Received: 23 April 2025 | Revised: 13 August 2025 | Accepted: 21 August 2025

DOI: 10.1002/rfe.70024

ORIGINAL ARTICLE

REVIEW of FINANCIAL ECONOMICS | THE UNIVERSITY of NEW ORLEANS | WILEY

Measuring the performance of government bond portfolios with index-based level, slope, and curvature factors

Martin Rohleder[1,2]

[1]Faculty of Economics and Business Administration, University of Augsburg, Augsburg, Germany
[2]School of Economics, Utrecht University, Utrecht, The Netherlands

Correspondence
Martin Rohleder, Faculty of Economics and Business Administration, University of Augsburg, Universitätstr. 16, 86159 Augsburg, Germany.
Email: martin.rohleder@uni-a.de; m.rohleder@uu.nl

Abstract

This paper introduces a three-factor interest rate risk model to improve the measurement of active bond fund performance. Traditional models assume a linear relationship between risk exposure and expected returns, leading to biases. By incorporating level, slope, and curvature factors derived from Treasury index returns, the proposed model better captures the nonlinear nature of bond returns. Empirical tests on passive and active US government bond portfolios confirm its accuracy in estimating passive style returns and active alpha. The study also provides the first performance analysis of fixed-income separate accounts, revealing their economic significance and superior value-added performance over mutual funds.

KEYWORDS

bond funds, curvature, level, performance, separate accounts, slope, Treasury indexes, yield curve

JEL CLASSIFICATION

G20, G11, G23

1 | INTRODUCTION

The performance of active bond funds is usually measured as the intercept of a linear time series regression explaining the fund's excess returns with those of a broadly diversified bond or Treasury market index plus additional risk and style factors.[1] In such a regression, the broad Treasury index primarily captures the interest rate risk premium, the respective beta coefficient represents the portfolio's risk exposure, and their product represents the portfolio's expected passive return from harvesting the premium. The average of the remaining return—alpha—is the portfolio's active performance from bond selection and market timing.[2] This logic is based on the crucial assumption that the interest rate risk premium is in fact a linear function of exposure. However, it is well known that expected bond (portfolio) returns normally follow a concave function of exposure,[3] like the normally decreasing upward-sloping yield curve. Seminal papers in the bond pricing and yield curve analysis literature show that the term structure of expected bond returns may be best described by nonlinear multipolinomial functions (e.g., Nelson & Siegel, 1987; Svensson, 1994). The first three factors–usually referred to as level, slope, and curvature—have been shown to explain the vast majority of term structure variations (e.g., Litterman & Scheinkman, 1991), with the curvature factor representing the main source of nonlinearity.

Bond portfolio performance models which do not account for this nonlinearity thus underestimate or overestimate the expected passive return and, consequentially, overestimate or underestimate active portfolio performance. Natter et al. (2021) propose a solution to this problem by matching funds and benchmark indexes via their reported durations. However, the matching requires data on the durations of funds and indexes, which are not always available, and there may not be a suitable index for each fund. Therefore, I propose a generalized three-factor interest rate risk model that explicitly considers the shape of the return curve via its main parameters level, slope, and curvature (e.g., Litterman & Scheinkman, 1991). I construct intuitive, in principle investable factors capturing these parameters using the returns of three Treasury indexes and show that the model produces unbiased estimates of expected passive style return and active alpha.

Figure 1 illustrates the construction of the factors. It shows a hypothetical normal return curve (solid line) with the expected return as a decreasingly upward-sloping function of risk exposure (e.g., beta or duration). I define the level of the return curve as the return of a short-term Treasury index (dashed line).[4] The slope of the return curve is the return difference between a long-term index and the short-term index (e.g., Fama & French, 1993). Comparing the upward-sloping dotted line and the return curve illustrates that considering only level and slope underestimates the expected returns of intermediate-term portfolios. This would result in a positive alpha which is, however, based on miscalculation and not on active management. Likewise, the expected returns of very long-term indexes are overestimated and, consequentially, result in negative alphas. Therefore, I consider the curvature of the return curve as the return difference between an intermediate-term index and the linear combination of the long-term index and the short-term index.[5] The expected curvature beta of an intermediate-term index is positive; that of a very long-term index is negative.

Hence, only three indexes are necessary to calculate the factor returns. This intuitive and simple construction produces factors that are passive diversified alternatives to an active portfolio strategy. Further, they are, at least in the later part of my sample period, investable via ETFs and index mutual funds, which is an important feature in portfolio performance measurement (e.g., Berk & van Binsbergen, 2015; Cremers et al., 2012).

The first part of my empirical analysis tests the effectiveness of the three-factor interest rate risk model in avoiding the miscalculation of passive style return and active performance. I apply the model to 75 AAA-rated passive US Treasury bond indexes with different stated maturity ranges. Their expected returns are exclusively determined by interest rate risk, and they have zero active alpha by definition, which makes them ideal test assets. I benchmark the model against three other models commonly used in the previous literature.[6] The tests show that my model produces the highest R^2, very small deviations from the ideal alpha of zero, and the expected pattern in the betas to the three factors. This holds in time series regressions and rolling window regressions. Further, with Fama and MacBeth (1973) tests, I show that all three factors are priced and thus important in determining expected passive style returns. Finally, I show that the alphas from the three-factor interest rate risk model cannot be explained by economic conditions—changes of the US Treasury yield curve parameters—while the alphas from other models can. This shows that the alphas from other models are potentially biased by external conditions while the alphas from my model represent unbiased estimates of passive expected returns and active management performance.

In the second part of the empirical analysis, I apply the model to active US domestic government bond portfolios: 125 mutual funds (MFs) and 129 separate accounts (SAs). These two types of investment vehicles have in common that

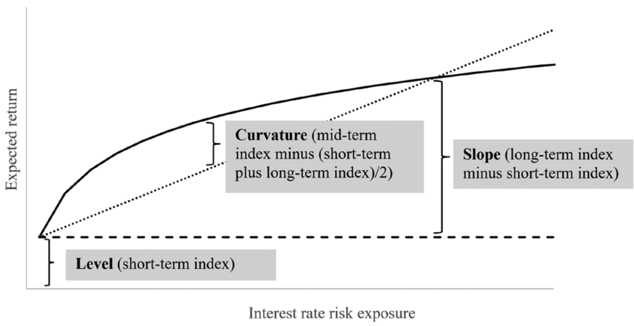

FIGURE 1 Illustration of the interest rate risk factor construction. This figure shows a hypothetical normal expected return curve with expected returns as a decreasingly upward-sloping function of interest rate risk exposure (e.g., beta or duration). Level is the expected return of a short-term index, Slope is the difference between the expected returns of a long-term index and a short-term index, and Curvature is the difference between the expected return of an intermediate-term index and the expected return of a linear combination of the short-term and the long-term index.

they represent managed diversified bond portfolios. However, they also have distinct differences. MFs mainly target retail investors and are usually held by thousands of investors with relatively small investment amounts via shares of a common separate estate. They are highly standardized, and investors have no means of interfering in the management of the portfolio. SAs mainly target institutional or very wealthy individual investors, usually requiring very high initial investments. For each investor, an individual portfolio is structured, usually following a model portfolio representing a specific investment style. SAs are held directly by the investor and are thus easily individualized to the investor's needs, for example, in terms of risk and return preferences or individual tax optimization. Investors may also influence the management of the portfolio. The degree of individualization determines, among other factors, the management fee applicable for the individual SA. My level of observation represents anonymized pools of SAs managed following the same style-specific model portfolio. I will discuss further differences in Section 3.1 with the specific summary statistics of the two types of portfolios.

