,MM}$ is the respective regression coefficient.[14]

The matrix $\mathbf{X}_{i,t}$ represents control variables, and **c** is a vector of associated coefficients. We included the previous two ratings instrumented by their respective lags to control for endogeneity due to the dynamic panel. As further controls, we included the fund characteristics listed in Table 1. We ran the regression with fund (π), time (μ), and style fixed effects (φ).

Panel A of **Table 4** presents two hypothesis tests: (1) The average difference in coefficient estimates ($b_{Broad,MM}$ – $b_{MM,Broad}$) is zero, and (2) the proportion of positive differences is 50%. For government bond funds, the average coefficient difference is positive but statistically insignificant. The proportion of positive differences is 60%. This finding indicates that Morningstar ratings of government bond funds are slightly better predicted by the broad alpha than by the MM alpha. However, it seems that the Morningstar ratings of government bond funds account reasonably well for the style return from passively harvesting the IR risk premium.

In contrast, the average coefficient difference of corporate bond funds is positive, large, and statistically significant. Also, the proportion of positive coefficient differences of 80% is statistically significant. Thus, the Morningstar ratings of corporate bond funds seem not to account properly for return expectations due to passive IR risk exposure because the broad alpha is distinctly better at predicting Morningstar ratings than the MM alpha is.

Investor Flows and the Alpha Deviation

Finally, we analyzed whether investor decisions are affected by alpha deviations, which would be the case if investor flows were more sensitive to broad alpha than to MM alpha. Once again, following Barber et al. (2016), we ran a horse race between the two measures, this time for attracting future investor flows. Therefore, we replaced the dependent variable in Equation 3 with $Flow_{i,t+1}$, which is fund i's cumulative monthly fund flow over the next 12 months. To control for the dynamic panel, we included the last two annual flows instrumented by their respective lags. We assigned the alpha decile ranks separately within six groups to account for the fact that different types of investors may use different performance measures for their investment decisions. These six groups are institutional versus retail government bond funds, institutional versus retail corporate bond funds, and high-yield versus investment-grade corporate bond funds.[15]

Table 4. Predicting Morningstar Ratings and Investor Flows

A. Morningstar ratings

	Government	Corporate
Average coefficient difference	0.0181	0.1266**
t-Statistic	0.92	3.70
% of coefficient differences > 0	60.00	80.00**

B. Investor flows

	Government		Corporate			
	Institutional	Retail	Institutional	Retail	High Yield	Investment Grade
Average coefficient difference	0.1445**	0.0148	0.0032	0.0868**	–0.0607	0.1118**
t-Statistic	3.05	0.60	0.11	2.48	–1.30	3.99
% of coefficient differences > 0	66.67*	33.33	55.56	68.89**	46.67	73.33**

Notes: Panel A presents horse races between the broad alpha and the MM alpha to predict the Morningstar ratings of government and corporate bond funds. Panel B presents similar horse races to predict the investor flows of institutional and retail government bond funds, as well as institutional, retail, high-yield, and investment-grade corporate bond funds. Following Barber et al. (2016), we estimated the relation between future Morningstar rating (flow over the next 12 months) and two-dimensional decile ranking based on the broad alpha and the MM alpha by using Equation 3. As control variables, we included the previous two ratings (annual flows) instrumented by their respective lags to control for endogeneity due to the dynamic panel, log size, log age, turnover ratio, expense ratio, percentage cash, percentage corporate bonds, percentage government bonds, and indicators for institutional and high-yield funds. We also included fund (π), time (μ), and style fixed effects (φ) based on the Morningstar Fixed-Income Style Box. We compared the coefficients for which the decile ranks are of the same magnitude but for which the ordering is reversed. For example, we compared $b_{10,1}$ (mean rating or flow for a top-decile broad alpha and bottom-decile MM alpha fund) with $b_{1,10}$ (mean rating or flow for a bottom-decile broad alpha and top-decile MM alpha fund). To estimate the model, we excluded the dummy variable $D_{5,5,i,t}$. The table presents the results of two hypothesis tests for each horse race: (1) H0: The average difference in coefficient estimates is zero (*t*-test), and (2) H0: the proportion of positive differences is 50% (binomial test).

*Statistically significant at the 5% level.

