On performance effects of mutual fund derivative use and other sources of non-linearity

Kumulative Dissertation

der Wirtschaftswissenschaftlichen Fakultät

der Universität Augsburg

zur Erlangung des akademischen Grades eines Doktors

der Wirtschaftswissenschaften

(Dr. rer. pol.)

vorgelegt von

Herrn Markus Natter, M. Sc.

Erstgutachter: Prof. Dr. Marco Wilkens

Zweitgutachter: Prof. Dr. Andreas Rathgeber

Drittgutachter: Prof. Dr. Wolfgang Schultze

Abgabedatum: 24.05.2017

Datum der mündlichen Prüfung: 21.07.2017

Contents

1 Introduction

1.1 Motivation

Recent official reports in the US illustrate that mutual funds' investment practices become more and more subject to financial regulators' concerns. In 2016, the Financial Stability Report (FSR) as well as documents published by Financial Stability Oversight Council (FSOC) raise concerns over the role of mutual funds for the stability of the financial system. Most concerns encompass liquidity and redemption risk but also risks arising from financial leverage through investments in derivatives and other leveraged instruments attract attention (OFR, 2016; FSOC, 2016a, and FSOC, 2016b). As a comprehensible consequence, the SEC aims to tighten the regulation on mutual funds' use of derivatives. Thus, the regulator invites researchers to comment on mutual funds' derivatives use and to review existing rules.[1] The idea of this procedure is to protect private as well as institutional investors and to mitigate any adversarial effects of leveraged investments for the financial system as a whole.

Mutual fund performance and risk have been subject to scientific research for a long time. Jensen (1968), Fama and French (1993) as well as Carhart (1997), among others, develop models that have now become state-of-the-art in research. Other important topics in this wide research area are, for example, measuring mutual fund timing (e.g. Treynor and Mazuy, 1966; Henriksson and Merton, 1981; Busse, 1999; Becker et al., 1999; Jiang et al., 2007; Elton et al., 2012). Others investigate how certain fund characteristics, for example size or cash, are associated with performance and risk (Chen et al. 2004; Pastor et al., 2015; Simutin, 2014). With regard to regulators' concerns, researchers examine the effects of flows into and out of mutual funds on performance as documented by Rakowski (2010), Coval and Stafford (2007)

[1] https://www.sec.gov/rules/concept/2011/ic-29776.pdf.

or Rohleder et al. (2017). The tenor of this stream of literature is that there is a negative relation between both inflows and outflows and the profitability of mutual funds.

With respect to leverage generating investment practices and their impact on fund performance and risk, however, there is no scientific consensus. In the US, the topic becomes particular important after the abolishment of the “short-short” rule with the “American taxpayer relief act” in 1997. With the repealing of this statute, the use of derivative securities and other leverage generating investment practices has grown but only a few studies address this topic. Lynch-Koski and Pontiff (1999) are the first to examine derivatives use by equity mutual funds in the US. They find no clear relation between derivatives and mutual fund performance or risk. Chen et al. (2013) document superior manager skill for short selling mutual funds. Evans et al. (2015) find underperformance for security lenders. Further work on mutual funds derivative use is provided by Cao et al. (2011), Cici and Palacios (2015), and Rohleder et al. (2017) for equity funds. The literature regarding this topic in connection with bond funds is much smaller as there are only two studies published (Deli and Varma, 2002 and Adam and Guettler, 2015).

One possible reason for the disaccord of research on mutual fund derivative use could be that performance and risk of funds using leverage generating instruments are difficult to measure. Many studies address problems and hurdles, which can arise from such investment practices (see e.g. Leland, 1999; Lhabitant, 2000, Goetzmann et al., 2007).

As a result, this dissertation aims to close two research gaps. First, it aims at improving the understanding of relations between both equity and bond mutual funds and complex investment practices. The second goal is to contribute to the literature on biases in performance and risk measures for mutual funds. Chapter II thus answers the question if option use by US domestic equity mutual funds is beneficial or harmful for investors and whether SEC’s worries about the use of these instruments are justified. Chapter III examines relations between the use of a broader range of complex investment practices by bond mutual funds. The focus of Chapter IV lies on the second main goal of this dissertation and reveals a systematic bias in bond fund performance measurement that arises from the non-linear relationship between changes in the term structure and bond returns, which has been neglected until now. Chapter V reviews the performance of option strategy benchmark indices provided by the Chicago Board of Options Exchange (CBOE) and uncovers the complexity of different approaches to measuring performance and risk of portfolios containing options. The last chapter VI sums up the results of this dissertation and gives an understanding of research ideas that might be relevant in the future based on the insights provided in this dissertation. The rest of Chapter I ends with the following brief summaries of the research articles provided in this dissertation.

1.2 Overview over papers included

Paper title	Co-authors	Published?	Journal	Date
The benefits of option use by mutual funds	Martin Rohleder Dominik Schulte Marco Wilkens	Yes	Journal of Financial Intermediation (A), Vol. 26, pp. 142–168[2]	2016
Bond mutual funds and complex investments	Martin Rohleder Dominik Schulte Marco Wilkens	Yes	Journal of Asset Management (B), forthcoming[3]	2017
Duration-adjusted bond fund performance	Martin Rohleder Marco Wilkens	No[4]	WP, University of Augsburg[5]	2017
Option-based benchmark indices – A review of performance and (in)appropriate measures	–	Yes	Journal of Futures Markets (B), forthcoming[6]	2017

1.2.1 Article I – The benefits of option use by mutual funds

The first article of this dissertation focuses on the performance and risk of mutual funds investing in options. Analyzing a sample of 2,576 actively managed U.S. domestic equity funds

[2] doi: 10.1016/j.jfi.2016.01.002.
[3] doi:10.1057/s41260-017-0046-7.
[4] Accepted at the 4th European Retail Investment Conference 2017, Stuttgart, Germany, the 2017 FMA European Conference, Lisbon, Portugal, the World Finance Conference, Sardinia, Italy, and the 79th Annual VHB Meeting, St. Gallen, Switzerland for presentation.
[5] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2877630.
[6] doi: 10.1002/fut.21865.

and a period from 1999 to the end of 2014, we find that option users among equity funds generate higher risk-adjusted performance as well as lower systematic risk compared to their non-using peers. Moreover, we show that observed relations of option use and performance as well as risk, respectively, are directly attributable to the employment of options and are not driven by other fund characteristics. Systematically lower systematic risk is mainly a result of hedging efforts, as mutual funds mainly employ protective put and covered call strategies. We strengthen our results by controlling for non-linearities in fund returns and skewed return distributions with models by Leland (1999), Bawa and Lindenberg (1977) as well as a novel investable option-factor approach (IOS) we develop. Various robustness tests indicate that our findings are not a result of known anomalies such as the low beta anomaly (Frazzini and Pedersen, 2014) or the low volatility anomaly (e.g. Baker et al, 2011) and are devoid of endogeneity concerns. Our overall conclusion is that the use of options by mutual funds is beneficial for investors and is not harmful for the stability of the financial system as a whole.

1.2.2 Article II – Bond mutual funds and complex investments

After dealing with the use of options by equity mutual funds, this paper examines the use of complex investments of active US bond mutual funds. We define complex investments as all investment practices that are not coercively common among mutual funds such as derivatives and practices that generate economic leverage, e.g. short selling or margin purchases. The majority of previous research papers analyze complex investment use of equity mutual funds, whereas only very few studies examine funds investing in fixed income securities. Looking at a sample consisting of 997 active bond mutual funds from 1990 to 2014, we find that complex investments are much more common among bond funds than among equity funds. However, it seems that most complex investment practices are unrelated to fund performance and risk. Bond funds investing in interest rate futures significantly underperform their non-using peers by more than 50 basis point p.a., though. Furthermore, these futures users exhibit higher exposure to the term-factor, i.e. they might employ these derivatives to speculate on interest rate changes.

1.2.3 Article III – Duration-adjusted bond fund performance

The second main topic in this dissertation is the methods used for performance measurement. In particular, this article focuses on systematic and mechanical biases arising when measuring the performance of fixed income mutual funds. All previous studies analyzing bond funds neglect this topic and thus, the topic is of vital importance. The main finding of this study is that the prevailing methods for performance measurement and their implicit assumption of a linear relationship between return and time to maturity respectively duration lead to severe and systematic biases. At first, we provide theoretical evidence and test our theory on a set of passive indices, where we explain the excess return of an index with the excess return of another index that has a different effective duration. We can show that the more the two durations differ the more severe the bias is in performance measures. In addition, we translate this methodology to a sample consisting of 127 actively managed US domestic government bond funds and 291 active US domestic corporate bond funds in the period from 1990 to 2014. The pattern observed with passive bond indices can also be found among bond funds. Since the duration of mutual funds is lower than for broad bond indices used in previous research, we can show that performance in existing studies is systematically overestimated. Our solution for this duration bias in performance measures is to choose a best-fit-benchmark that matches the duration of the respective fund as good as possible. In further analyses, we can show that results of previous work are at least partly driven by the duration bias.

1.2.4 Article IV – Option-based benchmark indices – A review of performance and (in)appropriate measures.

The Chicago Board of Options Exchange (CBOE) has invented a variety of passive option strategy benchmark indices to provide sufficient benchmarks for investors trading options.

Many studies attest relatively high returns and lower risk in terms of volatility for these indices and thus, find significant outperformance on a risk-adjusted basis. Measuring the performance of portfolios containing options, however, is highly complex. Options added to portfolios may generate asymmetric return distributions due to their non-linear payoff structures (e.g. Leland, 1999; Lhabitant, 2000). Skewed returns may lead to positively or negatively biased performance measures. Further biases might arise from time-varying betas, which are grounded in the self-financing replicating portfolio of options (Black and Scholes, 1973). This study employs different approaches to account for these pitfalls by controlling for higher moments, allowing time-varying betas and, moreover, develops a novel option straddle-factor model inspired by the hedge fund related work of Agarwal and Naik (2004). Although results do not change remarkably by controlling for these shortcomings, it is nevertheless possible that performance measures are biased. This is the first study that examines performance and risk for many different option strategy benchmark indices simultaneously. An interesting finding of this article is that outperformance found by previous studies is driven by past times and limited sample periods. A possible reason for the diminishing performance over time might be falsely priced options as documented by Constantinides et al. (2009) as well as Chambers et al (2014). In addition to analyses on raw data of option strategies, the profitability of direct investments in one of the strategy indices via an Exchange Traded Fund and an Exchange Traded Note is examined. The conclusion that the reader can draw from this analysis is that potential benefits of option strategies are consumed by the cost investment providers charge, as the investment products underperform their underlying by up to almost one percentage point p.a.

Bibliography

Adam, T. and 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.

Agarwal, V. and Naik, N. Y. (2004) Risks and portfolio decisions involving hedge funds. Review of Financial Studies 17 (1): 63–98.

Baker, M., Bradley, B. and Wurgler, J. (2011) Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly. Financial Analysts Journal 67 (1): 40–54.

Bawa, V. S. and Lindenberg, E. B. (1977) Capital market equilibrium in a mean-lower partial moment framework. Journal of Financial Economics 5 (2): 189–200.

Becker, C., Ferson, W., Myers, D.H. and Schill, M. J. (1999) Conditional market timing with benchmark investors. Journal of Financial Economics 52 (1): 119–148.

Black, F. and Scholes, M. (1973) The pricing of options and corporate liabilities. Journal of Political Economy 81 (3): 637–654.

Busse, J. A. (1999) Volatility timing in mutual funds: evidence from daily returns. Review of Financial Studies 12 (5): 1009–1041.

Cao, C., Ghysels, E. and Hatheway, F. (2011) Derivatives do affect mutual fund returns: Evidence from the financial crisis of 1998. Journal of Futures Markets 31 (7): 629–658.

Carhart, M. M. (1997) On persistence in mutual fund performance. Journal of Finance 52 (1): 57–82.

Chambers, D. R., Foy, M., Liebner, J. and Lu, Q. (2014) Index option returns: Still puzzling. Review of Financial Studies 27 (6): 1915–1928.

Chen, J., Hong, H., Huang, M. and Kubik, J. D. (2004) Does Fund Size Erode Mutual Fund Performance? The Role of Liquidity and Organization. American Economic Review 94 (5): 1276–1302.

Chen, H., Desai, H. and Krishnamurthy, S. (2013) A first look at mutual funds that use short sales. Journal of Financial and Quantitative Analysis 48 (3): 761–787.

Cici, G. and Palacios, L.-F. (2015) On the use of options by mutual funds: Do they know what they are doing? Journal of Banking and Finance 50: 157–168.

Constantinides, G. M., Jackwerth, J. C. and Perrakis, S. (2009) Mispricing of S&P 500 index options. Review of Financial Studies 22 (3): 1247–1277.

Coval, J. and Stafford, E. (2007) Asset fire sales (and purchases) in equity markets. Journal of Financial Economics 86 (2): 479–512.

Deli, D. N. and Varma, R. (2002) Contracting in the investment management industry: evidence from mutual funds. Journal of Financial Economics 63 (1): 79–98.

Elton, E. J., Gruber, M. J. and Blake, C. R. (2012) An examination of mutual fund timing ability using monthly holdings data. Review of Finance 16 (3): 619–645.

Evans, R. B., Ferreira, M. A. and Porras Prado, M. (forthcoming) Fund performance and equity lending: Why lend what you can sell? Review of Finance, in press. (doi: 10.1093/rof/rfw059)

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

Frazzini, A. and Pedersen, L. H. (2014) Betting against beta. Journal of Financial Economics 111 (1): 1–25.

Financial Stability Oversight Council (2016a) Annual Report, Washington: FSOC, June 21, 2016.
www.treasury.gov/initiatives/fsoc/studies-reports/Documents/FSOC%202016%20Annual%20Report.pdf
Accessed 29 March 2017.

Financial Stability Oversight Council (2016b) Update on Review of asset management products and activities, Washington: FSOC, April 18, 2016.

http://www.treasury.gov/initiatives/fsoc/news/Documents/FSOC%20Update%20on%20Review%20of%20Asset%20Management%20Products%20and%20Activities.pdf Accessed 29 March 2017.

Goetzmann, W., Ingersoll, J., Spiegel, M. and Welch, I. (2007) Portfolio performance manipulation and manipulation-proof performance measures. Review of Financial Studies 20 (5): 1503–1546.

Henriksson, R. D. and Merton, R. C. (1981) on market timing and investment performance. ii. statistical procedures for evaluating forecasting skills. Journal of Business 54 (4): 513–533.

Jensen, M. C. (1968) The performance of mutual funds in the period 1945-1964. Journal of Finance 23 (2): 389–416.

Leland, H. E. (1999) Beyond Mean-Variance: Performance measurement in a nonsymmetrical world. Financial Analysts Journal 55 (1): 27–36.

Jiang, G. J., Yao, T. and Yu, T. (2007) Do mutual funds time the market? Evidence from portfolio holdings. Journal of Financial Economics 86 (3): 724–758.