While bond MFs have been widely studied, bond SAs have received very little attention in the literature thus far. This is astonishing given that the SAs in my sample are nearly 2.5 times as large as the MFs, which gives them greater economic importance. I show that the average alpha before expenses in the period 1977–2024 was significantly negative for MFs with −0.9443% p.a. and for SAs with −0.5842% p.a. Looking at the "value added" (Berk & van Binsbergen, 2015) MFs lost 1.84 million US$ p.a. on average while the larger SAs gained 6.06 million US$ p.a. on average. Dividing the portfolios into groups based on their average durations shows that MFs with durations between 7 and 12 years lost 6.75 million US$ p.a. while MFs with durations higher than 12 years gained 7.63 million US$ p.a. Similarly, SAs with durations higher than 12 years managed to extract 38 million US$ p.a. on average from capital markets. This shows that very long-term portfolios, which are also the largest among the groups on average, are probably managed by more specialized and capable managers, while intermediate-term portfolios are probably managed by lesser managers.

My investigation contributes to the literature on fund performance in general and bond portfolio performance measurement in particular. The first bond fund performance studies occurred in the early 1990s with the studies by Cornell and Green (1991) who used a broad Treasury bond index to account for interest rate risk, and Blake et al. (1993) who used a broad bond index. Around the same time, Fama and French (1993) proposed a zero-investment "term" (respectively "slope") factor to account for interest rate risk in expected portfolio returns. Since then, most studies on bond and bond portfolio performance rely either on single-index or single-factor interest rate risk models. Only very few exceptions use two indexes or factors (see Table A1 in Appendix). I show that three factors—level, slope, and curvature—are necessary to account for the nonlinearity in expected bond returns and propose an intuitive set of factors based on only few common Treasury indexes, which are investable via low-expense ETFs and index mutual funds.

I apply the model to active government bond portfolios and, to the best of my knowledge, present the first analysis of fixed-income SAs. While bond MFs have been intensively investigated in previous studies—albeit not as intensively as equity MFs–fixed-income SAs have been ignored in the literature so far. This is surprising given their relative importance. During my sample period, I observe 125 government bond MFs and 129 SAs, with the latter being almost 2.5 times as large. In addition, government bond SAs outperform MFs on an alpha basis, consistent with previous findings for equity SAs by, for example, Elton et al. (2014), and on the basis of "value added" (Berk & van Binsbergen, 2015). My analysis thus allows valuable new insights into this important class of actively managed bond portfolios.

2 | INDEX ANALYSIS

2.1 | Data

To show the importance of using three factors—level, slope, and curvature—to account for interest rate risk in bond portfolio performance measurement, I start with an analysis based on AAA-rated passive US Treasury bond indexes with different stated maturity ranges as test assets. This has the advantage that I do not have to consider other risks, like, for example, default risk, at this stage and that the indexes have zero active alpha by construction. Hence, any alpha unequal to zero must stem from miscalculated expected passive style returns. I obtain index data from two major fixed-income index providers via Factset: Intercontinental Exchange (ICE) and Bloomberg. For both, I obtain monthly returns, durations, and yields in the period from 1973 to 2024. To keep the analysis as straightforward as possible, I exclude all indexes involving Treasury strips, inflation adjustments, callable bonds, and positive as well as negative leverage. Further, I exclude indexes producing an R^2 of less than 0.67 against a broad Treasury index, which results in a total set of 75 test assets, 42 from ICE and 33 from Bloomberg.

TABLE 1 Government bond index summary statistics.

				Percentiles		
	N	Mean	SD	1st	50th	99th
Effective duration	26,852	6.62	4.63	0.00	5.37	20.01
Maturity	29,537	9.45	7.91	1.41	6.56	29.89
Yield to maturity	23,213	3.69	2.36	0.19	3.34	10.20
Total return	30,919	5.34	24.76	−62.78	3.93	83.31

Note: This table shows summary statistics for 75 US Treasury bond indexes from ICE and Bloomberg in the period from 1973 to 2024. All indexes are rated AAA. Duration and maturity are denoted in years, yields and returns are denoted in % p.a.

Table 1 shows pooled summary statistics. The effective duration, which equals an option-adjusted modified duration, is widely dispersed between 0 years (1st percentile) and 20 years (99th percentile) with an average of 6.62 years. The average maturity is 9.45 years. The average yield is 3.69% p.a. and the average annualized monthly total return is 5.34% p.a.

2.2 | Empirical expected return curve

To verify that the relation between interest rate risk exposure and expected return is indeed nonlinear, Figure 2 plots the mean total returns of all Treasury indexes (gray dots) against their effective durations (upper plot) and their broad Treasury betas (lower plot). Both plots also include the best linear fit (solid line). It is clearly visible that most very short-term index returns are below the line, most medium-term indexes are above the line, and most very long-term indexes are once again below the line. Moreover, the plots also show that the best multipolynomial fit (dashed line) is a decreasingly upward-sloping line as expected. This clearly supports the notion that the nonlinearity of the expected return curve must be considered in bond portfolio performance regressions. Otherwise, the alphas from performance models not considering the nonlinearity will lead to negative alphas for very short-term and very long-term indexes, and to positive alphas for medium-term indexes.

2.3 | Factor construction and performance models

To construct the factors, I define Level as the excess return of a short-term index (Equation 1), Slope as the return difference between a long-term and a short-term index (Equation 2), and curvature ("Curve") as the return difference between an intermediate-term index and the linear combination of the short-term index and the long-term index (Equation 3) (see Figure 1). I chose indexes with inception dates between 1973 and 1977 to ensure a long return history. Further, I chose indexes that produce an approximately duration-neutral curvature factor.[7]

$$\text{Level} = \text{ICE } 1 - 3 \text{ years US Treasury minus risk} - \text{free rate} \tag{1}$$

$$\text{Slope} = \text{ICE } 10 + \text{years US Treasury minus ICE } 1 - 3 \text{ years US Treasury} \tag{2}$$

$$\text{Curve} = \text{ICE } 7 - 10 \text{ years US Treasury minus } (\text{ICE } 10 + \text{plus ICE } 1 - 3)\,/\,2 \tag{3}$$

Panel A of Table 2 shows summary statistics of the effective durations of the three indexes. The numbers show that the duration ranges of the three indexes do not overlap and that the durations of the intermediate-term index and the linear combination are very similar in the mean, the median, and in the 1st percentile. Panel B shows summary statistics on the total excess returns of the factors as well as on the broad Treasury index for comparison. All four have positive average returns. Panel C shows correlations between the factors as well as variance inflation factors (VIF). The factors are strongly positively correlated as expected (e.g., Wilkens, 1994), the VIFs are far below the rule-of-thumb critical value of 5, which indicates that multicollinearity is of no concern.