**Statistically significant at the 1% level.

Panel B of Table 4 shows the results. For institutional government bond funds, the average coefficient difference is positive, large, and statistically significant. Moreover, the proportion of positive differences is also significant, at 66.67%. Thus, broad alpha is significantly better at predicting or attracting flows. This finding has two possible causes: Either institutional investors in government bond funds do not account properly for the IR risk exposure in determining active bond selection performance, or they do not care about a proper separation of active and passive return contributions and instead invest on the basis of total return.

For retail investors in government bond funds, the results are ambiguous: The average coefficient difference is positive but very small and insignificant, and the proportion of positive differences is less than 50%. Thus, both measures seem to do equally well in predicting or attracting investor flows.

For corporate bond funds, these findings seem to be reversed. For institutional investors in corporate bond funds, the average coefficient is near zero and the proportion of positive differences is near 50%. Thus, the models do equally well in predicting flows. Retail investors in corporate bond funds, however, have a positive and significant average coefficient difference and a significant proportion of positive differences, at 68.89%. Thus, investment decisions of retail corporate bond fund investors are significantly better predicted by the maturity-unmatched broad alpha. Again, two possible explanations apply: Investors either do not account properly for active versus passive returns or do not care about the separation.

Finally, looking at high-yield corporate bond fund investors, we found a slightly negative but insignificant average coefficient difference and a proportion of positive differences near 50%, which indicates that the models do equally well in predicting flows. Considering that these funds should be the ones for which IR risk is the least important, this result is unsurprising. For investment-grade corporate bond funds, for which IR risk should be an important determinant of expected return, we once again found a large, positive, and significant average coefficient difference and a significant proportion of positive differences, at 73.33%. This finding indicates once more that the broad alpha is the better predictor of flows, because investors either do not know or do not care about the alpha deviation or a proper separation of active and passive return contributions.

Conclusion

We documented systematic alpha deviations in measured active bond fund selection performance. These alpha deviations originate from using linear regressions to measure alpha despite the known nonlinearity between interest rate risk exposure and expected return. Our simple and practical remedy for the problem is to individually match funds and benchmark indexes using their durations, thereby assigning to each fund just the appropriate IR risk premium without the need to lever it linearly to the IR risk exposure of the fund. This approach combines the firmly established and flexible regression method with the holdings-based idea of matching bonds and benchmarks on their durations and requires much less data. If the fund's duration is unknown, we identify beta and R^2 as promising candidates for alternative matchings.

During our sample period, the alpha deviation led to an overestimation of the average bond fund's performance. On the individual fund level, many funds had high positive or negative alpha deviations, depending systematically on the size and direction of their maturity mismatch with the usual broad, intermediate-term benchmark index applied to all funds. We also showed that the popular Morningstar rating and investor flows are influenced by the alpha deviation, because both had stronger responses to maturity-mismatched performance measures than to maturity-matched performance measures.

Appendix A. Bond Fund Performance Literature and US Treasury Index Summary Statistics and Expected Returns

Table A1 provides a summary of the bond fund performance literature. **Table A2** shows US Treasury index summary statistics. **Figure A1** shows expected returns of the shortest- and longest-maturity-range US Treasury indexes.

Table A1. Bond Fund Performance Literature

Authors	Year	Journal	Indexes/Factors Used to Capture Interest Rate Risk
Cornell and Green	1991	JF	Broad Treasury index (t − 1, t, t + 1)
Fama and French	1993	JFE	Term factor (long-term − short-term Treasury)
Blake, Elton, and Gruber	1993	JB	Broad government index, broad bond index, intermediate + long-term government indexes
Elton, Gruber, and Blake	1995	JF	Broad bond index
Zhao	2005	FAJ	Broad bond index as in Elton et al. (1995)
Ferson, Hendry, and Kisgen	2006	RFS	Level, slope, and curvature (yields) as state variables in conditional stochastic discount factor (SDF) models
Huij and Derwall	2008	JBF	Broad bond index as in Elton et al. (1995)
Bessembinder, Kahle, Maxwell, and Xu	2009	RFS	Broad bond index as in Elton et al. (1995), term factor as in Fama and French (1993)
Comer, Larrymore, and Rodriguez	2009	RFS	Intermediate + long-term government indexes as in Elton et al. (1995)
Chen, Ferson, and Peters	2010	JFE	Level, slope, and curvature (yields) as state variables in conditional SDF models

(continued)

Table A1. Bond Fund Performance Literature (continued)