Lhabitant, F. S. (2000) Derivatives in portfolio management: Why beating the market is easy. Derivatives Quaterly 7 (2): 39–45.

Lynch-Koski, J. and Pontiff, J. (1999) How are derivatives used? Evidence from the mutual fund industry. Journal of Finance 54 (2): 791–816.

Office of Financial Research (2016): Financial Stability Report, Washington: OFR, December 13, 2016. https://www.financialresearch.gov/financial-stability-reports/files/OFR_2016_Financial-Stability-Report.pdf Accessed 29 March 2017.

Pastor, L., Stambaugh, R. F. and Taylor, L. A. (2015) Scale and skill in active management. Journal of Financial Economics 116 (1): 23–45.

Rakowski, D. (2010) Fund flow volatility and performance. Journal of Financial and Quantitative Analysis 45 (1): 223–237.

Rohleder, M., Schulte, D. and Wilkens, M. (2017) Management of flow risk in mutual funds. Review of Quantitative Finance & Accouting 48 (1): 31–56.

Simutin, M. (2014) Cash holdings and mutual fund performance. Review of Finance 18 (4): 1425–1464.

Treynor, J. and Mazuy, K. (1966) Can mutual funds outguess the market? Harvard Business Review 44 (4): 131 –136.

2 Article I: The Benefits of Option Use by Mutual Funds

Markus Natter, Martin Rohleder, Dominik Schulte, and Marco Wilkens: The Benefits of Option Use by Mutual Funds. Journal of Financial Intermediation (2016) 26: 142–168
(doi: 10.1016/j.jfi.2016.01.002).

VHB-Jourqual 3: A

3 Article II: Bond Mutual Funds and Complex Investments

Markus Natter, Martin Rohleder, Dominik Schulte, Marco Wilkens: Bond Mutual Funds and Complex Investments. Journal of Asset Management (2017, forthcoming).

(doi: 10.1057/s41260-017-0046-7)

VHB-Jourqual 3: B

4 Article III: Duration-adjusted bond fund performance

Markus Natter[a], Martin Rohleder[b], and Marco Wilkens[c]

Working Paper, University of Augsburg

This draft: May 5, 2017

Abstract. We uncover a previously neglected mechanical bias in bond fund performance due to the use of benchmarks with non-matching durations. We show that the *duration bias* is caused by the non-linear reaction of bonds with different durations to interest rate changes. We find empirically that the normal use of a broad bond index in previous research leads to a significant overestimation of average bond fund performance and spurious findings of performance persistence. The key takeaway of our research is thus that bond fund performance should be duration-adjusted by choosing for each fund the benchmark index which best matches its duration.

Keywords: Bond funds, performance, duration, bond index, persistence

JEL Codes: G20, G11, G23

4.1 Introduction

In this paper, we uncover a previously neglected mechanical bias in measured bond fund performance caused by the use of benchmark indices which do not match the durations of the funds. This "duration bias" emerges because the non-linear reaction of bonds with different durations to changes of the term structure is not adequately recognized by the usual regression based performance models. This finding is of vital importance because the usual practice of using a broad market index leads to a significant overestimation of average bond fund performance as the durations of bond funds are shorter on average than the durations of those broad market indices. Further, this mechanical bias explains at least partly the consistent finding of performance persistence or "hot hands" in bond funds. The key takeaway of our paper is thus that bond fund performance should be measured in a consistent, duration-adjusted way by choosing for each fund the benchmark index which best matches the duration of the fund.

Ever since Jensen (1968), the dominant approach to measuring the risk-adjusted performance of mutual funds has been to relate fund returns linearly to the returns of one or multiple passive benchmark factors via OLS regression and interpret the constant α as the value added by active management. For equity funds, this approach is theoretically grounded on the CAPM by Sharpe (1964), Lintner (1965) and Mossin (1966) where the expected return of a stock is linearly linked to the expected return of the efficient equity market portfolio. An equity fund's systematic market risk is thus measured by its sensitivity *beta* with which any security can be linearly scaled up and down the equity security market line (*SML*).

For bonds, already early work by Macaulay (1938) shows that the duration D may be interpreted as a measure of a bond's sensitivity towards systematic term-risk, i.e. the risk of an instantaneous bond price change $dP_{i,t}/P_{i,t}$ due to a change of the interest rate r at time t (e.g., Hopewell and Kaufman, 1973).

$$\frac{dP_{i,t}}{P_{i,t}} = -D_{i,t}\ \frac{dr_{i,t}}{(1+r)}$$

Combining both approaches – the linear market model and the interpretation of duration as a bond's sensitivity to term-risk – Boquist et al. (1975) argue for default-free bonds that *beta* is a linear function of duration. Jarrow (1978) extends this relation to bonds with default risk.[7]

$$\beta_{i,t} = -D_{i,t}\,\frac{Cov(dr_{i,t}, R_{m,t})}{Var(R_{m,t})}$$

Consequentially, when bond fund performance research took off in the early 1990s with the studies by Cornell and Green (1991), Blake et al. (1993) and Elton et al. (1995), these studies adopted the linear market model approach from equity fund studies to explain the returns of bond funds with those of broad (Treasury) bond market indices. The implicit assumption is that a bond fund's systematic term risk can be linearly scaled up and down the "bond *SML*" using the *beta*.[8] Distinct advantages of adopting this already established methodology are that the results may be conveniently interpreted in the usual way and that the regression models allow a straightforward incorporation of further relevant risk factors such as, e.g., default risk, option related risk, equity-related risk or illiquidity risk.

However, the underlying assumption of linearity in term-risk may be only made for a parallel shift of a flat yield curve while the shape of the term structure, as well as changes thereof, are usually more complicated than that (e.g., Cox et al., 1979). In this context, Dietz et al. (1981) show via cross-sectional regressions that the return duration relationship in the US government bond market is, in fact, non-linear. To confirm and illustrate this non-linearity, we present different example term structure scenarios from a sensitivity analysis which is based on

[7] The parameter *D* in Jarrow (1978) is an approximation of Macaulay's duration.

[8] More recent examples of studies using this approach are, e.g., Huij and Derwall (2008), Gutierrez et al. (2009) Amihud and Goyenko (2013), and Chen and Qin (2016).

the proposition by Litterman and Scheinkman (1991) that level, slope and curvature sufficiently describe the shape of the term structure.

As a consequence, our primary research hypothesis is, that the usual approach of relating the returns of all bond funds linearly to those of the same broad (Treasury) bond market index to measure the term risk of funds will lead to a systematic bias in measured performance, if the average duration of the index does not match the average durations of the funds. We term this bias the "duration bias" and propose the usage of a "duration-adjusted benchmark model" which merely identifies and uses for each bond fund the one index from a set of alternative indices with the closest proximity in duration. This approach solves the duration bias problem while at the same time maintaining the above advantages of the usual regression models.

To test our hypothesis of a significant duration bias we employ a broad range of empirical tests. First, using ten US Treasury indices with focuses on different maturity ranges in the period from 1990 to 2014, we find statistically significant and positive duration biases if the benchmark indices' durations are too long and statistically significant and negative duration biases if the benchmark indices' durations are too short. Second, using monthly information on the average durations of 127 active US domestic government bond funds and 291 corporate bond funds, as well as information on the monthly durations of the ten US Treasury indices, we show a similar duration bias pattern in bond fund performance. Third, we show that this overall pattern is especially pronounced during months with a widening term-spread and reversed during months with a contracting term-spread. The reversal behavior is especially relevant during the build-up periods to the two economic crises during our sample period, consistent with the usual interpretation of an inverted term structure as a predictor of an upcoming recession (e.g., Estrella and Hardouvelis, 1991).

Important implications of these findings are, first, that the average duration of the government and corporate bond funds in both our samples is approximately four years while the duration of the broad Treasury index is close to 5.5 years. This disparity suggests a significant and systematic overestimation of the average measured bond fund performance in previous studies usually using a broad (Treasury) bond market index. Second, via panel regressions, we show that alpha performance using the usual broad index is significantly driven by term structure changes and thus a biased measure of managers' selection performance. Our duration-adjusted alpha, however, is unrelated to term structure changes and hence unbiased. Third, we find that the consistent finding of performance persistence in bond funds is at least partly driven by the systematic overestimation of shorter-term funds and the systematic underestimation of longer-term funds due to the duration bias and not due to "hot hands." This is especially true for government bond funds where empirical tests show no persistence after controlling for the duration-adjusted benchmark. Fourth, by analyzing the names of bond funds' self-stated prospectus benchmark indices, we find that the majority of funds provide dependable information regarding the average maturity range of their portfolio holdings. This raises the question why previous research has not used this information in bond fund performance measurement.

The paper proceeds as follows: Section 4.2 presents a sensitivity analysis to illustrate the non-linear relationship between the returns of bonds with different durations. Section 4.3 presents a bond index based analysis of the duration bias. Section 4.4 tests the duration bias in government and corporate bond funds. Section 4.5 tests if our duration-adjusted model yields unbiased estimates of fund performance. Section 4.6 presents further implications of our findings for prior and future bond fund research, specifically on the persistence in bond fund performance and on the information content of bond funds' prospectus benchmarks. Section 4.7 concludes.

4.2 Sensitivity analysis

This section shows a sensitivity analysis based on the proposition by, e.g., Litterman and Scheinkman (1991) that three common factors – economically interpreted as level L, slope S and curvature C – are sufficient to characterize the shape of the term-structure to illustrate the non-linear relation between the returns of bonds with different durations. For our analysis, we modify the term structure model used by the German central bank (Deutsche Bundesbank, 1991) via canonical variable transformation. As a result, Eq. (1) relates the internal rate of return r_T on bonds with maturity in T to the three common factors which we define as follows: level $L = r_1$, slope $S = (r_{10} - r_1) / 9$ and curvature $C = r_5 - r_1 - 4\,S$.[9]

$$r_T = L + S\,(T - 1) + C\,(0.436542 - 0.436542\,T + 1.70629 \ln T) \tag{1}$$

Compared to other term structure models like, e.g., the Nelson and Siegel (1987) function, Eq. (1) has the advantage that it conveniently allows analyzing isolated changes in all three common factors. This is an important feature for the sensitivity analysis presented in this section as well as in further analyses presented in section 5.

Based on our term structure model, we generate the spot rates of twenty zero-bonds with integer maturities in T = 1–20 years for a basic setup where L = 1.0%, S = 0.5%, and C = 0.75%. Using these spot rates, we calculate the respective zero-bond prices $P_T = 1 / (1 + r_T)^T$. Then, we assume different changes of the term-structure as isolated or combined changes of the level (dL), slope (dS) and curvature (dC) and calculate the instantaneous returns $R_T = dP_T/P_T$ of the twenty zero-bonds. Figure 1 presents six selected example scenarios for term-structure changes (left plots) and the corresponding zero-bond returns (right plots, solid

[9] This modification is based on Wilkens (1994). For details on the derivation, see Appendix A.

lines) as well as the linear approximation of the returns using the zero-bonds with maturities in T = 1 and T = 10 years (right plots, dotted lines).[10]

[Insert Figure 1 here.]

The plots for all example scenarios show that the relation between the returns of bonds with different durations is clearly non-linear so that linear approximations lead to serious over- and underestimations of bond returns. Even for isolated level shifts as presented in Scenario (i), the relation is not perfectly linear and thus may cause a systematic over- or underestimation of bond returns via linear approximation in extreme cases. For some term-structure changes like the one presented in scenario (vi) where a level increase is combined with a decrease in slope and curvature, the return relation for bonds with different durations may take extreme non-linear shapes. This scenario presents a rather usual case because central bank decisions change short-term interest rates very quickly, but transmission to long-term rates takes time. In that scenario, the linear approximation using the 10-year zero-bond leads to a clear overestimation of the returns of all zero-bonds with term-to-maturity of 2–9 years and to a clear underestimation of the returns on all zero-bonds with term-to-maturity of above 10 years. Thus, we consider our hypothesis warranted that the use of bond market indices which do not match the duration of the analyzed bond fund might lead to a significant duration-bias.

4.3 Treasury bond index analysis

4.3.1 Data

Before directly testing our hypothesis of a significant duration bias using bond funds, we first want to test the hypothesis in an empirical setting where we have complete control over the

[10] The implementation of this simulation in MS Excel is straightforward. However, we are happy to provide further example scenarios upon request. See also spot rates and bond prices for the basic term-structure and scenario (iii) in Appendix B.

alpha generated by active management.[11] To do so, we use ten Bank of America Merrill Lynch (BofAML) US Treasury bond total return indices focusing on different maturity ranges. Using Treasury indices has the further advantage that they allow analyzing term risk effects exclusively without distractions by other bond related risks like, e.g., default risk.[12] We obtain monthly returns of the Treasury indices in the period from 01/1990 to 12/2014 from Morningstar Direct and monthly statistics on their average quality and average effective duration in the period 01/1997 to 12/2014 from the BofAML Global Index System.[13]

Panel A of Table 1 reports summary statistics on quality and durations of the indices. All Treasury indices are rated AAA. The numbers for durations show that the indices cover a broad range of different durations. Moreover, the durations of most of the indices display relatively small variation. Also, the short-term indices are especially distinctive as minimum and maximum durations of adjacent indices show small overlap. The broad Treasury index, which is used in many studies to approximate term-risk, is situated between the indices focusing on maturity ranges from 5–7 and 5–10 years. Panel B shows average durations in non-overlapping 3-year sub-periods. The durations of all indices display an upward trend over time as could be expected by the decreasing nominal interest rates during this period, especially since the financial crisis in 2008.

[Insert Table 1 here.]

[11] Other studies like, e.g., Kosowski et al. (2006) and Fama and French (2010) use different bootstrapping techniques for this purpose.

[12] Alternative analyses using BofAML investment grade corporate bond and government/corporate bond indices with different maturity ranges lead to similar findings. They are available upon request.

[13] http://www.mlindex.ml.com/gispublic/bin/MLIndex.asp#. Data regarding quality and duration of the indices before 1997 is not available via the homepage.

In addition to these indices, we also use a *term* factor – the return difference between a long-term index (10+ years) and a short-term index (1–3 years). Such a factor is used in many studies to capture term risk in the returns of bonds or bond funds (e.g., Fama and French, 1993; Bessembinder et al., 2008; Amihud and Goyenko, 2013).

4.3.2 Duration bias in Treasury bond index performance

4.3.2.1 Overall period

To test the duration bias in Treasury bond index measurement in the overall period from 01/1990 to 12/2014, we run pairwise time series regressions of each index against all other indices following Eq. (2):

$$er_{treasury\ d,t} = \alpha_d + \beta_{d,\ treasury}\ er_t^{treasury\ D \neq d} + \varepsilon_{d,t} \tag{2}$$

where $er_{treasury\ d,t}$ is the return of Treasury index with duration *d* in month *t* in excess of the risk free rate (1m T-bill), $er_t^{treasury\ D \neq d}$ is the excess return of a Treasury index with another duration ($D \neq d$), α_d represents index *d*'s mean abnormal return, $\beta_{d,treasury}$ is the sensitivity of index *d*'s returns to the returns of the Treasury index and ε_{dt} is an error term with $E(\varepsilon_{dt}) = 0$.