To assess model performance, I benchmark the three-factor interest rate risk model against other models frequently used in bond portfolio performance evaluation. The most widely used model uses only a broad bond or Treasury market

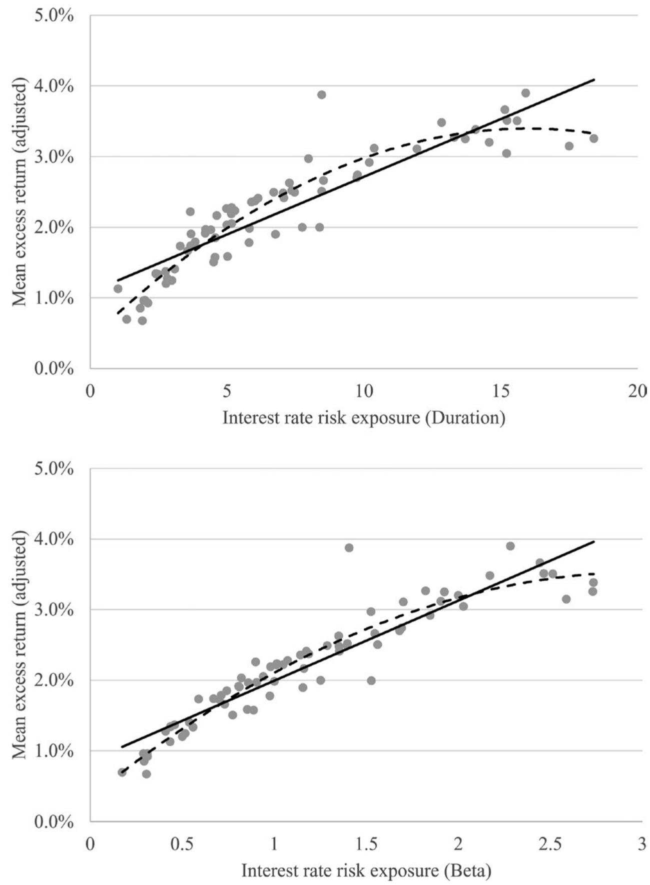

FIGURE 2 Empirical expected return curve. This figure shows the nonlinear relation between mean excess returns (gray dots) and interest rate risk exposure (effective duration and beta against a broad Treasury index) in the period 1973–2024. The solid line represents the best linear fit; the dashed line represents the best multipolynomial fit.

index (e.g., Blake et al., 1993; Cornell & Green, 1991). The most popular alternative is the use of a single slope factor (e.g., Fama & French, 1993). Very seldom, two-factor models with short-term and long-term Treasury indexes or with a level and a slope factor are used. Thus, I compare the four performance models displayed in Equations (4)–(7).

$$\text{Model 1: } \mathrm{er}_{i,t} = \alpha_{(1)i} + \beta_{\mathrm{Broad},i}\ \text{Broad Treasury}_t + \varepsilon_{i,t} \tag{4}$$

$$\text{Model 2: } \mathrm{er}_{i,t} = \alpha_{(2)i} + \beta_{\mathrm{Slope},i}\ \mathrm{Slope}_t + \varepsilon_{i,t} \tag{5}$$

$$\text{Model 3: } \mathrm{er}_{i,t} = \alpha_{(3)i} + \beta_{\mathrm{Level},i}\ \mathrm{Level}_t + \beta_{\mathrm{Slope},i}\ \mathrm{Slope}_t + \varepsilon_{i,t} \tag{6}$$

$$\text{Model 4: } \mathrm{er}_{i,t} = \alpha_{(4)i} + \beta_{\mathrm{Level},i}\ \mathrm{Level}_t + \beta_{\mathrm{Slope},i}\ \mathrm{Slope}_t + \beta_{\mathrm{Curve},i}\ \mathrm{Curve}_t + \varepsilon_{i,t} \tag{7}$$

2.4 | Time series regression alphas

In a first attempt to compare model performance, I run index-by-index time series regressions of all 75 indexes against the four models. An ideal model should produce high R^2, zero alpha on average, and single-index alphas that are small and symmetrically distributed around zero. Table 3 thus shows average alpha, average absolute alpha ($|\alpha|$), median alpha, counts of statistically significant positive and negative alphas at the 5% level, and average R^2. Model 1 produces an average R^2 of 88.02%, an average alpha of −0.0276% p.a. and a median alpha of 0.1580% p.a. Thus, the alphas are positively skewed as indicated by 8 positive and 3 negative significant alphas. The average absolute alpha is 0.3960% p.a., indicating that single alphas deviate strongly from zero.

Model 2 produces very poor results with an average R^2 of 70.66%, an average (absolute) [median] alpha of 0.8023% p.a. (0.8116% p.a.) [0.8670% p.a.] and 42 significant alphas, all of them positive. Model 3 performs much better with an

TABLE 2 Factor summary statistics.

			Percentiles		
	Mean	SD	1st	50th	99th
Panel A. Index effective durations					
ICE 1–3 years US Treasury	1.01	0.88	0.00	1.63	1.92
ICE 10+ years US Treasury	13.69	2.86	10.06	13.71	18.43
ICE 7–10 years US Treasury	7.06	0.64	5.96	7.35	8.07
(1–3 years plus 10+ years)/2	7.73	1.48	5.84	7.79	10.16
Panel B. Factor total excess returns					
Broad Treasury	2.02	19.87	−41.12	1.16	56.38
Level	1.14	9.95	−21.11	0.55	37.69
Slope	2.17	33.77	−83.45	0.00	104.77
Curvature	0.31	6.99	−17.55	0.00	16.89
	Broad	**Level**	**Slope**	**Curvature**	**VIF**
Panel C. Factor correlations and variance inflation factors					
Broad Treasury	1.00				
Level	0.87	1.00			2.13
Slope	0.88	0.56	1.00		1.52
Curvature	0.50	0.56	0.19	1.00	1.50

Note: This table shows summary statistics on the effective durations (in years) of the Treasury bond indexes involved in the factor construction (Panel A), total excess returns (in % p.a.) of the factors used in the performance models (Panel B), and correlation coefficients as well as variance inflation factors (VIFs) between the factors in the period from 1973 to 2024.

TABLE 3 Index time series regression alphas.

	Mean α	Mean $\|\alpha\|$	Median α	Significant	Positive	Negative	Adj. R^2
Model 1	−0.0276	0.3960	0.1580	11	8	3	0.8802
Model 2	0.8023	0.8116	0.8670	42	42	0	0.7066
Model 3	−0.0310	0.2134	−0.0488	9	1	8	0.9365
Model 4	0.0582	0.1922	0.0336	12	5	7	0.9610

Note: This table shows summary statistics on time series alphas (α) from Models 1 to 4, absolute alphas ($|\alpha|$), counts of on the 5% level statistically significant (positive and negative) alphas, and model adjusted R^2 for 75 US Treasury bond indexes in the period from 1977 to 2024. Alphas are denoted in % p.a.

average R^2 of 93.65% and an average alpha of −0.0310% p.a. The average (absolute) alpha is 0.2134% p.a. and the median alpha is −0.0488% p.a. However, of the 9 significant alphas, 8 are negative and thus unsymmetrically distributed around zero. Finally, Model 4 produces the highest average R^2 of 96.10% as well as average and median alphas close to zero with 0.0582% p.a. and 0.0336% p.a., respectively. Also, the average absolute alpha is smallest with 0.1922% p.a. Moreover, the significant alphas are almost equally distributed around zero (5 positive and 7 negative). This shows that Model 4 is an improvement over Model 3 and is thus closest to an ideal model.

2.5 | Time series regression betas

After showing that Model 4 is close to an ideal model in terms of alpha and R^2, I want to take a closer look at patterns in respective factor betas and analyze how important the single factors of the model are. Therefore, Table 4 shows average Model 4 betas and counts of significant betas for the 75 US Treasury indexes in my sample. The betas on all three factors are positive on average and in the median, which is consistent with economic intuition. Especially regarding the curvature beta, I expect a positive mean because most indexes have a medium-term maturity range (see Figure 2). The results further indicate that all three factors are very important as the minimum number of indexes with a significant beta is 61

TABLE 4 Index time series regression betas.

	Mean	Median	Significant	Positive	Negative
Level beta	0.9613	1.0046	64	64	0
Slope beta	0.4566	0.3577	69	66	3
Curvature beta	0.4414	0.4552	61	54	7

Note: This table shows summary statistics on time series betas from Model (4) and counts of on the 5% level statistically significant (positive and negative) betas for 75 US Treasury bond indexes in the period from 1977 to 2024.