Authors	Year	Journal	Indexes/Factors Used to Capture Interest Rate Risk
Ammann, Kind, and Seiz	2010	JBF	Term factor as in Fama and French (1993)
Cici, Gibson, and Merrick	2011	JFE	Holdings-based, matching bonds and characteristic-based benchmarks on duration
Cici and Gibson	2012	JFQA	Holdings-based, matching bonds and characteristic-based benchmarks on duration
Amihud and Goyenko	2013	RFS	Broad bond index as in Elton et al. (1995), term factor as in Fama and French (1993)
Huang and Wang	2014	MS	Term factor as in Fama and French (1993)
Adam and Guettler	2015	JFE	Broad bond index as in Elton et al. (1995)
Moneta	2015	JEF	Holdings-based, matching bonds and characteristic-based benchmarks on duration
Goldstein, Jiang, and Ng	2017	JFE	Broad bond index as in Elton et al. (1995)
Chen and Qin	2017	MS	Broad government index as in Cornell and Green (1991), broad bond index as in Elton et al. (1995)
Bai, Bali, and Wen	2019	JFE	Broad bond index as in Elton et al. (1995)
Palhares and Richardson	2020	FAJ	Medium-duration Treasury index

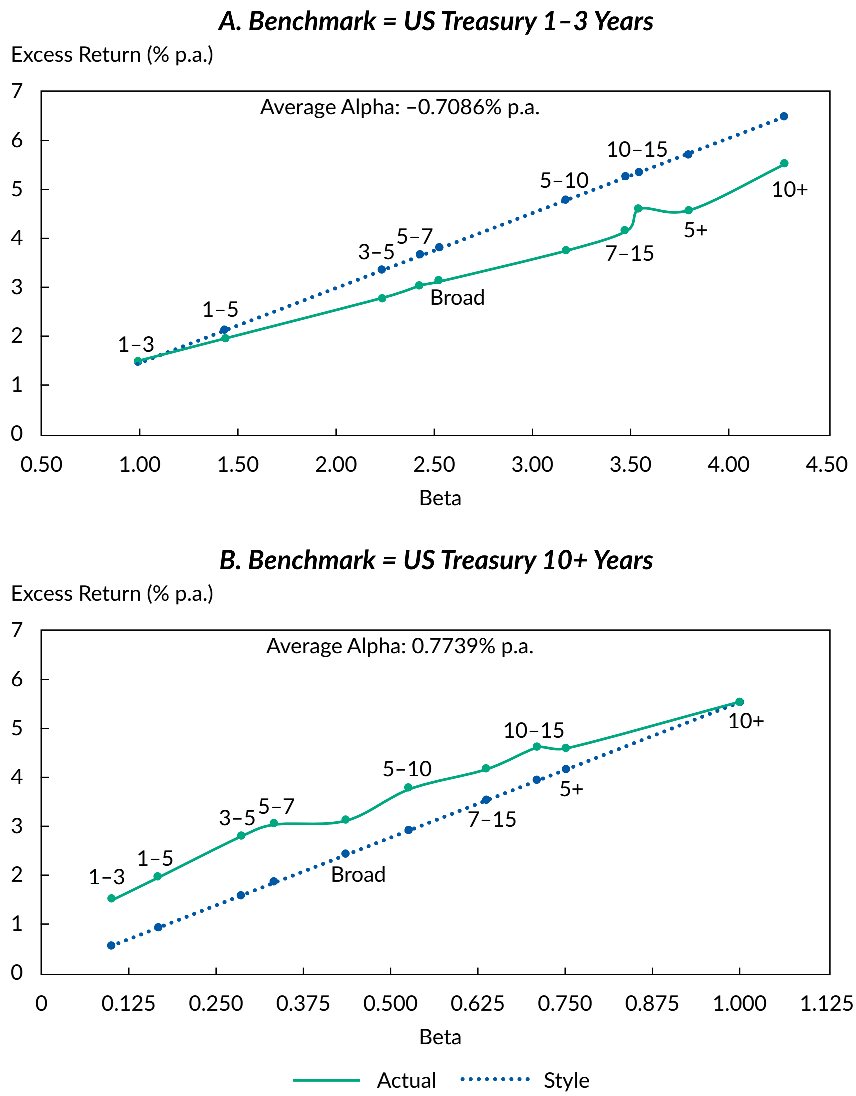
Notes: JF is the *Journal of Finance*. JFE is the *Journal of Financial Economics*. JB is the *Journal of Business*. FAJ is the *Financial Analysts Journal*. RFS is the *Review of Financial Studies*. JBF is the *Journal of Banking & Finance*. JFQA is the *Journal of Financial and Quantitative Analysis*. MS is *Management Science*. JEF is the *Journal of Empirical Finance*.