Panel A of Table 2 shows regression alphas such that the columns show the dependent index and the rows show the benchmark indices. The results clearly show the existence of a statistically and economically significant duration bias. All alphas below the diagonal are positive indicating that the use of an index with a longer than adequate duration leads to a positive alpha, respectively duration bias, i.e. an overestimation of performance. At the same time, all alphas above the diagonal are negative indicating that the use of an index with shorter than adequate duration leads to a negative duration bias, i.e. an underestimation of performance.

[Insert Table 2 here.]

Further, the two bottom rows of Panel A show regressions against the broad Treasury index and the term factor. As expected, the duration bias is smallest for the index with maturity range 5–10. All shorter indices show a significantly positive alpha and those above show a negative alpha. More astonishingly, however, the term factor leads to a positive, very high and statistically significant alpha independent of the respective index's duration. Thus, we conclude that the use of a long-minus-short term-factor has critical drawbacks compared to using long-only benchmark indices.

4.3.2.2 Distinguishing different market climates

For the overall period from 1990 to 2014, the previous section shows a significantly positive duration bias if short-term indices are benchmarked against long-term indices and vice versa. However, it is ex-ante unclear whether this pattern will hold in the future since our sample period can be characterized by an exceptional overall decrease of the interest rate level as documented by Panel A of Figure 2. To test if the direction of the duration bias depends on the bond market climate, we perform our pairwise index regressions during different level-independent market phases for which different duration biases may be expected. Specifically, we use the slope of the term structure, the spread between the yields of 10 years and 1 year Treasury bonds, to split the sample period into those months exhibiting widening and contracting term-spreads.[14] This separation leads to slightly more observations with a widening than a contracting term-spread as shown in Table 3.

[Insert Figure 2 and Table 3 here.]

During phases of a widening term spread, usually associated with a more positive or "normal" term structure and improving economic conditions (e.g., Estrella and Hardouvelis, 1991), we

[14] Daily yield curve data are obtained from the US Department of the Treasury (https://www.Treasury.gov) and aggregated to the monthly frequency using averages.

expect to find a similar pattern as in the overall period. Scenario (ii) of Figure 1 illustrates this expectation. Here, longer-term bonds underestimate the returns on shorter-term bonds resulting in a positive alpha, and shorter-term bonds overestimate the returns on longer-term bonds resulting in a negative alpha. Consistent with this intuition, Table 3 reports for widening term-spreads that short-term bonds have distinctively higher returns than long-term bonds. Moreover, Panel B of Table 2 shows pairwise index regressions exclusively using months with a widening term spread. The results confirm our expectation by showing the pattern very clearly with positive and significant duration biases below the diagonal and negative, significant and high duration biases above the diagonal.

Conversely, during phases of a contracting term spread, usually associated with a flattening or reversing term structure and worsening economic conditions (e.g., Estrella and Hardouvelis, 1991), we expect to find a reversed pattern. Scenario (iii) of Figure 1 illustrates this expectation. Here, longer-term bonds overestimate the returns on shorter-term bonds resulting in a negative alpha, and shorter-term bonds underestimate the returns on longer-term bonds resulting in a positive alpha. Consistent with this intuition, short-term bonds in Table 3 show distinctively lower returns than long-term bonds. Panel C of Table 2 shows the results of pairwise index regressions which confirm our expectation by showing the reversed pattern very clearly with negative and partly significant duration biases below the diagonal and positive, significant and high duration biases above the diagonal.

Regarding the duration biases with respect to the broad index and the term factor during both market phases, we again find that the broad index behaves like the 5–10 years index during both market phases. For the term factor, Panel B shows positive, very high and statistically significant alphas for all indices during months with a widening spread. During phases of a contracting spread, Panel C still shows positive but insignificant duration biases. This augments our previous conclusion that the term factor is inadequate as a benchmark for bond indices of all durations and that it is also inadequate during both market climates.

4.3.2.3 Rolling window analysis

In the previous section, we show that the duration bias has different signs during different very ideal bond market climates. However, the month-by-month separation of the sub-samples with widening and contracting term spreads is rather artificial as shown in Panel B of Figure 2 and may have only limited real implications. To test if the implications of the previous subsection are also economically relevant, we analyze the development of the duration bias over time using a monthly rolling 5-year window approach.[15]

Figure 3 plots the duration bias over time. In Panel A, the dependent index with the maturity range of 1–3 years is benchmarked against all other indices with longer maturity ranges. As expected from the results displayed in Panel A of Table 2, the figure shows positive duration biases for most of the windows as well as increasing biases with increasing benchmark durations. However, the plots also show that the duration bias is not constant through time. Specifically, during the build-up of the tech bubble in the late 1990s and early 2000s, the duration bias becomes almost zero. Even more dramatically, during the build-up of the financial crisis before 2008, the duration bias becomes negative, displaying the pattern known from Panel C of Table 2. Overall, this is consistent with the interpretation of a flattening or inverting term-structure as a strong predictor of an upcoming economic recession shown in previous studies (e.g., Estrella and Hardouvelis, 1991).

[Insert Figure 3 here.]

Panel B of the figure shows similar plots for the dependent index with a maturity range of 5–10 years benchmarked against all other indices. As expected, all indices with shorter durations

[15] A similar analysis using monthly rolling 3-year windows yields economically similar results.

show on average negative duration biases while those with longer durations show on average positive biases. Consistent with Panel A, the patterns flatten or reverse during the build-ups to the two economic crises occurring during our sample period. Moreover, the broad Treasury index shows an almost zero duration bias over time due to its proximity to the 5–10 years index. Finally, Panel C displays plots for the dependent index with the maturity range of 10+ years benchmarked against all other indices.[16] As expected, the plots show on average negative duration biases against all indices (except for the term factor) as well as the same phases with flattening or reversing biases.

Overall, we can thus conclude that the separation of the overall period into different market climates makes sense not only in theory but also in practice. Deviating from the global pattern shown in Panel A of Table 2, there are phases where a flattened or even reversed duration bias pattern emerges. These phases are economically linked to the build-up of financial and economic crises, consistent with a flattening or inverting of the term-structure.

4.3.3 Elton, Gruber and Nabar (1988)

In this subsection, we use another approach to demonstrate the existence of a significant duration bias in a more theoretical way. Therefore, we refer to the study by Elton et al. (1988) who derive their first return generating process, the single-factor duration model "Dur-1", from the assumption that the price change of a bond due to a random interest rate shock can be expressed as in Eq. (3):

$$\frac{\widetilde{P}_{i,t+1}-\overline{P}_{i,t+1}}{\overline{P}_{i,t+1}} = -D_{i,t+1}\frac{\Delta_{t+1}}{1+r} \tag{3}$$

[16] Further panels for the remaining indices show the expected patterns and are available upon request.

where $\widetilde{P}_{i,t+1}$ is the actual (but unknown) price of bond i at time t+1, $\overline{P}_{i,t+1}$ is the expected price, Δ_{t+1} is the interest rate shock and r is the interest rate before the shock. From this relation, the authors derive the Dur-1 model as:[17]

$$(R_{i,t} - R_{f,t}) = \frac{D_{i,t}}{D_{p,t}}\left(R_{p,t} - R_{f,t}\right) + e_{i,t} \tag{4}$$

where $R_{i,t}$ is the return of bond i in period t, $D_{p,t}$ is the duration of an arbitrary diversified bond portfolio p and $R_{p,t}$ is that portfolio's return.

The critical parameter in Eq. (4) is $e_{i,t}$. It captures the pricing error which, according to the authors, is independent and zero on average. However, given the non-linearity of the price reaction to interest rate shocks for bonds with different durations illustrated by our sensitivity analysis in Section 4.2, we argue that $e_{i,t}$ depends systematically on the discrepancy between D_i and D_p and may, therefore, be non-zero on average. For example, take a look at Scenario (ii) of Figure 1, where portfolio p is the zero-bond with maturity in 10 years, and the risk-free asset is the zero-bond with maturity in 1 year. Now, if $1 < D_{i,t} < 10$ the expected pricing error will be positive and if $D_{i,t} > 10$ the expected pricing error will be negative.

Thus, we apply model Dur-1 as given in Eq. (4) to the ten Treasury indices used in the previous subsections. Therefore, we first calculate the duration ratios $D_{i,t}/D_{p,t}$ for all index pairs $D_{i,t} \neq D_{p,t}$ in every month t. Then, we calculate the monthly expected index excess returns and subtract them from the actual index returns to arrive at time-series of pricing errors for all index pairs. Table 4 shows t-tests against the Null that the pairwise pricing errors are zero on average. Panel A reports the results for the overall period from 1990 to 2014.[18] The patterns displayed

[17] The exact notation used in the paper is $R_i = R_{30} + \frac{D_i}{D_p}\left(R_p - R_{30}\right) + e_i$ where R_{30} is the return on the 30-day T-bill. For details on the derivation of Dur-1, see Elton et al. (1988), p. 130-132.

[18] An additional panel for the period from 1997 to 2014 shows economically similar results and is available upon request.

in the table are similar to those displayed for the pairwise index regressions in Panel A of Table 2 in that the average pricing errors below the diagonal are positive and in many cases statistically significant for benchmark indices with too long durations. Further, the pricing errors are negative on average and in many cases statistically significant above the diagonal for benchmark indices with too short durations.

[Insert Table 4 here.]

Moreover, consistent with the relations illustrated in Figure 1 where the difference between the zero-bond returns and the linear approximation are larger if we approximate a longer index with a too short one, the average pricing errors are larger above the diagonal. This is even more pronounced in Panels B and C which show results for widening and contracting term spreads in the period from 1990 to 2014, thereby further strengthening our hypothesis of a significant duration bias in bond mutual fund performance.

For a possible explanation why the Dur-1 model of Elton et al. (1988) leads to non-zero average pricing errors if $D_i \neq D_p$ we once again refer to the works of Hopewell and Kaufman (1973) and Boquist et al. (1975). They define the instantaneous return of a bond i in reaction to a change in the term-structure as in Eq. (5) where the second term on the right-hand side is termed the “change in yield-to-maturity.”

$$\frac{dP_{i,t}}{P_{i,t}} = -D_{i,t}\frac{dr_{i,t}}{(1+r)} \tag{5}$$

However, as Figure 1 shows, the change in the spot rates of the zero-bonds due to changes of the term-structure clearly depend on the respective term to maturity and are thus specific to each zero-bond (except for isolated level changes). Thus, acknowledging that Δ_{t+1} in Eq. (3) should actually be bond specific $\Delta_{\boldsymbol{i},t+1}$, the derivation of Dur-1 as presented in Elton et al. (1988) does not work out.

4.4 Bond fund analysis

4.4.1 Data description and duration-adjusted benchmark definition

For our bond fund analysis, we obtain mutual fund data from two different databases, CRSP and Morningstar, which we match following Pastor et al. (2015) and Berk and van Binsbergen (2015). From the CRSP database, we select two groups of mutual funds reporting on average more than 50% holdings in either government bonds or corporate bonds and exclude from those all funds having in either database an objective class other than “US government bond” or “US corporate bond,” respectively. Further, we exclude all passive funds by analyzing the funds’ names. We obtain monthly returns, monthly total net assets (TNA) and information on further fund characteristics like, e.g., turnover, expenses, and age from CRSP. From monthly TNA, we calculate monthly implied net flows following Sirri and Tuffano (1998). From Morningstar, we obtain monthly information on average holdings durations. We aggregate share-class information on fund level using the match provided by Morningstar. Excluding all funds with incomplete information or less than 12 monthly observations once they surpass the threshold TNA of 5 million USD (e.g., Fama and French, 2010) leaves us with samples of 127 active US domestic government bond funds and 291 active US domestic corporate bond funds in the sample period from 01/1990 to 12/2014.

Table 5 reports detailed summary statistics on fund characteristics for all funds with Panels A and B showing figures for government bond funds and corporate bond funds separately. The average effective duration of government (corporate) bond funds is 4.12 (4.09) years which is rather low compared to the average duration of the broad Treasury index of 5.44 years reported in Table 1. We attribute this disparity to the fact that bonds with long durations are predominantly held by insurance companies and pension funds due to their rather long-term investment horizons. Thus, broad bond market indices cannot reflect the average duration structure of bond funds. Further, the average fund characteristics like TNA, age, turnover, expense ratio and flow are in line with previous research (e.g., Comer and Rodriguez, 2013). Finally, many of the government bond funds in our sample have considerable holdings in corporate bonds and vice versa, indicating the necessity of incorporating respective additional risk factors in the performance models.

[Insert Table 5 here.]

For each of these funds, we identify the duration-adjusted benchmark by the minimum gap between the fund's and the Treasury index's average holdings duration during the fund's existence within our sample period.[19]

4.4.2 Performance measures

To measure bond fund performance, we use several regression models of which the most basic one uses only a single Treasury index or factor to explain government bond returns. Eq. (6) represents the *SF*-model:

$$er_{i,t} = \alpha_i^{SF} + \beta_{i,\,treasury}^{SF}\, er_t^{treasury\ D} + \varepsilon_{i,t} \tag{6}$$

Where $er_{i,t}$ represents fund i's return in month t in excess of the risk-free rate, $er_t^{treasury\ D}$ represents the excess return of a Treasury index with duration D, α_i^{SF} represents fund i's mean abnormal return, $\beta_{i,treasury}^{SF}$ is the sensitivity of fund i's returns to the returns of the Treasury index and $\varepsilon_{i,t}$ is an error term with $E(\varepsilon_{i,t}) = 0$.