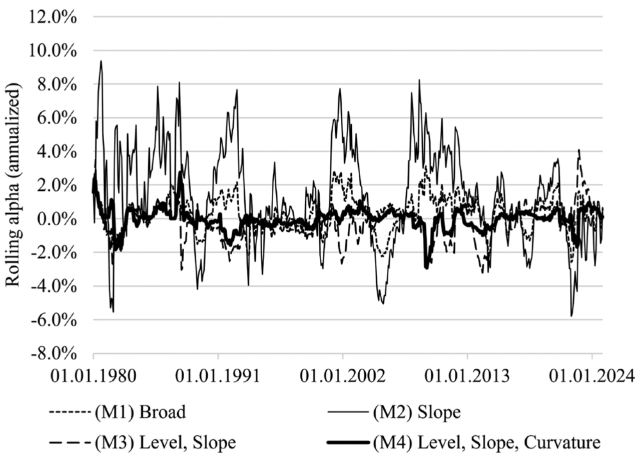

FIGURE 3 Rolling alphas of the ICE 3–7 years US Treasury bond index. This figure shows monthly rolling 12-month alphas of the ICE 3–7 year US Treasury bond index in the period from 1980 to 2024. The alpha estimates are from Models 1 to 4 and denoted in % p.a.

for any given factor. Moreover, of the significant curvature betas, 7 are negative. This is to be expected for very short-term and very long-term indexes as indicated by Figure 2.[8] Thus, this closer look at the patterns in factor betas also confirms the relevance and plausibility of all three interest rate risk factors.

2.6 | Rolling regression alphas

To assess model performance over time, I estimate alphas for the 75 indexes using a monthly rolling 12-month window.[9] As an example, Figure 3 plots rolling alphas from Models 1 to 4 over time for the ICE 3–7 years US Treasury index in the period 1980–2024. The figure shows that Model 4 alpha (thick line) is apparently the most stable and closest to zero over time. Table 5 confirms this notion by showing that Model 4 has an average (median) alpha closest to zero, the lowest standard deviation as well as the least extreme 1st and 99th percentiles. Thus, Model 4 also produces the best results in the rolling window analysis. Similar results apply for the other indexes.[10]

2.7 | Pricing tests of level, slope, and curvature

So far, I used Model 4 to measure the performance of bond portfolios. However, it is also an asset pricing model. An important feature of a good asset pricing model is that it uses only priced factors—that is, all factors must carry a significantly positive risk premium. The usual method to test for a risk premium is to run cross-sectional regressions that explain average returns with factor betas. The analysis presented in Table 6 thus shows the results of Fama and MacBeth (1973) regressions of monthly rolling contemporary (t) and future ($t+1$ year) 12-month mean index excess returns on 12-month level, slope, and curvature betas as shown in Equation (8).[11] Newey and West (1987) standard errors control for heteroscedasticity and autocorrelation.

$$\text{Mean return}_{i,t} = \pi_0 + \pi_1\,\beta_{\text{Level},i,t} + \pi_2\,\beta_{\text{Slope},i,t} + \pi_3\,\beta_{\text{Curve},i,t} + \nu_{i,t} \tag{8}$$

TABLE 5 Rolling alphas of the ICE 3–7 years US Treasury bond index.

			Percentiles		
	Mean (%)	**SD (%)**	**1st (%)**	**50th (%)**	**99th (%)**
Model 1	0.2947	1.0500	−2.1916	0.3468	2.6500
Model 2	1.3928	2.9051	−4.9987	1.1956	7.9521
Model 3	−0.2329	1.0431	−2.7370	−0.1900	2.6607
Model 4	−0.0358	0.6313	−1.8490	0.0099	1.7957

Note: This table shows summary statistics on the monthly rolling 12-month alphas of the ICE 3–7 years US Treasury bond index in the period from 1980 to 2024. The alpha estimates are from Models 1–4 and denoted in % p.a.

TABLE 6 Fama and MacBeth regressions.

Dependent: Mean return	**Contemporary (t)**	**Future ($t+1$ year)**
Level beta	0.6130***	0.2707
	(3.21)	(1.35)
Slope beta	1.8785**	2.1607**
	(2.06)	(2.32)
Curve beta	0.3109**	0.4017***
	(2.13)	(2.70)
Constant	0.6184**	1.0182***
	(2.54)	(4.49)
N	29,925	29,055
Avg. R^2	0.86	0.84

Note: This table shows the results from Fama and MacBeth (1973) regressions of mean returns against level, slope, and curvature betas from monthly 12-month rolling regressions applying model (4) on the excess returns of 75 US Treasury bond indexes in the period 1977–2024. The dependent variable of the cross-sectional regressions is either the contemporary (t) or future ($t+1$ year) 12-month mean index return. The independent variables are the corresponding 12-month level, slope, and curvature betas. The time series t-tests account for heteroscedasticity and autocorrelation via Newey-West standard errors with 5 lags. t-Statistics are in parentheses. ***, **, and * indicate statistical significance on the 1%, 5%, and 10% level, respectively.

The results clearly show that all three factors are relevant in determining the expected passive returns of government bond portfolios. Only the level factor shows an insignificant coefficient on the future mean return. However, the curvature factor is significant in both regressions. Moreover, the average R^2 from the regressions is very high at 86% and 84%, respectively, showing that the three factors combined explain mean index returns very well. This further strengthens my argument that all three factors should be considered in bond portfolio performance measurement as all of them represent priced risks.

2.8 | Rolling regression alphas and economic conditions

As a last index-based analysis, I test if rolling index alphas can be explained by relevant economic conditions. If this is the case, the deviation of index alpha from the ideal value of zero is externally driven and not entirely the result of internal management processes once applied to active portfolios. If alphas cannot be explained by economic conditions, all alphas deviating from zero must stem from internal sources and the performance model works as an assessment tool for active portfolio performance. As economic conditions, I define 12-month changes in level, slope, and curvature parameters of the US Treasury yield curve, as used by, for example, Chen et al. (2010), following Equations (9)–(11).[12]

$$\text{LEVEL} = 1 - \text{year US Treasury yield} \tag{9}$$

$$\text{SLOPE} = 10 - \text{year minus } 1 - \text{year US Treasury yield} \tag{10}$$

$$\text{CURVE} = 5 - \text{year US Treasury yield minus } (1 - \text{year plus } 10 - \text{year yield}) / 2 \tag{11}$$

TABLE 7 Alpha and economic conditions.

Dependent: Alpha	Model 1	Model 2	Model 3	Model 4
ΔLEVEL × Duration	0.0108	−0.0848***	0.0100	−0.0037
	(1.57)	(−7.81)	(1.25)	(−1.17)
ΔSLOPE × Duration	−0.1154***	0.1233***	0.0075	−0.0027
	(−4.28)	(4.43)	(0.70)	(−0.34)
ΔCURVATURE × Duration	0.2677***	−0.2893***	−0.0841**	−0.0046
	(4.25)	(−4.20)	(−2.19)	(−0.26)
Constant	−0.0675	0.8013***	−0.1473**	−0.0326
	(−0.72)	(9.01)	(−2.14)	(−0.52)
N	25,751	25,751	25,748	25,745
Adj. R^2	0.13	0.35	0.03	0.00
F-statistic	14.46***	43.10***	3.79**	0.91

Note: This table shows pooled panel regressions of monthly rolling 12-month index alphas on contemporary interaction terms between the indexes' effective duration and US Department of the Treasury yield curve parameter changes (ΔLEVEL, ΔSLOPE, and ΔCURVATURE) of 75 US Treasury bond indexes in the period from 1990 to 2024. The alphas are from Models 1 to 4. The test statistics control for heteroscedasticity and autocorrelation via index and time-clustered standard errors. *t*-Statistics are in parentheses. ***, **, and indicate statistical significance on the 1%, 5%, and 10% level, respectively.