Table A2. US Treasury Index Summary Statistics

		Expected Excess Return		Linear Regression Parameters	
	Duration	Actual	Style	Alpha	Beta
1–3 years	1.75	1.51	0.88	0.64	0.28
1–5 years	2.43	1.97	1.39	0.59	0.44
3–5 years	3.55	2.81	2.32	0.50	0.74
5–7 years	4.99	3.06	2.62	0.43	0.84
Broad	5.44	3.13	3.13	0.00	1.00
5–10 years	6.02	3.78	3.86	–0.09	1.23
7–15 years	7.16	4.17	4.50	–0.32	1.43
10–15 years	7.99	4.62	4.85	–0.23	1.55
5+ years	8.95	4.60	5.17	–0.57	1.65
10+ years	12.09	5.54	6.43	–0.88	2.05

Notes: This table shows mean durations (in years), excess returns over the one-month T-bill (in percentage per annum), and linear regression parameters alpha and beta of 10 ICE BofA Merrill Lynch US Treasury total return indexes from January 1990 to December 2014. Nine of the indexes hold Treasury bonds with maturities within specific ranges; one broad index holds Treasury bonds of all maturities. "Actual" is the arithmetic mean and indicates the expected return under the independently and identically distributed assumption. "Style" is the product of beta and the actual mean of the broad index. Linear regression parameters alpha and beta are estimated using Equation 1.

Figure A1. Expected Returns of US Treasury 1–3 and 10+ Year Indexes



Notes: See notes to Figure 1. This figure plots the actual mean excess returns over the one-month T-bill (in percentage per annum) of various maturity-range-specific passive US Treasury bond total return indexes against their betas from linear regressions using Equation 1 and applying the 1–3 year index (Panel A) and the 10+ year index (Panel B) as the common benchmark.

Notes

1. Sharpe (1992) defined the style benchmark return as the sum of beta-weighted factor returns. Daniel, Grinblatt, Titman, and Wermers (1997) used the term *average style return*. Other popular terms, such as *smart beta* (e.g., Kahn and Lemmon 2016) and *factor investing* (e.g., Clarke, de Silva, and Thorley 2016), describe the same kind of investment decision. In the context of corporate bond funds specifically, Choi and Kronlund (2018) showed that many funds reach for yield by choosing high long-term exposures to interest rate risk and default risk.

2. A wide range of studies have found that mutual funds in general and bond funds in particular show no relevant market-timing performance (e.g., Treynor and Mazuy 1966; Henriksson and Merton 1981; Chen, Ferson, and Peters 2010). Bunnenberg, Rohleder, Scholz, and Wilkens (2019)

showed that constant-beta models, such as Jensen's (1968) alpha, capture the aggregate of selection and timing as total active performance (approximately). Therefore, we concentrate on alpha and trust that it captures total active performance, including timing.

3. Another popular approach is to use a zero-investment factor to capture IR risk exposure constructed as the return difference between a long-term and a short-term Treasury bond index (e.g., Fama and French 1993). However, in unreported tests, we found that such a factor amplifies the problem discussed in our article rather than solving it. See Table A1 in Appendix A for an overview of the bond fund performance measures applied in the most influential bond fund research papers of the past 30 years.

4. $\varepsilon_{i,t}$ is an error term with $E(\varepsilon_{i,t}) = 0$.

5. We obtained the index data for January 1990 to December 2014 from www.theice.com/market-data/indices.

6. Table A2 in Appendix A shows the details of the various expected excess returns and linear regression parameters (alpha, beta), as well as mean durations, of the 10 US Treasury total return indexes.

7. This finding corresponds to durations (maturity ranges) shorter than that of the broad index; see Table A2 in Appendix A.

8. Panel A of Figure A1 in Appendix A shows a similar exercise using the shortest-maturity-range index as the benchmark. All betas are less than 1, all alphas are negative, and average alpha is –0.7086% p.a. Likewise, in Panel B, with the longest-maturity-range index as the benchmark, all betas are greater than 1, all alphas are positive, and average alpha is 0.7739% p.a.

9. We used duration instead of maturity for the matching because it is the better measure of IR risk exposure (Macaulay 1938).