Because the bond funds in our sample also have considerable holdings in asset classes other than government bonds as shown in Table 5, we augment the regression model to capture

[19] Information regarding the average durations of the US Treasury indices is available beginning in 1997. As the variation in average duration over time is rather low as shown in Table 1, we use the average holdings duration reported for each index in 01/1997 as the best estimate for the average holdings duration in the period from 01/1990 to 12/1996.

other systematic risks. Thus, model *MF*1 (Eq. 7) additionally includes a default factor def_t which is constructed as the return of a BofAML high yield index minus the return of an intermediate term Treasury index (5–7y). This model is used by, e.g., Gutierrez et al. (2009) and Huij and Derwall (2008).

$$er_{i,t} = \alpha_i^{MF1} + \beta_{i,treasury}^{MF1}\, er_t^{treasury\,D} + \beta_{i,\,def}^{MF1}\, def_t + \varepsilon_{i,t} \qquad (7)$$

In addition to investing in bonds with considerable default risk, government bond funds may also invest in mortgage-backed securities which have non-linear, option-like features. Thus, in model *MF*2 (Eq. 8), we include the return of a high yield index, $er_{hy,t}$, and the return of a mortgage-backed bond index, $er_{gnma30,t}$, in addition to the Treasury index. This model is used by, e.g., Huij and Derwall (2008).

$$er_{i,t} = \alpha_i^{MF2} + \beta_{i,treasury}^{MF2}\, er_{treasury,t}^{D} + \beta_{i,hy}^{MF2}\, er_{hy,t} + \beta_{i,gnma}^{MF2}\, er_{gmna30,t} + \varepsilon_{i,t} \qquad (8)$$

Further augmenting the multi-factor models to capture also equity holdings of government bond funds, *MF*3 (Eq. 9) includes the default factor, the mortgage-backed security index and the CRSP value-weighted stock index in addition to the Treasury index.

$$er_{i,t} = \alpha_i^{MF3} + \beta_{i,\,treasury}^{MF3}\, er_{treasury,t}^{D} + \beta_{i,\,def}^{MF3}\, def_t + \beta_{i,gnma}^{MF3}\, er_{gmna30,t} \qquad (9)$$

$$+ \beta_{i,mkt}^{MF3}\, er_{mkt,t} + \varepsilon_{i,t}$$

Our final model *MF*4 (Eq. 10) captures the same risks as model *MF*3, however, instead of the mortgage-backed security index, we use an option factor $option_t$ which is constructed as the GNMA index minus an intermediate-term US Treasury index.

$$er_{i,t} = \alpha_i^{MF4} + \beta_{i,treasury}^{MF4}\, er_{treasury,t}^{D} + \beta_{i,def}^{MF4}\, def_t + \beta_{i,\,option}^{MF4}\, option_t \qquad (10)$$

$$+ \beta_{i,mkt}^{MF4}\, er_{mkt,t} + \varepsilon_{i,t}$$

4.4.3 Government bond funds

4.4.3.1 Average performance

Table 6 shows average alpha estimates aggregated over all government bond funds and benchmarked separately against all ten US Treasury indices, against the term factor as well as against the funds' individually identified duration-adjusted benchmark. Panel A shows results for net returns in the overall sample period, while Panel B shows similar results using gross returns. The first interesting finding is that the average alpha of the funds in our sample increases with the duration of the benchmark index used with, e.g., α_i^{SF} in Panel A increasing almost monotonically from -0.3290% p.a. using the 1–3 index as the benchmark to 0.4981% p.a. using the 10+ index. This is in line with our index based findings. Further, the average alpha using the broad index is similar to performance measured using the 5–10 index and the term factor produces the highest average alpha; both results are also in line with our index based analysis. The duration-adjusted average alpha with -0.1053% p.a. is below the average alpha measured using the broad index, which is in line with the majority of funds having average durations below that of the broad index which means that using the broad index for all funds systematically overestimates average government bond fund performance.

[Insert Table 6 here.]

Comparing the different factor models reveals that the SF model likely does not capture all relevant risks in government bond fund performance because the average alphas generated by the MF models are distinctively lower with duration-adjusted performance around -1.1% p.a. In the case of the MF4 model, government bond fund performance is significantly negative using all Treasury indices, which is in line with the overall finding that on average active management does not add value for investors. The only exception is the average alpha using the term factor, which is significantly positive; another indication of the problems caused by using a term factor.

Comparing the results for net returns in Panel A with those for gross returns in Panel B reveals that, while the average gross alpha using the broad index is only slightly negative and statistically insignificant, the average gross duration-adjusted alpha is significantly negative with -0.3724% p.a. indicating that government bond funds do not even earn the fees they charge. This is in contrast to numerous findings on equity funds which show that average gross performance is zero on average (e.g., Sharpe, 1991; Fama and French, 2010).

About different market phases, Panels C and D show average government bond alphas using net returns during months with widening and contracting term spreads, respectively. The relations are in line with our index based analysis. Panel C shows that average alpha increases with higher benchmark duration while Panel D shows decreasing alpha with higher benchmark duration. Further, the differences between the average alphas using different indices are much higher, thereby creating even clearer patterns in both panels compared to Panel A. Regarding the average performance during both periods, the results using the duration-adjusted benchmark indicate that average performance during phases with widening term spread is much lower than the average performance during phases with contracting term spread. Moreover, this difference is more pronounced for the duration-adjusted alpha than for the broad Treasury alpha. This is consistent with the overall impression in Table 3 that bond returns are lower during phases of widening term-spread compared to phases with a contracting term-spread.

4.4.3.2 Duration bias in government bond fund performance

The results in the previous subsection on the average alpha performance of all funds show that there is a duration bias in government bond fund performance. To get into more detail on the magnitude and significance of the bias, Table 7 shows the results of an analysis similar to our pairwise Treasury index regressions presented in Table 2. Therefore, we sort all funds into groups according to their duration-adjusted benchmark. Then, for each group with at least five funds, we perform paired mean-comparison t-tests between the funds' duration-adjusted MF4-

model alphas and the MF4-model alphas using the other nine Treasury indices.[20] The figures represent the average alpha differences, i.e. duration-biases, for each duration group with respect to each Treasury index.

[Insert Table 7 here.]

Panel A shows results for the overall period while Panels B and C show results for the different market phases. Overall, the relations are as expected so that in Panels A and B using an index with a higher than appropriate duration leads to a positive duration bias while using a shorter than appropriate one leads to a negative duration bias. Panel C shows the reversed relation for market phases with contracting yield spread, consistent with our index based analysis in Table 2. Regarding the magnitude of the bias in different fund groups, we primarily look at the results for the broad Treasury index. While the bias decreases from left to right with growing bond fund duration, it is positive and significant for the first four groups which cover 96 of the total 127 funds in our sample (75.6%). This explains why the overall duration-adjusted performance in Table 6 is slightly lower than the alpha using the broad index as a benchmark.

Regarding single groups, the average duration-bias in Panel A may be as high as 0.5520% p.a. for the 24 funds sorted in group 1–3 which is clearly of economic relevance. During phases of widening term spreads in Panel B, the duration bias of group 1–3 increases to 1.4663% p.a. while the largest duration bias during phases of contracting term-spread is reported in Panel C for the group 1–5 with -0.6104% p.a. Thus, depending on the market phase, using an index with an inappropriate duration as benchmark leads to economically significant duration-bias, consistent with our main research hypothesis.

[20] Results for the SF- and the other MF-models are economically similar and available upon request. The same applies to results based on gross returns.

4.4.4 Corporate bond funds

Given that the Treasury indices we use capture only term-structure risk, we do not use the SF-model for corporate bond funds. Moreover, as Table 6 shows, the results for the MF-models are relatively similar we exclusively use *MF*4 (Eq. 6) in our corporate bond fund analysis.[21] Table 8 reports average alphas of all funds against all Treasury bond indices as well as against the term factor and the duration-adjusted benchmark. For net returns in the overall period, Panel A shows the usual pattern of increasing alpha with increasing index duration. Performance measured using the broad index is between the performance using the 5–7 and the 5–10 indices as expected and the highest alpha is displayed for the term factor. The duration-adjusted alpha is below the broad index alpha indicating a general overestimation of corporate bond funds performance in previous research. Compared to government bond funds, the average performance of corporate bond funds is higher. Panel B reports average gross alphas. As expected, the results are higher leading to a significantly positive average duration-adjusted alpha of 0.5015% p.a.

[Insert Table 8 here.]

Panels C and D report corporate bond fund performance results separately for different market phases. Generally, the results show the same patterns like government bond funds in that the overall pattern is pronounced in times of a widening term-spread and reverted in times of a contracting term-spread. However, the differences are much smaller than displayed for government bond funds, especially in Panel D. The duration-adjusted alpha is more or less the same during both market phases. However, the alpha using the broad index is very high for a widening term spread and low during periods with contracting term-spread causing false interpretations regarding the average performance during both phases.

[21] Similar analyses using the other MF-models yield economically similar results and are available upon request.

For a deeper insight into the relevance and the significance of the duration bias of single funds, Table 9 shows the duration bias for all duration-groups with at least five funds. The patterns in Panels A and B for the overall period and phases with a widening term-spread are as expected. The patterns in Panel C for phases with a contracting term-spread are again less pronounced and less clear compared to government bond funds, especially below the diagonal. Looking at the most important duration bias between the duration-adjusted alpha and the broad index, Panel A shows significantly positive bias estimates for four of five groups which represent 253 of 291 funds (86.9%). This explains the overall positive duration bias. However, the bias for the single groups is not as high as for government bond funds with the highest overall bias displayed for the 5–7 group with 0.3063% p.a. Still, we consider a bias of this magnitude as economically important. During different market phases, the duration bias in corporate bond fund performance can be as high (low) as 1.3750% p.a. for the 1–5 group in Panel B and -0.5294% p.a. for the 1–3 group in Panel C. This also confirms the existence of an economically relevant duration bias in corporate bond fund performance and further promotes the necessity of using the duration-adjusted benchmark approach.

[Insert Table 9 here.]

4.5 Justification of the duration-adjusted performance model

In the previous sections, we show that the alpha performance of bond funds depends on the duration of the benchmark index. This creates a significant duration bias which we define as the difference between our duration-adjusted alpha and any alpha using an inappropriate index. However, the identification of the duration-adjusted benchmark as the one with the closest duration to that of the fund is more or less heuristic. Whether our model itself yields unbiased measures of selection performance, is thus unclear.

Therefore, we run pooled panel regressions with two-dimensionally clustered standard errors (Petersen, 2009) where we explain alphas from the broad Treasury index and the duration-adjusted benchmark with changes of the term-structure.[22] Thus, we first estimate bond fund alphas in overlapping 12-month rolling windows using the *SF* and the *MF*4-models. For the same 12-month windows, we calculate changes of the level (*dL*), slope (*dS*) and curvature (*dC*), where *L*, *S*, and *C* are defined as in the sensitivity analysis in Section 4.2. Further, sign and magnitude of the duration bias in different market phases depend on the duration gap between the benchmark and the fund, as our previous results show. Therefore, we use interactions of the term structure changes with the average duration gap (DG) between the index and the fund during the 12-month window. Finally, because these interaction variables are highly correlated we also construct orthogonalized variables following the democratic, simultaneous orthogonalization procedure of Klein and Chow (2013).[23]

The results for government bond funds are presented in Super-Panel I of Table 10 where Panel A use alphas from the *SF*-model and Panel B uses the *MF*4-model. The regressions using the term structure variables separately show that alphas from the broad index on the left are significantly driven by changes of all three term structure parameters, especially by changes of level and slope. This is also true when using the three orthogonalized term structure variables in combination. Further, looking at the adjusted R^2 of 2% for level and slope reveals that the term structure variables have some explanatory power over alpha. These results vanish almost completely if we look at the results for the duration-adjusted alphas. Here, the term structure variables display statistically insignificant coefficients. Moreover, the adjusted R^2 is zero in all

---

[22] Alternative panel regressions using time-fixed effects yield similar results and are available upon request.

[23] The correlations are as high as: ρ_{dLdS} = -79%; ρ_{dLdC} = 38%; ρ_{dSdC} = 82%. The procedure by Klein and Chow (2013) is designed to construct uncorrelated variables which, however, maintain very high correlations with the initial factors and thus allow similar economic interpretations. Alternative results using a conventional orthogonalization with level as the base variable are economically similar and available upon request.

cases showing that term structure changes have no explanatory power over our duration-adjusted performance.

[Insert Table 10 here.]

The results for corporate bond funds in Super Panel II are similar, albeit a bit weaker which is in line with our previous finding that the duration bias is generally less strong in corporate bond funds compared to government bond funds. Overall, the results of this analysis prove that bond fund alphas using the usual broad index are driven by term structure changes while our duration-adjusted model provides an unbiased measure of bond fund performance.

4.6 Implications of duration-adjusted performance for prior and future research

4.6.1 Performance persistence

From the previous sections it is by now clear that for both types of funds, government and corporate bond funds, the usual performance measurement using a broad bond index to capture term-structure risk suffers from an economically relevant duration bias. Overall and in "normal" phases of widening term-spreads, this bias leads to significant overestimation of the performance of short-term bond funds and underestimation of the performance of long-term bond funds. During phases with a contracting term-spread, it is consistently the other way around.

Therefore, we next analyze if another usual finding by previous bond fund studies that bond fund performance is persistent over time (e.g. Huij and Derwall, 2008) could at least partly be driven by this systematic over- and underestimation of specific fund groups. Closely following the methodology of Huij and Derwall (2008), we use bond fund alphas from non-overlapping 12-months rolling windows using the four MF-models introduced in Eqs. (7) to (10). As term-risk benchmarks, we alternatively use the broad index, the term factor as well as two versions of the duration-adjusted benchmark. Version “average” uses the same index as in Sections 4 and 5 where we identify the duration-adjusted benchmark based on the average lifetime duration of the funds. Version “monthly” creates a new hypothetical index for each fund by using in each month the return of the index with the nearest duration to that of the fund during this month.

Table 11 shows the results of performance persistence tests similar to those by Huij and Derwall (2008), where Super-Panel I reports government bond funds and Super-Panel II corporate bond funds. The respective Panel A presents the average slope coefficients and R^2 statistics of cross-sectional Fama and MacBeth (1973) regressions of contemporary alpha on one-year lagged alpha while Panels B report respective Spearman rank correlations. The Fama and MacBeth regressions in Panel I.A for government bond funds using the broad index and the term factor show the usual finding that performance persists with relatively high coefficients and R^2 statistics. These parameters clearly reduce if either version of the duration-adjusted benchmark is used in the performance models, especially in MFs 2–4 where the coefficients are not statistically significant anymore. Similar results are displayed in Panel I.B for the Spearman rank correlations which are higher when using the broad index or the term factor and distinctively lower and less significant when using either version of the duration-adjusted benchmark. This leads us to the conclusion that for government bond funds, a considerable portion of the usual finding that bond fund performance persists is due to the use of benchmarks with an inappropriate duration that induces persistence mechanically through the duration bias shown in the previous sections.

[Insert Table 11 here.]

For corporate bond funds in the lower half of the table, these findings are less pronounced. Firstly, the coefficients and R^2 statistics of the Fama and MacBeth-regressions in Panel II.A are overall lower than in Panel I.A indicating lower overall persistence. Secondly, probably due to the larger sample size, the coefficients are statistically significant for all benchmarks. However, coefficients and R^2 statistics still reduce from using the broad index or the term factor to using either version of the duration-adjusted benchmark. The same applies for the Spearman rank correlations in Panel II.B. On the one hand, this finding is consistent with some part of the persistence in corporate bond funds being mechanically driven by duration bias. On the other hand, it is in accordance with our finding in previous sections that the duration bias is less relevant for corporate bond funds than for government bond funds because of the higher relative importance of other risks. Overall, this confirms our expectation that performance persistence in bond funds is at least partly driven by the mechanical duration bias through the use of inappropriate benchmarks and less by "hot hands."