Further, I acknowledge that the direction of deviations from an ideal alpha of zero depends on the interest rate risk exposure of the index.[13] Therefore, I run pooled panel regressions[14] of 12-month rolling alphas of all n models on contemporaneous interaction terms of the indexes' effective durations ($\text{Duration}_{i,t}$) and the yield curve parameter changes (Δ) following Equation (12).[15] The regressions consider index and time-clustered standard errors to account for heteroscedasticity and autocorrelation.

$$\alpha_{(n)i,t} = \varphi_0 + \varphi_1\ \Delta\text{LEVEL}_t \times \text{Duration}_{i,t} + \varphi_2\ \Delta\text{SLOPE}_t \times \text{Duration}_{i,t} + \varphi_3\ \Delta\text{CURVE}_i \times \text{Duration}_{it} + \eta_{i,t} \tag{12}$$

The results in Table 7 show that the alphas from Models 1 to 3 are significantly explained by economic conditions as evidenced by significant coefficients, high R^2, and significant F-statistics. Hence, the alphas are externally influenced. In contrast, the alphas from Model 4 cannot be explained by economic conditions as evidenced by exclusively insignificant coefficients, zero R^2, and an insignificant F-statistic. Moreover, alphas from Model 3, which consider level and slope but not curvature, are significantly driven by changes in the yield curve's curvature, once more emphasizing that the inclusion of a curvature factor in performance regressions is important. This final index-based test thus confirms that Model 4 can be expected to yield unbiased measures of management performance when applied to active government bond portfolios, whereas the other models may yield biased alpha estimates.

3 | THE PERFORMANCE OF ACTIVE BOND PORTFOLIOS

3.1 | Mutual fund and separate account data

After showing that the consideration of three interest rate risk factors—level, slope and curvature—is important for the unbiased estimation of active performance, I apply the model to active US domestic government bond portfolios: mutual funds (MFs) and separate accounts (SAs). I obtain the data from Morningstar. As search criteria, I used "Base currency = US dollar," "Global Broad Category = Fixed Income," "Investment Area = United States," and "Fixed-Income Primary Sector Government (%) Net >75%" as of February 2025. Then, I manually excluded all non-Treasury and non-government portfolios based on portfolio names, which results in the final samples of 125 mutual funds in the period 1980–2024 and 129 separate accounts in the period 1975–2024.

Low frequency or occasionally missing data like, for example, classifications, holdings percentages in different asset classes, total net assets (TNA), turnover, and expenses are extrapolated to the monthly frequency. Most variables are winsorized to control for outliers and implausible values (e.g., holdings in specific asset classes are winzorized to a maximum of 100%). Returns are censored at the 0.05% and 99.95% percentiles. Returns net and gross of fees are used to calculate

TABLE 8 Mutual fund and separate account summary statistics.

	Mutual funds (125)			Separate accounts (129)		
	N	Mean	SD	*N*	Mean	SD
Net return (% p.a.)	25,458	3.9819	20.6401	22,228	3.1530	20.5196
Gross return (% p.a.)	23,827	4.5684	20.9664	22,428	3.4320	20.5611
Morningstar rating	25,602	3.1154	0.9935	12,886	2.8047	1.1109
Total net assets (mil. US$)	28,416	934.44	3082.21	22,390	2353.06	10,358.36
Expense ratio (% p.a.)	25,816	0.6279	0.3805	22,653	0.4032	0.5141
Turnover ratio (% p.a.)	23,995	140.00	190.59	14,207	66.69	82.68
Credit quality	23,372	1.77 (AA)	1.08	14,868	1.55 (AA)	0.82
Effective duration (years)	21,767	6.27	4.52	14,584	6.64	5.87
Modified duration (years)	13,790	5.69	5.30	11,549	6.97	5.95
No. of holdings	26,265	96	176	16,602	79	192
% Government	26,239	72.52	31.17	16,518	77.43	27.25
% Corporate	7919	5.52	11.61	8253	5.89	11.71
% MBS	20,043	0.02	0.27	14,326	0.03	0.31
% Municipal	26,239	0.00	0.04	16,518	0.01	0.45

Note: This table shows summary statistics on the portfolio characteristics of US domestic government bond mutual funds in the period from 1980 to 2024 and separate accounts in the period from 1977 to 2024. Credit quality ranges from 1 (AAA) to 7 (below B).

expense ratios. Mutual fund share classes are aggregated to the fund level using previous-month share class size as a weighting factor.

Table 8 reports summary statistics of MF and SA characteristics. In comparison, MFs have higher returns (3.9819% p.a. net of fees) than SAs (3.1530% p.a.) and receive higher Morningstar ratings on average (3.1154 vs. 2.8047). SAs are larger by a factor of almost 2.5 ($2353 million vs. $934 million), highlighting their economic importance. Despite the strong standardization, MFs have higher expense ratios (0.6279% p.a. vs. 0.4032% p.a.) which is because they do not require the high initial investment amounts applicable for SAs. The higher variation in SA expense ratios (0.5141% p.a. vs. 0.3805% p.a.) is due to the dependence of the expense ratio on the degree of individualization. MFs have a distinctively higher turnover ratio (140% p.a. vs. 66.69% p.a.) which may indicate that SA investors' hold bonds until maturity while MFs regularly roll over their bonds prior to maturity to remain within their given maturity ranges. This would also be consistent with SAs having slightly better credit quality (1.55 vs. 1.77) to compensate for the higher risk exposure stemming from the longer holding periods. MFs hold a larger number of different bonds (96 vs. 79) and hold a lower percentage in government bonds on average (72.52% vs. 77.43%). The effective durations and percentage holdings in other asset classes such as corporate bonds, MBS, and municipal bonds are rather similar between the two types of investment vehicles.

3.2 | Time series government bond portfolio performance

As shown in Table 8, the government bond portfolios considered in my analysis may have considerable holdings in other fixed-income classes, especially in corporate bonds. Moreover, Chen et al. (2021) documented that bond funds may misclassify holdings to make them appear less risky. Therefore, I measure government bond portfolio performance using performance Model 4e, which is extended by three additional factors to consider default risk (*Def*), MBS option-related risk (Option), and equity-related risks (Equity) which are defined following Equations (13)–(15). This model resembles the "index-4" model from Elton et al. (1995) in spirit. Equation (16) shows the extended performance model.[16]

$$\begin{aligned} \text{Def} = \text{ICE } 5-10 \text{ years U.S. Corporate A-BBB index } [\text{Dur.} = 6.01 \text{ years}] \\ \text{minus ICE } 5-10 \text{ years U.S. Corporate AAA-AA index } [\text{Dur.} = 6.19 \text{ years}] \end{aligned} \tag{13}$$

$$\text{Option} = \text{ICE U.S. Mortgage-¬Backed Securities index [Dur.} = 3.66\text{ years]} \\ \text{minus ICE 3–5 year U.S. Treasury index [Dur.} = 3.67\text{ years]} \quad (14)$$

$$\text{Equity} = \text{US stock market index minus risk} - \text{free rate} \quad (15)$$

$$\text{Model 4e: } er_{i,t} = \alpha_{(4)i} + \beta_{\text{Level},i}\,\text{Level}_t + \beta_{\text{Slope},i}\,\text{Slope}_t + \beta_{\text{Curve},i}\,\text{Curve}_t + \beta_{\text{Def},i}\,\text{Def}_t + \beta_{\text{Option},i}\,\text{Option}_t + \beta_{\text{Equity},i}\,\text{Equity}_t + \varepsilon_{i,t} \quad (16)$$

Table 9 shows the results of portfolio-by-portfolio time series regressions applying Model 4e to net and gross portfolio returns for MFs in Panel A and for SAs in Panel B. The first two columns show results for all MFs and SAs using net and gross returns, respectively. Regarding the factor coefficients, all three interest rate risk factors are relevant and influence portfolio returns with a positive sign on average. Further, default risk and equity-related risk show positive and significant coefficients. The coefficients of option risk are insignificant on average. The average R^2 is 81% for MFs and 85% for SAs, which means improvements of +3% on average over a performance model including only level and slope but not curvature.