10. In unreported tests, we instead used returns gross of expenses. The results were similar.

11. According to Cici, Gibson, and Merrick (2011), corporate bond funds in particular may have substantial discretion in the valuation of their holdings, which could lead to autocorrelation due to smoothed returns. Details on the unsmoothing procedure are available upon request.

12. In unreported robustness checks, we alternatively used the added value proposed as a performance measure by Berk and van Binsbergen (2015), which scales gross alpha by TNA. The results are economically in line with our main results.

13. In unreported robustness checks, we extended the baseline model in Equation 2 with several additional bond risk factors. Many are freely available from the authors listed in this note but cover only (in some cases, very short) subperiods of our sample period. We used several alternative liquidity or illiquidity factors provided by Dick-Nielsen, Feldhütter, and Lando (2012); Schestag, Schuster, and Uhrig-Homburg (2016); and Bai, Bali, and Wen (2019). We used bond momentum and reversal factors from Jostova, Nikolova, Philipov, and Stahel (2013) and Bai et al. (2019). We used macro factors from Ludvigson and Ng (2009). We used two self-constructed factors capturing corporate issuer size (bond SMB) and constraints (constrained minus unconstrained, or CMU) based on ICE indexes. Using various combinations of these factors renders our main findings and conclusions economically and quantitatively unchanged.

14. For example, $D_{1,10,i,t}$ indicates that fund *i* is in Decile 1 according to broad alpha and in Decile 10 according to MM alpha in month *t*.

15. We considered funds to be institutional if more than 50% of their TNA are invested in their institutional share classes. The high-yield denomination is based on the Morningstar Broad Category Groups.

References

Adam, T., and A. Guettler. 2015. "Pitfalls and Perils of Financial Innovation: The Use of CDS by Corporate Bond Funds." *Journal of Banking & Finance* 55: 204–14.

Amihud, Y., and R. Goyenko. 2013. "Mutual Fund's R^2 as Predictor of Performance." *Review of Financial Studies* 26 (3): 667–94.

Ammann, M., A. Kind, and R. Seiz. 2010. "What Drives the Performance of Convertible-Bond Funds?" *Journal of Banking & Finance* 34 (11): 2600–13.

Bai, J., T. Bali, and Q. Wen. 2019. "Common Risk Factors in the Cross-Section of Corporate Bond Returns." *Journal of Financial Economics* 131 (3): 619–42.

Barber, B. M., X. Huang, and T. Odean. 2016. "Which Factors Matter to Investors? Evidence from Mutual Fund Flows." *Review of Financial Studies* 29 (10): 2600–42.

Berk, J. B., and J. H. van Binsbergen. 2015. "Measuring Skill in the Mutual Fund Industry." *Journal of Financial Economics* 118 (1): 1–20.

Bessembinder, H., K. M. Kahle, W. F. Maxwell, and D. Xu. 2009. "Measuring Abnormal Bond Performance." *Review of Financial Studies* 22 (10): 4219–58.

Blake, C. R., E. J. Elton, and M. J. Gruber. 1993. "The Performance of Bond Mutual Funds." *Journal of Business* 66 (3): 371–403.

Bunnenberg, S., M. Rohleder, H. Scholz, and M. Wilkens. 2019. "Jensen's Alpha and the Market-Timing Puzzle." *Review of Financial Economics* 37 (2): 234–55.

Chen, Y., W. Ferson, and H. Peters. 2010. "Measuring the Timing Ability and Performance of Bond Mutual Funds." *Journal of Financial Economics* 98 (1): 72–89.

Chen, Y., and N. Qin. 2017. "The Behavior of Investor Flows in Corporate Bond Mutual Funds." *Management Science* 63 (5): 1365–81.

Choi, J., and M. Kronlund. 2018. "Reaching for Yield in Corporate Bond Mutual Funds." *Review of Financial Studies* 31 (5): 1930–65.

Cici, G., and S. Gibson. 2012. "The Performance of Corporate Bond Mutual Funds: Evidence Based on Security-Level Holdings." *Journal of Financial and Quantitative Analysis* 47 (1): 159–78.

Cici, G., S. Gibson, and J. J. Merrick. 2011. "Missing the Marks? Dispersion in Corporate Bond Valuations across Mutual Funds." *Journal of Financial Economics* 101 (1): 206–26.

Clarke, R., H. de Silva, and S. Thorley. 2016. "Fundamentals of Efficient Factor Investing." *Financial Analysts Journal* 72 (6): 9–26.