4.6.2 Prospectus benchmark analysis

As we show in the previous sections, using the wrong benchmark can significantly bias the measured performance of mutual funds upwards, i.e. by using an index with a too long duration during "normal" periods. Thus, measured performance may be subject to manipulation if bond fund managers are aware of the duration bias and accordingly name indices with too long durations as their official benchmarks. To analyze whether bond fund managers use this opportunity, e.g. to attract investor flows, we use the name of the prospectus benchmark index provided via Morningstar for a subset of the funds in our sample. For these, we manually extract the indices' maturity ranges and compare those to the maturity ranges of the indices identified in Section 4.4.1 by average duration as the funds' duration-adjusted benchmarks.[24]

[Insert Table 12 here.]

[24] While this method lacks precision compared to a comparison directly using the prospectus benchmarks' durations, this latter information is unfortunately not given in the database.

Table 12 reports the results with Panel A showing government bond funds and Panel B corporate bond funds. The table shows for each duration-adjusted benchmark group in the columns the number of funds with a specific maturity-range according to prospectus benchmark. The first thing to notice is that a large number of funds do either not specify any maturity range or name an “intermediate” or “broad” index as their prospectus benchmark. More importantly, the second thing to notice is that the maturity ranges overall show an excellent match with the duration-adjusted benchmark groups as indicated by the shaded areas in both panels. But there are also extreme cases, e.g. for corporate bond funds, where the duration-adjusted benchmark is the 1–3 index and the prospectus benchmark specifies an index with a maturity range of 10+ years. However, those are very rare.

Thus, we conclude that managers are on average unaware of the opportunity to manipulate measured performance by naming a too long benchmark index. Or they are plain honest. This further raises the question why previous research does not properly account for the average duration or term to maturity of bond funds, given that the maturity range is more or less correctly specified by most government and corporate bond funds via their prospectus benchmarks? Finally, it confirms our overall argument that using a benchmark with an appropriate duration is important for both academic research and practical performance assessment.

4.7 Conclusion

Can we measure bond fund performance the same way as we measure equity fund performance? In this paper, we answer this important question with a clear: No and Yes! *No,* because the non-linear relation between bond returns and systematic term-risk as measured by duration leads to a significant mechanical bias in bond fund performance if the benchmark index’s duration does not match the duration of the fund. Hence using the same benchmark index for all bond funds like we do in equity fund studies may not be recommended. *Yes* because the solution to this problem is quite simple in that we propose using for each bond fund the one index from a variety of alternative indices which best matches the duration of the fund to measure its duration-adjusted performance. Thus, overall we can continue to use similar methods as with equity funds but still measure performance in a consistent, unbiased way.

In a broad range of empirical tests using US Treasury indices focusing on different maturity ranges and a comprehensive sample of US government and corporate bond funds, we show that the duration bias in measured bond fund performance is statistically and economically significant. About the usual use of a broad Treasury bond index, duration-adjusted performance is lower on average indicating a potential overestimation of average bond fund performance in previous research. Moreover, due to the systematic over- and underestimation of the performance of specific fund groups, the consistent previous finding of performance persistence in bond funds can be at least partly traced back to the duration bias. Finally, we show that the fund's self-stated prospectus benchmarks provide dependable information on the fund's maturity range facilitating the identification of each fund's specific duration-adjusted benchmark.

Appendix A – Derivation of the term structure function

The following derivation of a term structure model with economically interpretable parameters level (*L*), slope (*S*), and curvature (*C*) is based on Wilkens' (1994) modification of Eq. (A1) used by the German central bank (Deutsche Bundesbank, 1991), where $r_{T,K,i}$ is the empirical yield of bond *i* with maturity in *T* and coupon *K*.

$$r_{T,K,i} = b_0 + b_1\, T_i + b_2 \ln T_\mathrm{i} + b_3\, K_i + b_4 \ln K_i + e_i \tag{A1}$$

Taking expectations yields

$$r_{T,K} = b_0 + b_1\, T + b_2 \ln T + b_3\, K + b_4 \ln K. \tag{A2}$$

Assuming homogeneous coupons for similar maturities at each point in time allows replacing the coupon effect with a constant $k = b_3\, K + b_4 \ln K$.

$$r_T = b_0 + b_1\, T + b_2 \ln T + k \tag{A3}$$

As coefficients $b_0 – b_2$ have no economically meaningful interpretation, they are subsequently replaced with alternative coefficients via canonical variable transformation. Therefore, three explicit maturities, one for each coefficient, are inserted into Eq. (A3).

$$r_1 = b_0 + b_1\, 1 + b_2 \ln 1 + k$$

$$r_5 = b_0 + b_1\, 5 + b_2 \ln 5 + k$$

$$r_{10} = b_0 + b_1\, 10 + b_2 \ln 10 + k.$$

Now, b_0 is expressed as a combination of r_1 and b_1.

$$b_0 = r_1 – b_1 – k \tag{A4}$$

b_1 is expressed as a combination of r_5, r_1, and b_2.

$$r_5 = r_1 - b_1 - k + b_1\ 5 + b_2 \ln 5 + k$$
$$r_5 = r_1 + b_1\ 4 + b_2 \ln 5$$

$$b_1 = \frac{r_5 - r_1 - b_2 \ln 5}{4} \tag{A5}$$

b_2 is expressed as a combination of all three interest rates.

$$r_{10} = r_1 - b_1 - k + b_1\ 10 + b_2 \ln 10 + k$$
$$r_{10} = r_1 + b_1\ 9 + b_2 \ln 10$$
$$r_{10} = r_1 + \frac{r_5 - r_1 - b_2 \ln 5}{4}\ 9 + b_2 \ln 10$$
$$r_{10} = r_1 + \frac{9}{4} r_5 - \frac{9}{4} r_1 - \frac{9}{4} b_2 \ln 5 + b_2 \ln 10$$
$$r_{10} = r_1 + \frac{9}{4} r_5 - \frac{9}{4} r_1 - 1.31865021\ b_2$$
$$r_{10} = \frac{9}{4} r_5 - \frac{5}{4} r_1 - 1.31865021\ b_2$$

$$b_2 = \frac{2.25\ r_5 - 1.25\ r_1 - r_{10}}{1.31865021} \tag{A6}$$

Next, parameters $b_0 - b_2$ in Eq. (A3) are substituted with the respective Eqs. (A4) to (A6).

$$r_T = b_0 + b_1\ T + b_2 \ln T + k$$
$$r_T = r_1 - b_1 - k + b_1\ T + b_2 \ln T + k$$
$$r_T = r_1 - b_1 + b_1\ T + b_2 \ln T$$
$$r_T = r_1 + b_1 (T - 1) + b_2 \ln T$$
$$r_T = r_1 + \frac{r_5 - r_1 - b_2 \ln 5}{4} (T - 1) + b_2 \ln T$$
$$r_T = r_1 + \frac{r_5 - r_1}{4} (T - 1) - \frac{b_2 \ln 5}{4} (T - 1) + b_2 \ln T$$
$$r_T = r_1 + \frac{r_5 - r_1}{4} (T - 1) + b_2 \left(\ln T - \frac{\ln 5\ (T - 1)}{4}\right)$$
$$r_T = r_1 + \frac{r_5 - r_1}{4} (T - 1) + \frac{2.25\ r_5 - 1.25\ r_1 - r_{10}}{1.31865021} \left(\ln T - \frac{\ln 5\ (T - 1)}{4}\right)$$

$$r_T = r_1 - \frac{r_1}{4}(T-1) - 0.947939029\, r_1 \ln T + 0.381412253\, r_1 (T-1)$$
$$+ \frac{r_5}{4}(T-1) + 1.706290253\, r_5 \ln T - 0.686542055\, r_5 (T-1)$$
$$- \frac{r_{10} \ln T}{1.31865021} + 0.305129802\, r_{10} (T-1)$$

$$r_T = r_1 - \frac{r_1}{4}(T-1) - 0.947939029\, r_1 \ln T + 0.381412253\, r_1 (T-1)$$
$$+ \frac{r_5}{4}(T-1) + 1.706290253\, r_5 \ln T - 0.686542055\, r_5 (T-1)$$
$$- \frac{r_{10} \ln T}{1.31865021} + 0.305129802\, r_{10} (T-1)$$

$$r_T = r_1 \left(1 - \tfrac{1}{4}\, T + \tfrac{1}{4}(-0.947939029) \ln T + 0.381412253\,(T-1)\right)$$
$$+ r_5 \left(\tfrac{1}{4}(T-1) + 1.706290253 \ln T - 0.686542055\,(T-1)\right)$$
$$+ r_{10}\,(0.305129802\,(T-1) - 0.758351223 \ln T\,)$$

$$r_T = r_1\,(0.868587747 + 0.131412253\, T - 0.947939029 \ln T\,) \tag{A7}$$
$$+ r_5\,(0.436542055 - 0.436542055\, T + 1.706290253 \ln T\,)$$
$$+ r_{10}\,(-0.305129802 + 0.305129802\, T - 0.758351223 \ln T)$$

Next, explicit functional forms for level (L), slope (S) and curvature (C) are defined using the same three maturities as above.

$$L = r_1$$
$$S = \frac{r_{10} - r_1}{9} \qquad (r_{10} = L + 9\,S)$$
$$C = r_5 - r_1 - 4\,S \qquad (r_5 = L + C + 4\,S)$$

Finally, the interest rates in Eq. (A7) are replaced with the functions for L, S, and C.

$$r_T = r_1\ (0.868587747 + 0.131412253\ T - 0.947939029 \ln T\ )$$
$$+\ r_5\ (0.436542055 - 0.436542055\ T + 1.706290253 \ln T\ )$$
$$+\ r_{10}\ (-\ 0.305129802 + 0.305129802\ T - 0.758351223 \ln T\ )$$

$$r_T = L\ (0.868587747 + 0.131412253\ T - 0.947939029 \ln T\ )$$
$$+\ (L + C + 4\ S)\ (0.436542055 - 0.436542055\ T + 1.706290253 \ln T\ )$$
$$+\ (L + 9\ S)\ (-\ 0.305129802 + 0.305129802\ T - 0.758351223 \ln T\ )$$

$$r_T = 0.868587747\ L + 0.131412253\ T\ L - 0.947939029\ (\ln T)\ L$$
$$+\ 0.436542055\ L - 0.436542055\ T\ L + 1.706290253\ (\ln T)\ L$$
$$+\ 0.436542055\ C - 0.436542055\ T\ C + 1.706290253\ (\ln T)\ C$$
$$+\ 1.74616822\ S - 1.74616822\ T\ S + 6.825161012\ (\ln T)\ S$$
$$-\ 0.305129802\ L + 0.305129802\ T\ L - 0.758351223\ (\ln T)\ L$$
$$-\ 2.746168218\ S + 2.746168218\ T\ S - 6.825161007\ (\ln T)\ S$$

$$r_T = L + S\ (T - 1) + C\ (0.436542055 - 0.436542055\ T + 1.706290253 \ln T\ ). \qquad \text{(A8)}$$

Appendix B

Table A
Spot rates, bond prices and bond returns for Figure 1, Scenario (iii)

	Basic term-structure		After the term-structure change			
T	Spot rate r_T	ZB price P_T	Spot rate r_T	ZB price P_T	ZB return dP_T/P_T	Lin. Approx. with 10y
1	1.00%	0.9901	1.00%	0.9901	0.000%	0.000%
2	2.06%	0.9600	1.91%	0.9629	0.295%	1.527%
3	2.75%	0.9218	2.45%	0.9299	0.881%	3.054%
4	3.29%	0.8785	2.84%	0.8940	1.762%	4.582%
5	3.75%	0.8319	3.15%	0.8564	2.942%	6.109%
6	4.16%	0.7832	3.41%	0.8180	4.431%	7.636%
7	4.53%	0.7336	3.63%	0.7793	6.240%	9.163%
8	4.87%	0.6836	3.82%	0.7409	8.383%	10.690%
9	5.19%	0.6341	3.99%	0.7030	10.878%	12.218%
10	5.50%	0.5854	4.15%	0.6659	13.745%	13.745%
11	5.79%	0.5381	4.29%	0.6297	17.009%	15.272%
12	6.08%	0.4926	4.43%	0.5945	20.698%	16.799%
13	6.35%	0.4490	4.55%	0.5605	24.845%	18.327%
14	6.62%	0.4076	4.67%	0.5278	29.488%	19.854%
15	6.88%	0.3685	4.78%	0.4963	34.669%	21.381%
16	7.14%	0.3319	4.89%	0.4661	40.438%	22.908%
17	7.39%	0.2977	4.99%	0.4372	46.850%	24.435%
18	7.63%	0.2661	5.08%	0.4097	53.969%	25.963%
19	7.87%	0.2369	5.17%	0.3834	61.867%	27.490%
20	8.11%	0.2101	5.26%	0.3585	70.625%	29.017%

This table shows spot rates, zero-bond prices and zero-bond returns for the basic term-structure setup and Scenario (iii) of the sensitivity analysis presented in of Section 4.2.

References

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

Berk, J. B and 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. and Xu, D. (2008) Measuring abnormal bond performance. *Review of Financial Studies* 22, 219-258.

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

Boquist, J. A., Racette, G. A. and Schlarbaum, G. G. (1975) Duration and risk assessment for bonds and common stocks. *Journal of Finance* 30, 1360-1365.

Chen, Y., Ferson, W. and Peters, H. (2010) Measuring the timing ability and performance of bond mutual funds. *Journal of Financial Economics* 95, 72-89.

Chen, Y. and Qin, N. (2016) The behavior of investor flows in corporate bond mutual funds. *Management Science* (forthcoming). http://dx.doi.org/10.1287/mnsc.2015.2372.

Comer, G. and Rodriguez, J. (2013) A comparison of corporate versus government bond funds. *Journal of Economics and Finance* 37, 495-510.

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

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1979) Duration and the measurement of basic risk. *Journal of Business* 52, 51-61.

Czaja, M.-G., Scholz, H. and Wilkens, M. (2009) Interest rate risk of German financial institutions: The impact of level, slope, and curvature of the term-structure. *Review of Quantitative Finance and Accounting* 33, 1-26.

Dietz, P. O., Fogler, H. R. and Rivers, A. U. (1981) Duration, nonlinearity, and bond portfolio performance. *Journal of Portfolio Management* 7 (3), 31-41.

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

Elton, E. J., Gruber, M. J. and Nabar, P. G. (1988) Bond returns, immunization and the return generating process. *Studies in Banking and Finance* 5, 125-154.

Estrella, A. and Hardouvelis, G. A. (1991) The term-structure as a predictor of real economic activity. *Journal of Finance* 46, 555-576.

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

Fama, E. F. and French, K. R. (2010) Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance* 65, 1915-1947.

Fama, E. F. and MacBeth, J. D. (1973) Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81, 607-636.

Grinblatt, M. and Titman, S. (1993) Performance measurement without benchmarks: An examination of mutual fund returns. *Journal of Business* 66, 47-68.

Gutierrez, R. C., Maxwell, W. F. and Xu, D. (2009) On economics of scale and persistent performance in corporate bond mutual funds. Working paper, Universtiy of Oregon, Southern Methodist University, Gonzaga University.