Both MFs and SAs show negative and significant alphas net of total expenses, with −0.9443% p.a. for MFs and −0.5842% p.a. for SAs. For gross returns, MFs continue to show a negative alpha with −0.2917% p.a. on average, while that of SAs is statistically insignificant. To summarize, SAs on average outperform MFs, consistent with, for example, Elton et al. (2014). This is despite MFs having the higher total returns in Table 8. However, Table 9 shows that MFs also have higher factor exposures, driving up passive style returns.

3.3 | Index transaction costs and expenses

The indexes I use to construct the factors do not consider fees as would actual investable passive alternatives like ETFs and index funds (e.g., Berk & van Binsbergen, 2015).[17] Thus, the comparison of MFs and SAs with these indexes is not entirely fair. Therefore, I estimate the expense ratios (monthly gross return minus monthly net return) of 117 ETFs on US Treasury and government indexes obtained from Morningstar. The earliest availability of this data is 2002. The average total expense ratio across all ETFs over the period 2002–2024 is 0.1540% p.a. Adding the expense ratio to the fund returns still yields significantly negative alphas net of fees for MFs and SAs as shown in the line "Alpha (Fee adj.)" in Table 9 with −0.7953% p.a. and −0.4314% p.a., respectively. For returns gross of fees, both MFs and SAs have insignificant alpha on average, indicating that they perform on par with passive alternatives before expenses.

3.4 | The "value added" by active government portfolios

Berk and van Binsbergen (2015) argue that it is important to consider the money extracted from capital markets as a measure of portfolio management skill besides alpha. Further, their measure of "value added" by active portfolio management is an intuitive metric to assess economic significance. Therefore, I use the fee adjusted gross return results to additionally calculate the portfolios' value added following Equation (17), where $TNA_{i,t-1}$ is the total net asset value (in million US$) of portfolio i in month t–1. The results are also reported in Table 9.

$$V_i = \frac{1}{T}\sum\nolimits_{t=1}^{T}\left(\left(\alpha_i + \varepsilon_{i,t}\right)\text{TNA}_{i,t-1}\right) \quad (17)$$

The calculation reveals that an average government bond MF provides a negative significant value added of −1.84 million US$ p.a. which means that it lost money to the capital market rather than extracting money during my sample period. Conversely, an average SA shows a positive value added of 6.06 million US$ p.a.; however, not statistically significant.

This positive value added by SAs is due to the fee adjusted gross alpha, which is close to zero and a positive correlation between alpha and size of 12% (ρ(Alpha, TNA)). This means that larger SAs more often have higher alpha. This can be interpreted such that larger SAs are managed by more capable managers than smaller SAs. Similar correlations can be observed for MFs, but this relation does not suffice to generate positive value added on average.

TABLE 9 Active government bond portfolio performance.

	Net	Gross				
	All	All	Short (≤3 years)	Medium (3–7 years)	Long (7–12 years)	Very long (>12 years)
Panel A. Mutual funds						
Level	0.9556***	0.9704***	0.8863***	1.0223***	0.9251***	0.9985*
	(24.75)	(23.14)	(7.71)	(21.70)	(8.01)	(10.19)
Slope	0.3253***	0.3325***	0.0358***	0.3124***	0.4630***	1.0897**
	(12.53)	(12.48)	(4.15)	(21.68)	(6.88)	(15.90)
Curve	0.3769***	0.3791***	0.1126**	0.5125***	0.6361***	−0.0960
	(8.32)	(8.36)	(2.95)	(7.99)	(6.79)	(−0.71)
Def	0.3273***	0.3493***	0.1803**	0.3752***	0.6295***	0.3787
	(6.21)	(6.32)	(2.26)	(4.57)	(4.53)	(2.51)
Option	−0.0196	−0.0269	0.0216	−0.0319	−0.1311*	−0.0984
	(−0.75)	(−1.14)	(1.21)	(−1.03)	(−2.14)	(−1.52)
Equity	0.0407***	0.0419***	0.0166	0.0332***	0.0664***	0.0301
	(4.18)	(4.34)	(1.44)	(4.77)	(4.32)	(0.97)
Alpha	−0.9443***	−0.2917***	−0.7571**	−0.9901***	−1.2786***	−0.5041
	(−8.25)	(−2.90)	(−2.60)	(−7.08)	(−3.70)	(−3.71)
N	22,946	21,452	4454	11,300	2896	2082
#MFs	125	125	19	56	15	8
Adj. R^2	0.81	0.81	0.72	0.82	0.76	0.94
Δ Adj. R^2	+0.03	+0.02	+0.01	+0.03	+0.03	0.00
Alpha (fee adj.)	−0.7953***	−0.1403				
	(−7.01)	(−1.40)				
Value added		−1.84*	0.79	1.51	−6.75**	7.63*
ρ(Alpha, TNA)		0.12	0.03	0.20	−0.13	0.20
TNA		934.44	422.66	1082.91	634.18	1224.01
Curve significant			10	51	12	6
Curve positive			10	50	12	1
Curve negative			0	1	0	5
Panel B. Separate accounts						
Level	0.8885***	0.8902***	0.7710***	0.9073***	0.7372***	1.0672***
	(20.51)	(20.51)	(17.72)	(14.36)	(13.17)	(10.00)
Slope	0.3393***	0.3465***	0.7710***	0.9073***	0.7372***	1.0672***
	(8.41)	(8.52)	(17.72)	(14.36)	(13.17)	(10.00)
Curve	0.2791***	0.2738***	0.0598**	0.5179***	0.8092***	−0.2713*
	(6.37)	(6.19)	(2.08)	(13.49)	(19.28)	(−2.00)
Def	0.2509***	0.2498***	0.0581	0.2552**	0.7114***	0.2375***
	(4.95)	(5.04)	(1.27)	(2.69)	(5.98)	(3.54)
Option	0.0020	−0.0003	0.0534***	−0.0618	−0.1386*	0.0432
	(0.08)	(−0.01)	(2.84)	(−1.48)	(−1.99)	(0.76)
Equity	0.0273***	0.0280***	0.0096***	0.0379**	0.0589***	0.0084
	(4.43)	(4.57)	(2.75)	(2.54)	(5.58)	(0.61)
Alpha	−0.5842***	−0.1837	−0.4505***	−0.8193*	−0.3473	−0.0764
	(−3.20)	(−1.02)	(−3.10)	(−1.74)	(−1.41)	(−0.28)

TABLE 9 (Continued)

	Net	Gross				
	All	All	Short (≤3 years)	Medium (3–7 years)	Long (7–12 years)	Very long (>12 years)
N	20,955	21,209	4767	7804	3250	3282
#SAs	129	129	0.84	0.85	0.74	0.96
Adj. R^2	0.85	0.85	37	43	15	17
Δ Adj. R^2	+0.03	+0.03	+0.02	+0.03	+0.04	+0.01
Alpha (fee adj.)	−0.4314**	−0.0309				
	(−2.37)	(−0.17)				
Value added		6.06	9.42	9.21	1.76	38.30
ρ(Alpha, TNA)		0.12	0.006	0.19	−0.16	0.20
TNA		2353.06	1375.77	1404.18	1664.25	4775.92
Curve significant			19	40	15	11
Curve positive			15	40	15	2
Curve negative			4	0	0	9