Comer, G., N. Larrymore, and J. Rodriguez. 2009. "Controlling for Fixed-Income Exposure in Portfolio Evaluation: Evidence from Hybrid Mutual Funds." *Review of Financial Studies* 22 (2): 481–507.

Cornell, G., and K. Green. 1991. "The Investment Performance of Low-Grade Bond Funds." *Journal of Finance* 46 (1): 29–48.

Couts, S., A. S. Gonçalves, and A. Rossi. 2020. "Unsmoothing Returns of Illiquid Funds." Working paper (27 November). Available at https://ssrn.com/abstract=3544854.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers. 1997. "Measuring Mutual Fund Performance with Characteristic-Based Benchmarks." *Journal of Finance* 52 (3): 1035–58.

Dick-Nielsen, J., P. Feldhütter, and D. Lando. 2012. "Corporate Bond Liquidity Risk before and after the Onset of the Subprime Crisis." *Journal of Financial Economics* 103 (3): 471–92.

Elton, E. J., M. J. Gruber, and C. R. Blake. 1995. "Fundamental Economic Variables, Expected Returns, and Bond Fund Performance." *Journal of Finance* 50 (4): 1229–56.

Fama, E. F., and K. R. French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics* 33 (1): 3–56.

——. 2010. "Luck versus Skill in the Cross-Section of Mutual Fund Returns." *Journal of Finance* 65 (5): 1915–47.

Ferson, W., T. R. Hendry, and D. J. Kisgen. 2006. "Evaluating Government Bond Fund Performance with Stochastic Discount Factors." *Review of Financial Studies* 19 (2): 423–55.

Goldstein, I., H. Jiang, and D. T. Ng. 2017. "Investor Flows and Fragility in Corporate Bond Funds." *Journal of Financial Economics* 126 (3): 592–613.

Henriksson, R. D., and R. C. Merton. 1981. "On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills." *Journal of Business* 54 (4): 513–33.

Huang, J.-Z., and Y. Wang. 2014. "Timing Ability of Government Bond Fund Managers: Evidence from Portfolio Holdings." *Management Science* 60 (8): 2091–109.

Huij, J., and J. Derwall. 2008. "'Hot Hands' in Bond Funds." *Journal of Banking & Finance* 32 (4): 559–72.

Jensen, M. C. 1968. "The Performance of Mutual Funds in the Period 1945–1964." *Journal of Finance* 23 (2): 389–416.

Jostova, G., S. Nikolova, A. Philipov, and C. W. Stahel. 2013. "Momentum in Corporate Bond Returns." *Review of Financial Studies* 26 (7): 1649–93.

Kahn, R. N., and M. Lemmon. 2016. "The Asset Manager's Dilemma: How Smart Beta Is Disrupting the Investment Management Industry." *Financial Analysts Journal* 72 (1): 15–20.

Litterman, R., and J. Scheinkman. 1991. "Common Factors Affecting Bond Returns." *Journal of Fixed Income* 1 (1): 54–61.

Ludvigson, S. C., and S. Ng. 2009. "Macro Factors in Bond Risk Premia." *Review of Financial Studies* 22 (12): 5027–67.

Macaulay, F. R. 1938. *The Movements of Interest Rates, Bond Yields and Stock Prices in the United States since 1856*. New York: NBER.

Moneta, F. 2015. "Measuring Bond Mutual Fund Performance with Portfolio Characteristics." *Journal of Empirical Finance* 33 (September): 223–42.

Nelson, C. R., and A. F. Siegel. 1987. "Parsimonious Modeling of Yield Curves." *Journal of Business* 60 (4): 473–89.

Palhares, D., and S. A. Richardson. 2020. "Looking under the Hood of Active Credit Managers." *Financial Analysts Journal* 76 (2): 82–102.

Schestag, R., P. Schuster, and M. Uhrig-Homburg. 2016. "Measuring Liquidity in Bond Markets." *Review of Financial Studies* 29 (5): 1170–219.

Sharpe, W. F. 1992. "Asset Allocation: Management Style and Performance Measurement." *Journal of Portfolio Management* 18 (2): 7–19.

Treynor, J. L., and K. K. Mazuy. 1966. "Can Mutual Funds Outguess the Market?" *Harvard Business Review* 44 (4): 131–36.

Zhao, X. 2005. "Determinants of Flows into Retail Bond Funds." *Financial Analysts Journal* 61 (4): 47–59.