Hopewell, M. H. and Kaufman, G. G. (1973) Bond price volatility and term to maturity: A generalized respecification. *American Economic Review* 63, 749-753.

Huij, J. and Derwall, J. (2008) "Hot Hands" in bond funds. *Journal of Banking and Finance* 32, 559-572.

Klein, R. F. and Chow, K. V. (2013) Orthogonalized equity risk premia and systematic risk decomposition. *Review of Quantitative Finance and Accounting* 53, 175-187.

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

Jarrow, R. A. (1978) The relationship between yield, risk, and return of corporate bonds. *Journal of Finance* 33, 1235-1240.

Jensen, M. C. (1968) The performance of mutual funds in the period 1945-1964. *Journal of Finance* 23, 389-416.

Lintner, J. (1965) The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47, 13–37.

Litterman, R. and Scheinkman, J. (1991) Common factors affecting bond returns. *Journal of Fixed Income* 6, 54-61.

Macaulay, F. R. (1938) Some theoretical problems suggested by the movement of interest rates, bond yields and stock prices in the United States since 1856. NBER, Columbia University Press, New York.

Mossin, J. (1966) Equilibrium in a capital asset market. *Econometrica* 34, 768–783.

Nelson, C. R. and Siegel, A. F. (1987) Parsimonious modeling of yield curves. *Journal of Business* 60 (4), 473-489.

Pastor, L., Stambaugh, R. F. and Taylor, L. A. (2015) Scale and skill in active management. *Journal of Financial Economics* 116, 23-45.

Petersen, M. A. (2009) Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22, 435-480.

Sharpe, W. F. (1964) Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19, 425–442.

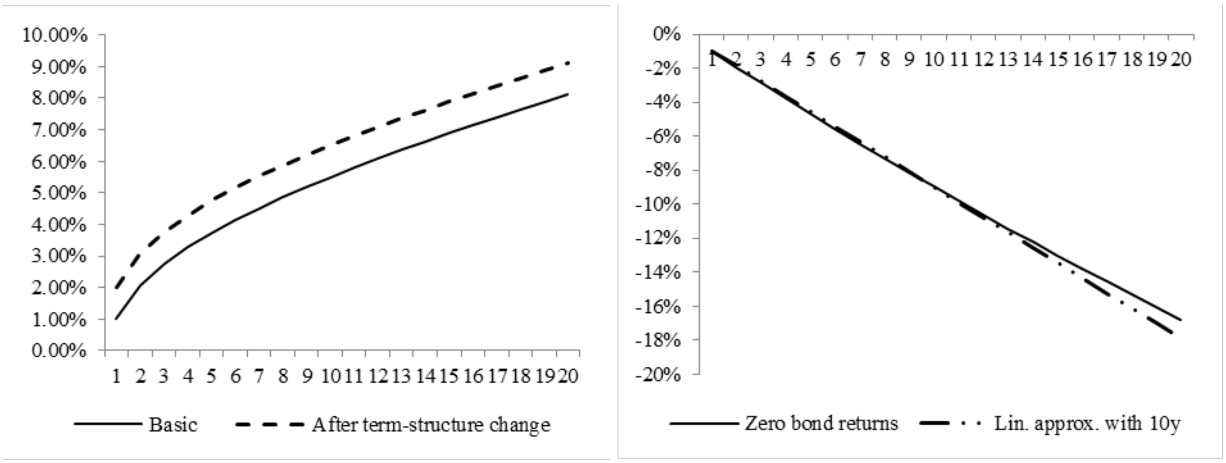
Sharpe, W. F. (1991) The arithmetic of active management. *Financial Analysts Journal* 47, 7-9.

Sirri, E. R. and Tufano, P. (1998) Costly search and mutual fund flows. *Journal of Finance* 53, 1589-1622.

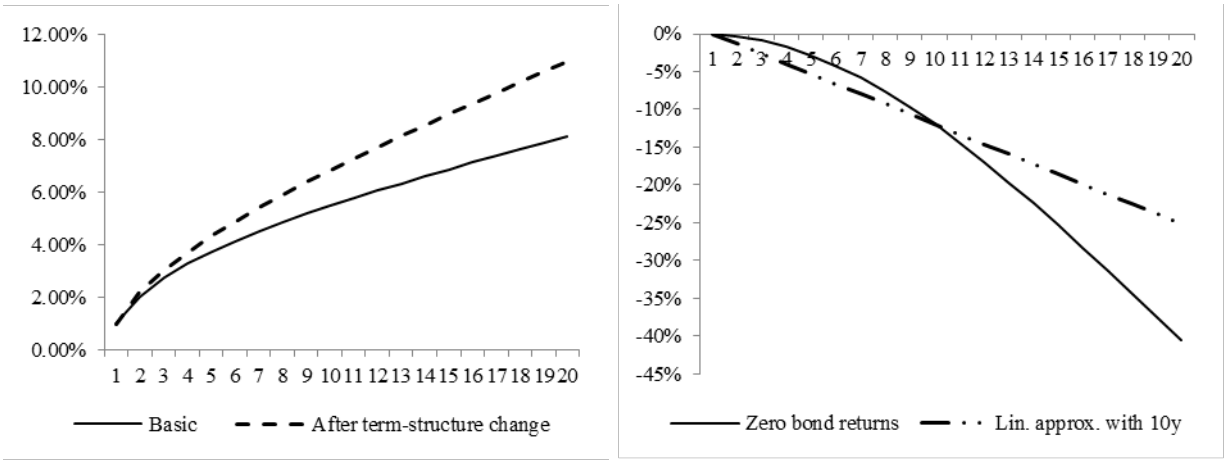
Wilkens, Marco (1994) Risiko-Management mit Zins-Futures in Banken. Neue Betriebswirtschaftliche Studienbücher Nr. 8, Verlag Otto Schwartz & Co., Göttingen.

Figures and Tables

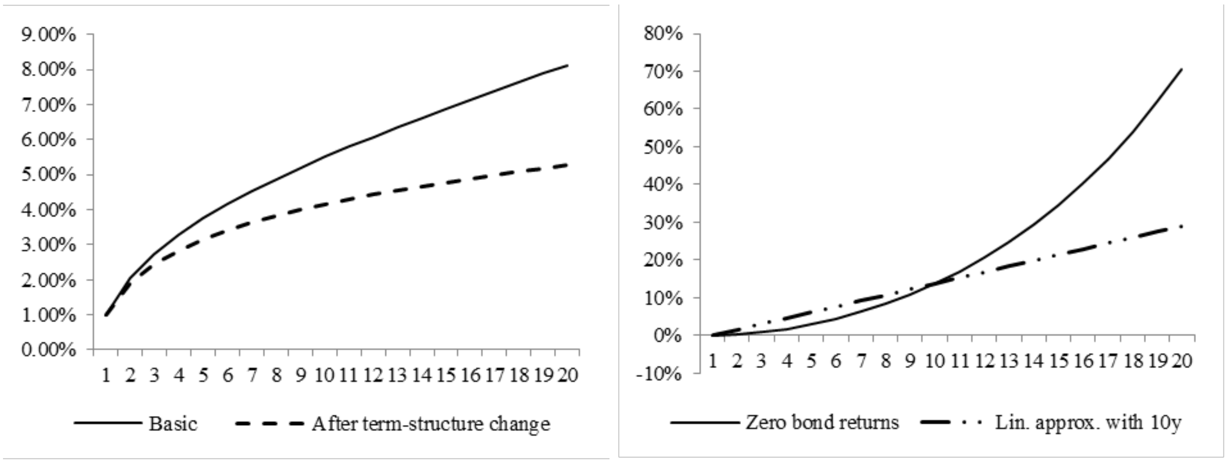
Term-structure **Instantaneous zero-bond returns**



Scenario (i) Isolated increase of the level (*dL* = +1.0%, *dS* = 0, *dC* = 0)

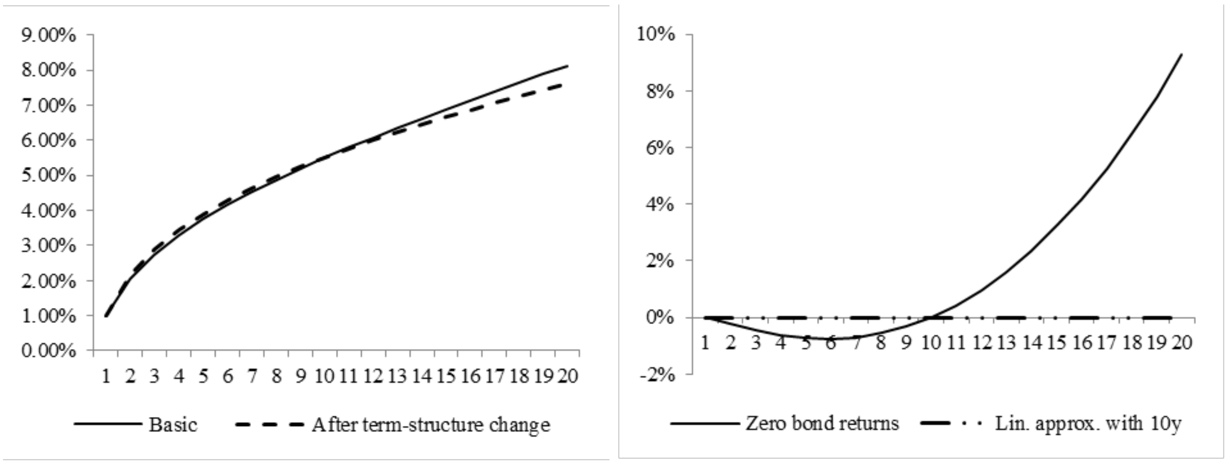


Scenario (ii) Isolated increase of the slope (*dL* = 0, *dS* = +0.15%, *dC* = 0)

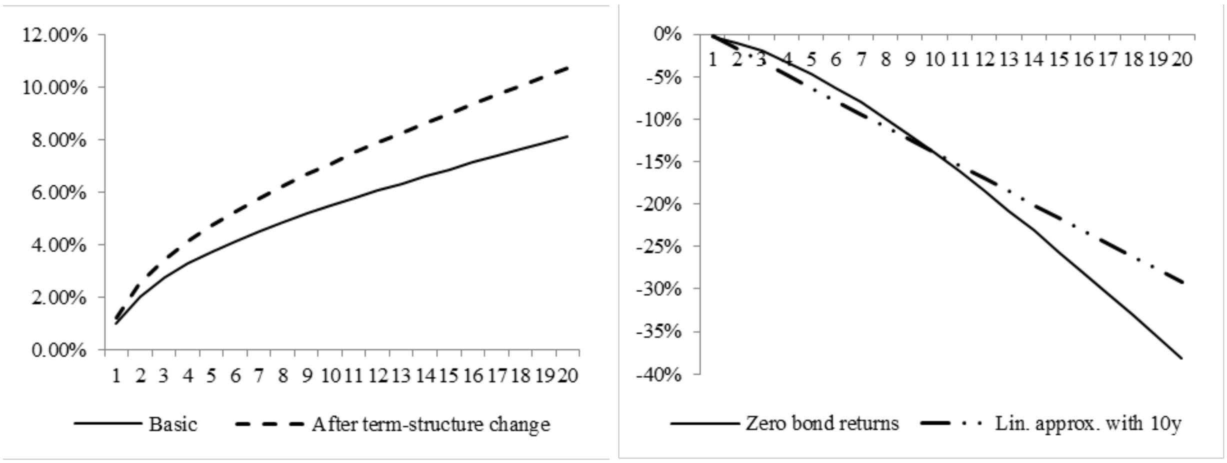


Scenario (iii) Isolated decrease of the slope (*dL* = 0, *dS* = –0.15%, *dC* = 0)

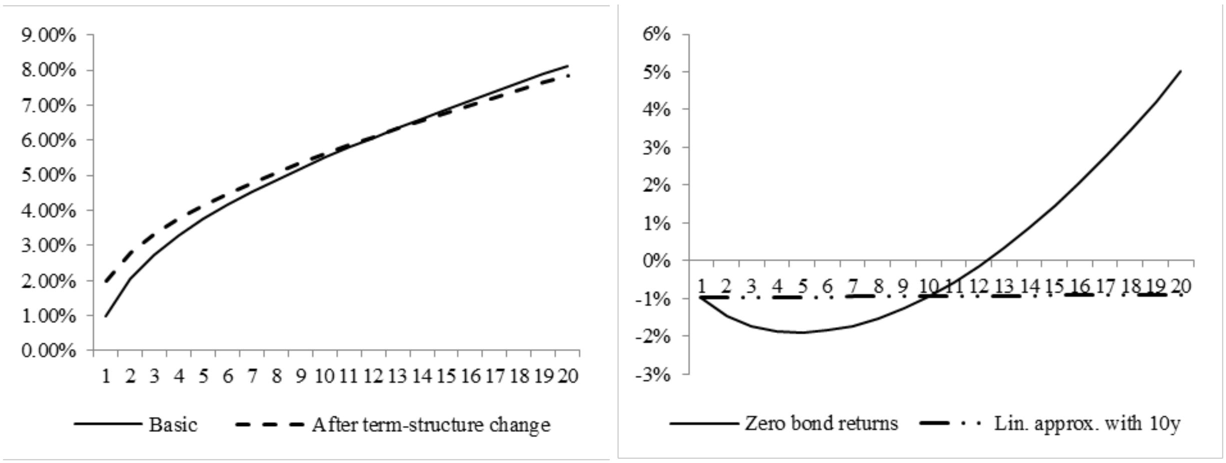
Term-structure **Instantaneous zero-bond returns**



Scenario (vi) Isolated increase of the curvature (dL = 0, dS = 0, dC = +0.15%)



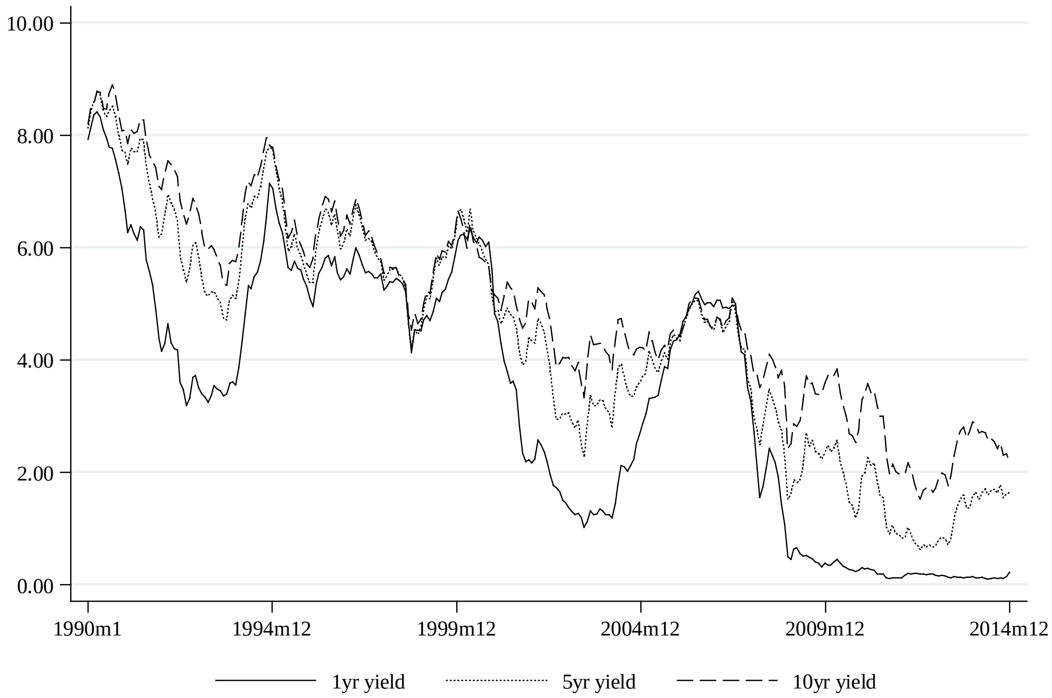
Scenario (v) Combined increase of all 3 factors (dL = +0.25%, dS = +0.15%, dC = +0.15%)