Note: This table shows average time series regression coefficients from performance Models 4e for US domestic government bond mutual funds (Panel A) and separate accounts (Panel B) in the period 1977–2024. Alpha is denoted in % p.a. Alpha (fee adj.) is calculated from returns where the average expense ratio of ETFs on US Treasury indexes is added to total portfolio returns. Δ Avg. adj. R^2 denotes the change in adj. R^2 by adding the curvature factor, that is, between models (3) and (4). Value added (in million US$ p.a.) is calculated following Berk and van Binsbergen (2015) from fee adjusted gross returns. ρ(Alpha, TNA) is the correlation between mutual funds' and separate accounts' average size and alpha. TNA (total net assets) is denoted in million US$. The MFs and SAs are also grouped by their average reported effective durations into short-term (≤3 years), medium-term (3–7 years), long-term (7–12 years) and very long-term (>12 years). Curve significant (positive, negative) denotes the count of portfolios with a significant (significantly positive, significantly negative) curvature beta. t-Statistics are reported in parentheses. ***, **, and indicate statistical significance on the 1%, 5%, and 10% levels, respectively.

3.5 | Active portfolio performance grouped by average duration

Finally, I group the MFs and SAs in my sample according to their average effective durations. The results of gross of fee performance regressions within these groups are reported in the four rightmost columns of Table 9. Specifically, these groups are short-term portfolios (≤3 years average duration), medium-term portfolios (3–7 years), long-term portfolios (7–12 years), and very long-term portfolios (>12 years). As durations are not available for all portfolios, this analysis slightly decreases the samples to 98 MFs and 112 SAs. For these groups, the table also reports the value added as well as counts of statistically significant curvature betas.

The largest group in both panels is the medium-term group with 56 MFs and 43 SAs. For SAs, this group is also the one with the worst alpha on average with −0.8193% p.a. gross of fees. For MFs, the worst alpha is displayed for the long-term group with −1.2786% p.a. This group also shows the worst value added with a significant −6.75 million US$ lost to the capital market. Very long-term MFs, on the contrary, show the highest significant value added with 7.63 million US$ extracted from capital markets. This number is topped by very long-term SAs with 38.30 million US$; however, it is not statistically significant due to a very high standard deviation among the portfolios.

The very long-term group is also the one with the largest portfolios with 1224 million US$ for MFs and 4775.92 million US$ for SAs. In combination with the insignificant average alpha, this explains the relatively high value added. In addition, the correlation between alpha and size is also very high at 20%. Thus, I conclude that very long-term portfolios are managed by highly specialized and capable portfolio managers, while medium-term portfolios are managed by lesser managers.

Finally, looking at the counts of significant curvature betas shows that curvature is a very important factor, as in each group more than half of the portfolios have a significant curvature beta. Further, the patterns are also consistent with economic intuition and the results displayed in Figure 2, as most curvature betas are positive in the groups of short-term to long-term portfolios. However, for very long-term portfolios, the curvature betas are predominantly negative, as expected.

4 | CONCLUSION

This study introduces a three-factor interest rate risk model that enhances the measurement of active bond fund performance by accounting for the nonlinear relationship between risk exposure and expected bond returns. By incorporating level, slope, and curvature factors, the model provides a more accurate assessment of passive style returns and active alpha. Empirical analyses on passive Treasury indexes and actively managed bond portfolios demonstrate the model's robustness, offering valuable insights into bond fund performance evaluation.

The findings have significant implications for both policy and practice. For regulators and policymakers, the study highlights the need for more precise risk-adjusted performance metrics in the oversight of fixed-income funds. Current methodologies that fail to capture return curve nonlinearity may misrepresent fund performance, leading to inefficient capital allocation and misinformed investor decisions. Implementing enhanced performance models in regulatory reporting could improve transparency and accountability in the bond fund industry.

For investment managers and institutional investors, the three-factor model provides a more reliable framework for evaluating bond portfolio performance. Managers can better distinguish between skill-based alpha generation and structural miscalculations caused by improper risk-adjustment techniques. Additionally, the study underscores the economic relevance of separate accounts, which, despite their size and significance, have received limited academic attention. Understanding their superior value-added performance relative to mutual funds can help institutional investors optimize their fixed-income allocations.

Overall, this research contributes to the ongoing discourse on bond fund performance by offering a refined, implementable model for more accurate return attribution. Future research could further explore its applicability across different bond market segments, asset classes, and economic conditions. Specific challenges, for example, when assessing the performance of corporate bond portfolios, include that corporate bonds are very heterogeneous due to the multitude of issuers from different industries. Moreover, they are distinctively less liquid than government bonds, especially due to generally lower nominal volumes and long-term investors, like insurance or pension funds, holding corporate bonds to maturity. Thus, a corporate bond illiquidity factor needs to be considered as well as strategies to account for stale prices, like the use of lagged factors even at the monthly return frequency. Further, the government bond portfolios assessed in this study hold corporate bonds of high credit quality as evidenced by the high average credit rating of AA. Thus, a single default factor representing the slope between BBB- and AAA-rated corporate bonds may suffice. However, it may be assumed that the default premium is no linear function of credit ratings, especially in the segment of high default risk, and might also require a multipolynomial function to fully capture expected returns. Thus, a similar approach to the one I propose for interest rate risk may be necessary to construct an additional curvature factor of default risk.

ACKNOWLEDGMENTS

I thank Rainer Baule, Henk Berkman, Stephen Brown, Oliver Entrop, Iraj Fooladi, Daniel Giamouridis, Alexander Hillert, Christian Koziol, Markus Natter, David Maslar, Steffen Mayer, Fabio Moneta, Sebastian Müller, Andreas Neuhierl, Nial O'Sullivan, Andreas Rathgeber, Hendrik Scholz, Stijn van Nieuwerburgh, Andreas Walter, Florian Weigert, Marco Wilkens, and David Yermack for very helpful comments and suggestions. I am responsible for all remaining errors. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from Morningstar, FactSet. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the author(s) with the permission of Morningstar, FactSet.

ORCID

Martin Rohleder https://orcid.org/0000-0003-3067-8640

ENDNOTES

[1] For example, Cornell and Green (1991), Blake et al. (1993), Elton et al. (1995), Bessembinder et al. (2009), Amihud and Goyenko (2013).

[2] Portfolio returns are affected by (1) predictions of securities' idiosyncratic deviations from expected returns (active selection), the harvesting of systematic risk premia with long-term exposures (passive style return, e.g., “reaching for yield,” Choi & Kronlund, 2018), and short-term deviations from long-term exposures (active timing). As Bunnenberg et al. (2019) assert that constant-beta measures like Jensen's (1968) alpha approximately capture the sum of selection and timing as “total active performance,” I concentrate on alpha.

[3] Among others, Blake et al. (1993) and Frazzini and Pedersen (2014) show that beta and duration are alternative measures of portfolios' interest rate risk exposure.

[4] Other approaches like, for example, that of Nelson and Siegel (1987) and Brooks and Moskowitz (2017) define the long end of the yield curve as its level. However, this long end is undefined (infinite) and thus not investable.

[5] The factor construction follows, for example, Wilkens (1994), Chen et al. (2010) and Brooks and Moskowitz (2017). However, their yield curve parameters are based on yields and do not represent the returns of diversified passive alternatives.

[6] See Table A1 in Appendix for an overview of the most important bond and bond fund performance studies of the past 30 years and the models they use.