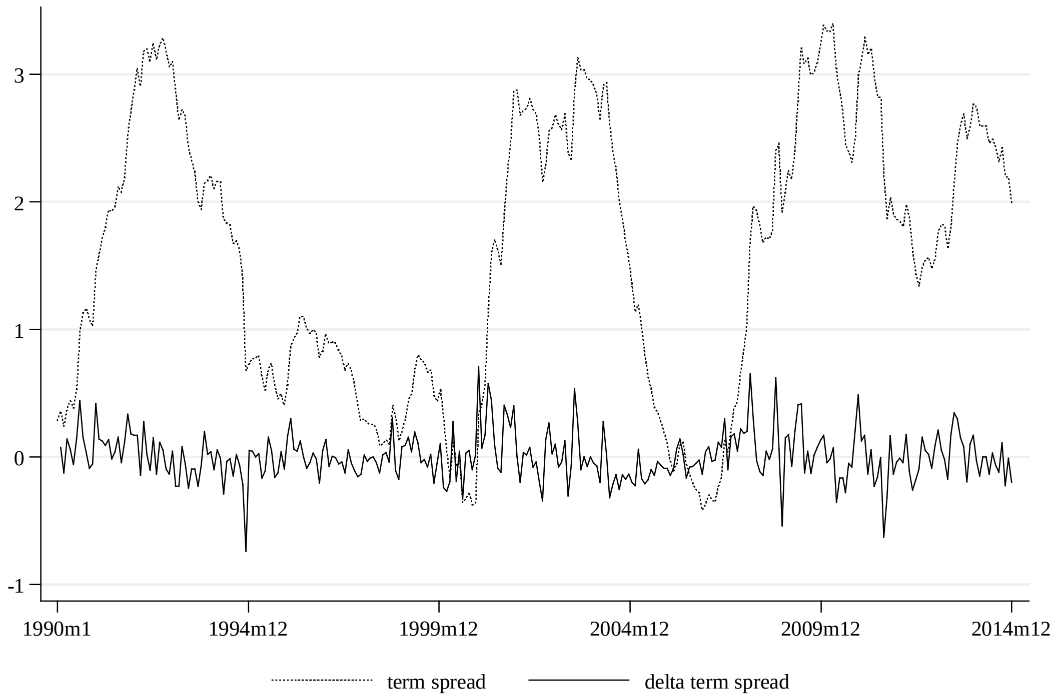
Scenario (vi) Level increase combined with decrease of slope and curvature (dL = +1.0%, dS = –0.1%, dC = –0.2%)

Figure 1. Sensitivity analysis

The left column of this table shows plots of different term structures before (Basic) and after term-structure changes based on Eq. (1) for different example term-structure change scenarios. The basic scenario is: L=1.0%, S=0.5%, C=0.75%. The right columns show the respective zero-bond returns as well as linear approximations thereof using the zero-bonds with maturity in 1 and 10 years.



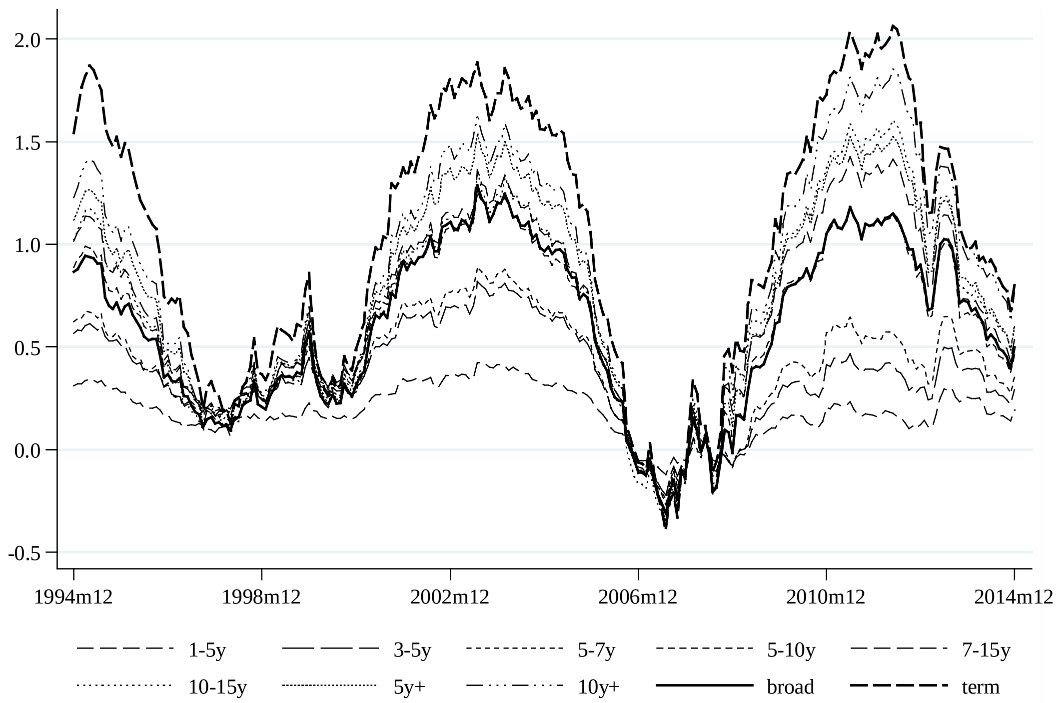
Panel A. Development of Treasury yields over time



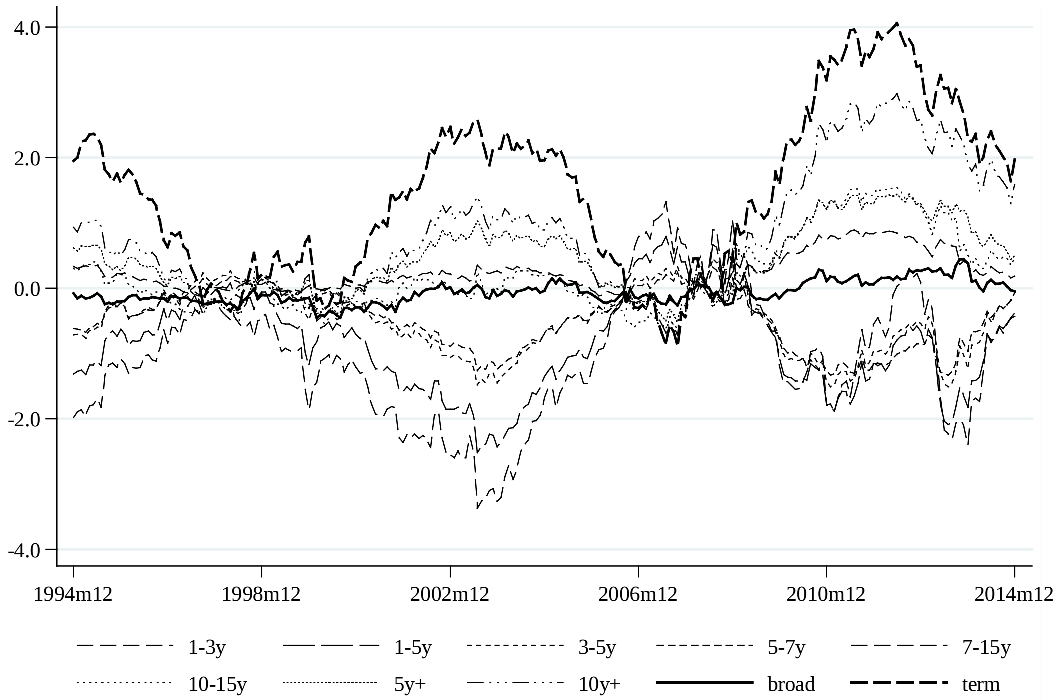
Panel B. Term spread (10yr-1yr) over time

Figure 2. Term-structure

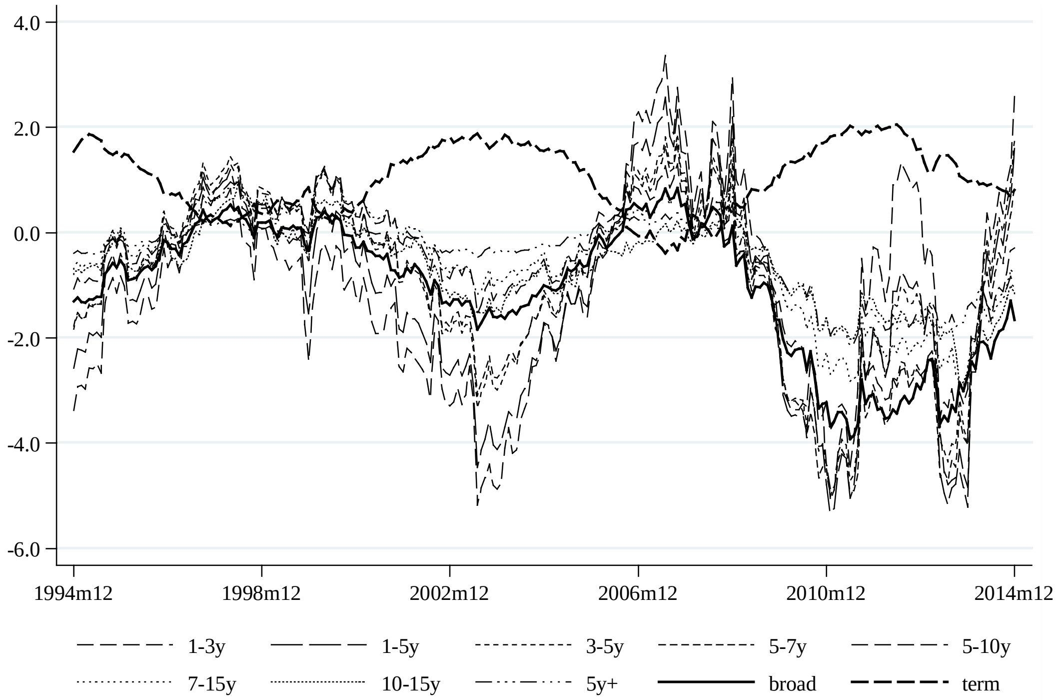
This figure shows the development of different Treasury yields (Panel A) as well as the development of the term spread (Panel B) over the sample period from 01/1990 to 12/2014. All figures are denoted in % p.a.



Panel A. 1–3y



Panel B. 5–10y



Panel C. 10y+

Figure 3. Rolling window duration bias

This figure shows the development of the duration bias in rolling 5-year windows over the sample period from 01/1990 to 12/2014. Duration bias is denoted in % p.a.

Table 1
US Treasury index duration statistics

Panel A. Duration and quality, overall period (1997-2014)

	Duration (in years)					Average
Maturity range	Average	Std. dev.	Min	Median	Max	Quality
1–3 years	1.75	0.11	1.58	1.70	1.94	AAA
1–5 years	2.43	0.19	2.11	2.37	2.75	AAA
3–5 years	3.55	0.18	3.18	3.50	3.85	AAA
5–7 years	4.99	0.41	4.33	4.81	5.69	AAA
5–10 years	6.02	0.40	5.28	6.03	6.63	AAA
7–15 years	7.16	0.71	5.80	7.32	8.09	AAA
10–15 years	7.99	1.36	5.37	8.41	9.85	AAA
5+ years	8.95	0.54	8.03	8.83	10.54	AAA
10+ years	12.09	2.14	9.99	11.01	16.68	AAA
Broad	5.44	0.35	4.69	5.40	6.13	AAA

Panel B. Average duration over time

Sub-period	1–3	1–5	3–5	5–7	5–10	7–15	10–15	5+	10+	Broad
1997-1999	1.66	2.25	3.42	4.72	5.57	6.07	5.67	8.72	10.54	5.20
2000-2002	1.64	2.23	3.35	4.72	5.51	6.48	6.96	8.96	10.73	5.68
2003-2005	1.70	2.36	3.52	4.62	6.00	7.29	8.30	8.95	10.89	5.39
2006-2008	1.71	2.38	3.45	4.83	6.09	7.40	8.56	8.37	10.88	5.07
2009-2011	1.88	2.64	3.75	5.51	6.49	7.79	8.90	8.82	13.35	5.40
2012-2014	1.90	2.72	3.80	5.60	6.45	7.92	9.57	9.90	16.17	5.90

This table shows summary statistics on the durations of ten BofAML US Treasury bond indices in the period from 01/1997 to 12/2014. Panel A displays different statistics in the overall period while Panel B displays statistics over time for different sub-periods. All figures are denoted in years.