[7] The risk-free rate (1-month T-Bill) is obtained from the Kenneth R. French data library.

[8] For example, the ICE 27.5+ years Off-the-Run index has a curvature beta of −0.2929**, the ICE 5–10 years index has a positive beta of 0.9283***, and the ICE 0–3 years index has a curvature beta of −0.0106***.

[9] Alternative analyses using 24 and 36-month windows show similar results (available upon request).

[10] Respective figures and tables are available upon request.

[11] Alternative regressions using 24 and 36-month windows show similar results (available upon request).

[12] Daily treasury yields are obtained from the US Department of the Treasury (available since 1990) and transformed to the monthly frequency using end-of-month values.

[13] Consider for example in Figure 1 an intermediate-term index and a very-long-term index. Not considering the curvature of the yield curve in the performance model, the expected alpha of the intermediate-term index is positive, that of the very-long-term index is negative.

[14] Index fixed effects regressions yield similar results and are available upon request.

[15] Alternatively, I use 24- and 36-month windows, which yield similar results and are available upon request.

[16] Obtained from the Kenneth R. French data library.

[17] As of April 2022, ICE fixed income indexes consider transaction costs.

REFERENCES

Adam, T., & Guettler, A. (2015). Pitfalls and perils of financial innovation: The use of CDS by corporate bond funds. *Journal of Banking and Finance*, *55*, 204–214.

Amihud, Y., & Goyenko, R. (2013). Mutual fund R^2 as predictor of performance. *Review of Financial Studies*, *26*(3), 667–694.

Ammann, M., Kind, A., & Seiz, R. (2010). What drives the performance of convertible-bond funds? *Journal of Banking & Finance*, *34*(11), 2600–2613.

Bali, T., Bai, J., & Wen, Q. (2019). Common risk factors in the cross-section of corporate bond returns. *Journal of Financial Economics*, *131*, 619–642.

Berk, J. B., & van Binsbergen, J. H. (2015). Measuring skill in the mutual fund industry. *Journal of Financial Economics*, *118*, 1–20.

Bessembinder, H., Kahle, K. M., Maxwell, W. F., & Xu, D. (2009). Measuring abnormal bond performance. *Review of Financial Studies*, *22*(10), 219–258.

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

Brooks, J., & Moskowitz, T. J. (2017). *Yield curve premia*. Working paper, AQR Capital Management, Yale School of Management.

Bunnenberg, S., Rohleder, M., Scholz, H., & Wilkens, M. (2019). Jensen's alpha and the market timing puzzle. *Review of Financial Economics*, *37*(2), 234–255.

Chen, H., Cohen, L., & Gurun, U. G. (2021). Don't take their word for it: The misspecification of bond mutual funds. *Journal of Finance*, *76*(4), 1699–1730.

Chen, Y., Ferson, W., & Peters, H. (2010). Measuring the timing ability and performance of bond mutual funds. *Journal of Financial Economics*, *98*, 72–89.

Chen, Y., & Qin, N. (2017). The behavior of investor flows in corporate bond mutual funds. *Management Science*, *63*(5), 1365–1381.

Choi, J., & Kronlund, M. (2018). Reaching for yield in corporate bond mutual funds. *Review of Financial Studies*, *31*(5), 1930–1965.

Cici, G., & Gibson, S. (2012). The performance of corporate bond mutual funds: Evidence based on security-level holdings. *Journal of Financial and Quantitative Analysis*, *47*(1), 159–178.

Cici, G., Gibson, S., & Merrick, J. J. (2011). Missing the Marks? Dispersion in corporate bond valuations across mutual funds. *Journal of Financial Economics*, *101*, 206–226.

Comer, G., Larrymore, N., & Rodriguez, J. (2009). Controlling for fixed-income exposure in portfolio evaluation: Evidence from hybrid mutual funds. *Review of Financial Studies*, *22*(2), 481–507.

Cornell, G., & Green, K. (1991). The investment performance of low-grade bond funds. *Journal of Finance*, *46*(1), 29–48.

Cremers, M., Petajisto, A., & Zitzewitz, R. (2012). Should benchmark indexes have alpha? Revisiting performance evaluation. *Critical Finance Review*, *2*, 1–48.

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

Elton, E. J., Gruber, M. J., & Blake, C. R. (2014). The performance of separate accounts and collective investment trusts. *Review of Finance*, *18*(5), 1717–1742.
Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.
Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, *81*(3), 607–636.
Ferson, W., Hendry, T. R., & Kisgen, D. J. (2006). Evaluating government bond fund performance with stochastic discount factors. *Review of Financial Studies*, *16*(2), 423–455.
Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, *111*(1), 25.
Goldstein, I., Jiang, H., & Ng, D. T. (2017). Investor flows and fragility in corporate bond funds. *Journal of Financial Economics*, *126*(3), 592–613.
Huang, J.-Z., & Wang, Y. (2014). Timing ability of government bond fund managers: Evidence from portfolio holdings. *Management Science*, *60*(8), 2091–2109.
Huij, J., & Derwall, J. (2008). 'Hot hands' in bond funds. *Journal of Banking & Finance*, *32*(4), 559–572.
Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *Journal of Finance*, *23*(2), 389–416.
Litterman, R., & Scheinkman, J. (1991). Common factors affecting bond returns. *The Journal of Fixed Income*, *1*(1), 54–61.
Moneta, F. (2015). Measuring bond mutual fund performance with portfolio characteristics. *Journal of Empirical Finance*, *33*, 223–242.
Natter, M., Rohleder, M., & Wilkens, M. (2021). Maturity-matched bond fund performance. *Financial Analysts Journal*, *77*(2), 83–96.
Nelson, C. R., & Siegel, A. F. (1987). Parsimonious modeling of yield curves. *Journal of Business*, *60*(4), 473–489.
Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708.
Pelhares, D., & Richardson, S. A. (2020). Looking under the Hood of active credit managers. *Financial Analysts Journal*, *76*(2), 82–102.
Svensson, L. E. O. (1994). *Estimating and interpreting forward interest rates: Sweden 1992–1994*. Technical report 114, International Monetary Fund, Washington, DC.
Wilkens, M. (1994). Risiko-management mit zins-futures in Banken. In *Neue betriebswirtschaftliche studienbücher nr. 8*. Verlag Otto Schwartz & Co.
Zhao, X. (2005). Determinants of flows into retail bond funds. *Financial Analysts Journal*, *61*(4), 47–59.

How to cite this article: Rohleder, M. (2025). Measuring the performance of government bond portfolios with index-based level, slope, and curvature factors. *Review of Financial Economics*, *00*, 1–17. https://doi.org/10.1002/rfe.70024

REVIEW of FINANCIAL ECONOMICS — WILEY

APPENDIX

Table A1 provides an overview of the bond and bond fund performance literature of the past 30 years and the indexes/factors they use to measure performance. The overview was adopted from Natter et al. (2021) with permission of the authors.

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 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
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 characteristics-based benchmarks on duration
Cici and Gibson	2012	JFQA	Holdings-based, matching bonds and characteristics-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	JBF	Broad bond index as in Elton et al. (1995)
Moneta	2015	JEF	Holdings-based, matching bonds and characteristics-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)
Pelhares and Richardson	2020	FAJ	Medium-Duration Treasury Index
Natter, Rohleder, and Wilkens	2021	FAJ	Single Treasury index, matched to the fund's duration

Abbreviations: FAJ, Financial Analysts Journal; JB, Journal of Business; JBF, Journal of Banking and Finance; JEF, Journal of Empirical Finance; JF, Journal of Finance; JFE, Journal of Financial Economics; JFQA, Journal of Financial and Quantitative Analysis; MS, Management Science; RFS, Review of Financial Studies.