Table 2
Regressions of Treasury bond indices

X ↓ *Y*→	1–3	1–5	3–5	5–7	5–10	7–15	10–15	5+	10+
Panel A: Overall period, all market climates									
1–3		-0.2100**	-0.5916**	-0.6241**	-1.0397*	-1.0902	-0.7429	-1.1546	-0.9404
1–5	0.1920***		-0.3406***	-0.3956**	-0.8630*	-0.9658	-0.6798	-1.0547	-0.9663
3–5	0.3868***	0.2492***		-0.0441	-0.4619	-0.5619	-0.2964	-0.6408	-0.5520
5–7	0.4352***	0.3027***	0.0685		-0.4792*	-0.6255	-0.4087	-0.7260	-0.7646
5–10	0.6589***	0.6134***	0.5404**	0.4781***		-0.1876	-0.0418	-0.3084	-0.4784
7–15	0.7514***	0.7442***	0.7411**	0.6837***	0.2175		0.0921	-0.1308	-0.3828
10–15	0.7963***	0.8027***	0.8261**	0.7598**	0.2623	-0.0090		-0.1705	-0.5671
5+	0.8167***	0.8423***	0.8952***	0.8538***	0.4261	0.2166	0.3006		-0.3092
10+	0.9544***	1.0463***	1.2245***	1.1982***	0.8486*	0.6428	0.6793*	0.4312	
Broad Treasury	0.6371***	0.5855***	0.4969**	0.4319*	-0.0880	-0.3235	-0.2275	-0.5654**	-0.8835
Term	1.1345***	1.3201***	1.6677***	1.6931***	1.5535***	1.4515***	1.5432***	1.3509***	1.1345***
Panel B: Market phases with widening term spread									
1–3		-0.6732***	-1.9310***	-2.5174***	-4.9499***	-6.2746***	-6.9057***	-7.6028***	-10.1889***
1–5	0.5050***		-0.9322***	-1.4395***	-3.6055***	-4.8096***	-5.3825***	-5.9987***	-8.3790***
3–5	0.9410***	0.6090***		-0.4195***	-2.2887***	-3.3575***	-3.8650***	-4.4034***	-6.5513***
5–7	1.1132***	0.8550***	0.3844***		-1.7450***	-2.7585***	-3.2424***	-3.7423***	-5.7991***
5–10	1.6059***	1.5749***	1.5274***	1.2778***		-0.7838***	-1.1384**	-1.5492***	-3.2176***
7–15	1.7712***	1.8217***	1.9259***	1.7332***	0.6642***		-0.2720	-0.6646**	-2.1310***
10–15	1.8027***	1.8700***	2.0060***	1.8315***	0.8295**	0.2175		-0.4099	-1.7822***
5+	1.8843***	1.9909***	2.2009***	2.0458***	1.1243***	0.5455**	0.3336		-1.2966***
10+	2.0292***	2.2121***	2.5624***	2.4712***	1.7908***	1.3668***	1.2835***	0.9754***	
Broad Treasury	1.5641***	1.5117***	1.4252***	1.1619***	-0.1598	-0.9608***	-1.3216***	-1.7304***	-3.4222***
Term	2.2433***	2.5616***	3.1607***	3.1855***	2.9508***	2.8041***	2.9371***	2.6868***	2.2433***

Table 2 cont'd.

X ↓ *Y*→	1–3	1–5	3–5	5–7	5–10	7–15	10–15	5+	10+
Panel C: Market phases with contracting term spread									
1–3		0.3773***	1.0557***	1.6494***	3.7043***	5.1678***	6.5734***	6.6660***	10.2137***
1–5	-0.2061***		0.3729***	0.8654***	2.5517***	3.8324***	5.1351***	5.1906***	8.3719***
3–5	-0.2829**	-0.1749**		0.4360***	1.9046***	3.0680***	4.3056***	4.3220***	7.2954***
5–7	-0.3895***	-0.3702***	-0.3433***		1.1808***	2.1826***	3.2988***	3.3145***	5.8359***
5–10	-0.4912**	-0.5900***	-0.7602***	-0.5660**		0.6511***	1.4989***	1.4341***	2.9910**
7–15	-0.4907**	-0.6267**	-0.8575**	-0.7350**	-0.4344**		0.6249**	0.5993*	1.5216
10–15	-0.4260	-0.5613*	-0.7782	-0.7093	-0.5719	-0.3325		0.1088	0.3639
5+	-0.4482*	-0.5681*	-0.7813	-0.6727	-0.5025	-0.1739	0.3286		0.3665
10+	-0.0738	-0.0308	0.0822	0.1689	0.3938	0.5929	0.8005	0.6112	
Broad Treasury	-0.5294**	-0.6321**	-0.8231**	-0.6384**	-0.1648	0.3749	1.0841**	0.8402***	1.9315*
Term	0.1846	0.3169	0.5976	0.6835	0.9455	1.1048	1.2147	1.0683	0.1846

This table shows the alphas from pairwise US Treasury index regressions in the period from 01/1990 to 12/2014. Panel A reports results using all returns during the sample period while Panels B and C split the sample into months with widening and contracting term-spreads. The term spread is defined as the monthly difference between the 10y and the 1y Treasury yields. All figures are denoted in % p.a. *,**,*** represent statistical significance on the 1%, 5% and 10% level, respectively.

Table 3
Return summary statistics for Treasury bond indices

	A. Widening term-spread				B. Contracting term-spread			
	N	Mean	Std.Dev	p50	N	Mean	Std.Dev	p50
1–3	183	1.3690	1.8092	0.7819	153	1.5209	1.2574	1.4456
1–5	183	1.2259	2.5383	0.8398	153	2.5885	1.8562	2.8655
3–5	183	0.9700	3.9647	0.5927	153	4.5305	3.0121	4.9886
5–7	183	0.6506	4.3893	0.3556	153	5.3901	3.3123	5.9965
5–10	183	-0.8737	5.9952	-0.4524	153	8.6352	4.6902	9.2566
7–15	183	-1.7822	6.8990	-1.1534	153	10.5464	5.4141	10.7318
10–15	183	-2.2102	7.5638	-0.9546	153	12.0315	5.9214	11.3210
5+	183	-2.6624	7.8119	-1.8034	153	12.5206	6.2208	13.3792
10+	183	-4.5420	9.6371	-2.6884	153	16.7278	8.4279	16.5168
Broad Treasury	183	-0.5882	4.8884	-0.6980	153	7.0897	3.7320	7.9734

This table shows average US Treasury index returns in the sample period from 01/1990 to 12/2014 split into those months with a widening (Panel A) and a contracting term-spread (Panel B). All figures are denoted in % p.a.

Table 4
Average error terms according to model "Dur-1" from Elton et al. (1988)

X ↓ *Y*→	1–3	1–5	3–5	5–7	5–10	7–15	10–15	5+	10+
Panel A. All market phases, 1990-2014									
1–3		-0.1107	-0.2471	-1.2830***	-1.4285**	-1.8369**	-1.7788*	-3.2128***	-4.4427***
1–5	0.0832		-0.0739	-1.0465***	-1.1495***	-1.5352**	-1.4964*	-2.7829***	-3.9827***
3–5	0.1172	0.0411		-0.9523***	-1.0459***	-1.4540***	-1.4691*	-2.6170***	-3.8568***
5–7	0.4485***	0.4992***	0.6692***		0.0996	-0.1177	-0.0300	-0.9086	-1.6655
5–10	0.4189**	0.4552**	0.6139***	-0.0829		-0.2410	-0.1932	-1.0735***	-1.9736**
7–15	0.4759**	0.5346**	0.7292***	0.0850	0.2025		0.0670	-0.7823**	-1.6126**
10–15	0.4520**	0.5043**	0.6805**	0.0240	0.1291	-0.0745		-0.9000***	-1.7543***
5+	0.6265***	0.7418***	1.0322***	0.5180	0.7191***	0.5931***	0.7024***		-0.6068**
10+	0.7183***	0.8696***	1.2162***	0.7863*	1.0411***	0.9698***	1.1145***	0.4544**	
Broad Treasury	0.5074***	0.5799***	0.7899***	0.1753	0.3098*	0.1335	0.2372	-0.6028***	-1.3311**
Panel B. Widening term spread, 1990-2014									
1–3		-0.7567***	-2.0142***	-3.7271***	-6.5252***	-8.4776***	-9.4794***	-11.3998***	-15.9167***
1–5	0.5499***		-0.8950***	-2.1566***	-4.6433***	-6.2859***	-7.1123***	-8.5824***	-12.2465***
3–5	0.9880***	0.5981***		-0.9134***	-3.1543***	-4.5712***	-5.2918***	-6.3380***	-9.3722***
5–7	1.3016***	1.0391***	0.6317**		-2.0519***	-3.2901***	-3.9025***	-4.7375***	-7.2767***
5–10	1.8975***	1.8638***	1.8471***	1.7072***		-0.8940***	-1.2944**	-1.7290***	-3.2705***
7–15	2.1159***	2.1694***	2.2908***	2.3377***	0.7659***		-0.3173	-0.6211	-1.8068**
10–15	2.1737***	2.2544***	2.4083***	2.5087***	0.9828**	0.2665		-0.3263	-1.3989*
5+	2.2363***	2.3347***	2.5297***	2.6802***	1.1799***	0.4601	0.1611		-1.0538**
10+	2.3948***	2.5571***	2.8491***	3.1376***	1.7392***	1.1119**	0.8785*	0.7982***	
Broad Treasury	1.8443***	1.7922***	1.7341***	1.5562***	-0.1739	-1.1044***	-1.5226***	-1.9833***	-3.6517***

Table 4 contd.

X ↓ *Y*→	1–3	1–5	3–5	5–7	5–10	7–15	10–15	5+	10+
Panel C. Contracting term spread, 1990-2014									
1–3		0.5099***	1.4508***	1.0654***	3.4683***	4.5433***	5.6198***	4.6532***	6.5813***
1–5	-0.3651***		0.7150***	0.0201	2.2073***	3.0292***	3.8992***	2.7891***	3.9570**
3–5	-0.7195***	-0.4942***		-0.9897***	0.9799**	1.5410**	2.2038**	0.9581	1.4422
5–7	-0.3712***	-0.0196	0.7051***		2.1666***	2.9304***	3.6905***	2.7701***	3.7257***
5–10	-1.0016***	-0.8981***	-0.5709**	-1.8028***		0.3865**	0.8648**	-0.4437	-0.7275
7–15	-1.0998***	-1.0362***	-0.7712**	-2.0794***	-0.3387**		0.4361**	-0.9371**	-1.4260
10–15	-1.2021***	-1.1771***	-0.9795**	-2.3633***	-0.6912**	-0.4022**		-1.4512***	-2.0957***
5+	-0.9202***	-0.7887***	-0.4065	-1.5594***	0.2763	0.7209**	1.2225***		-0.1773
10+	-0.8924***	-0.7517**	-0.3526	-1.4728***	0.3704	0.8333*	1.3411***	0.1242	
Broad Treasury	-0.7771***	-0.5849***	-0.1173	-1.1514***	0.7746***	1.3229***	1.9280***	0.7236**	0.8985

This table shows the average pricing errors from pairwise US Treasury Dur-1 calculations following Elton et al. (1988) in the period from 01/1990 to 12/2014. Panel A reports results using all returns during the sample period while Panels B and C split the sample into months with widening and contracting term-spreads. The term spread is defined as the monthly difference between the 10y and the 1y Treasury yields. All figures are denoted in % p.a. *,**,*** represent statistical significance on the 1%, 5% and 10% level, respectively.

Table 5
Bond fund summary statistics

	Average	Std. deviation	p5	p25	p50	p75	p95	Skewness	Kurtosis
Panel A. Government bond funds (127 funds)									
Effective duration (years)	4.12	2.70	1.00	2.11	4.07	5.30	7.68	3.00	24.44
TNA (mil US$)	598.32	2,028.76	9.50	51.16	127.90	403.40	2,156.50	8.20	84.15
Age (years)	8.88	5.91	1.08	4.08	7.83	12.83	20.17	0.63	2.65
Expense Ratio (p.a.)	0.70%	0.37%	0.26%	0.51%	0.68%	0.83%	1.13%	5.06	64.85
Turnover Ratio (p.a.)	154%	236%	13%	42%	90%	180%	480%	9.12	158.91
Implied net flow (p.a.)	1.30%	23.12%	-6.29%	-1.47%	-0.08%	1.70%	10.21%	47.39	3,160.29
Avg.% government bonds	71.63%	16.10%	52.03%	57.54%	67.89%	85.97%	99.37%	0.46	1.90
Avg.% corporate bonds	15.81%	12.24%	0.77%	5.33%	11.67%	25.08%	37.39%	0.51	2.06
Avg.% cash	2.38%	8.09%	-6.07%	1.48%	2.86%	4.72%	12.48%	-2.59	22.85
Panel B. Corporate bond funds (291 funds)									
Effective duration (years)	4.09	1.83	1.40	3.22	4.14	4.82	6.77	2.13	32.23
TNA (mil US$)	560.98	1,591.04	10.20	47.00	133.50	436.00	2,317.70	7.73	80.97
Age (years)	8.30	5.86	0.83	3.42	7.17	12.25	19.42	0.71	2.84
Expense Ratio (p.a.)	0.92%	0.38%	0.34%	0.69%	0.89%	1.11%	1.61%	0.46	3.97
Turnover Ratio (p.a.)	112%	136%	19%	42%	72%	124%	353%	4.63	39.14
Implied net flow (p.a.)	2.95%	89.61%	-6.07%	-1.17%	0.33%	2.63%	12.67%	142.10	23,936.13
Avg.% government bonds	10.85%	12.05%	0.00%	0.02%	5.50%	19.82%	33.40%	0.84	2.49
Avg.% corporate bonds	73.83%	14.74%	51.26%	59.96%	74.53%	88.26%	94.41%	-0.08	1.59
Avg.% cash	5.30%	6.10%	0.33%	2.64%	4.22%	5.56%	16.30%	4.60	35.86

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 01/1990 to 12/2014.

Table 6
Government bond fund performance

	SF-model			MF-model 1			MF-model 2			MF-model 3			MF-model 4		
Benchmark	Alpha	+	–	Alpha	+	–	Alpha	+	–	Alpha	+	–	Alpha	+	–
Panel A. Overall period, all market climates, net returns (% p.a.)															
1–3	-0.3209***	0	32	-1.3747***	0	61	-1.4390***	0	88	-1.4224***	0	86	-1.6043***	0	85
1–5	-0.2275***	5	25	-1.3915***	0	72	-1.3572***	0	82	-1.3625***	0	80	-1.4598***	0	85
3–5	-0.0367	8	14	-1.1696***	1	52	-1.2053***	1	65	-1.1944***	1	65	-1.1944***	1	65
5–7	-0.0752	9	15	-1.1692***	2	56	-1.2009***	1	63	-1.1859***	2	61	-1.1643***	2	60
5–10	0.1058	16	7	-0.7422***	9	29	-0.9116***	5	44	-0.8095***	7	33	-0.7461***	10	29
7–15	0.1799**	18	6	-0.5420***	12	16	-0.7903***	5	38	-0.6787***	7	25	-0.5773***	11	20
10–15	0.1570**	16	7	-0.4582***	13	14	-0.8005***	2	38	-0.6993***	6	25	-0.5390***	8	18
5+	0.2786***	23	6	-0.4100***	15	10	-0.7939***	4	35	-0.6804***	6	23	-0.4808***	13	12
10+	0.4981***	25	2	-0.0708	19	4	-0.7707***	2	30	-0.6545***	6	20	-0.1996*	17	7
Broad Treasury	0.0443	14	11	-0.8287***	6	30	-1.0025***	2	48	-0.9214***	5	42	-0.8653***	6	31
Term	0.8349***	39	2	0.3915***	34	2	-0.6907***	2	23	-0.5716***	6	15	0.2501**	26	3
Duration-adjusted	-0.1053	4	20	-1.0687***	0	59	-1.1558***	0	70	-1.0936***	0	68	-1.0906***	0	65
Panel B. Overall period, all market climates, gross returns (% p.a.)															
1–3	0.3979***	23	4	-0.6558***	8	11	-0.7201***	9	25	-0.7037***	9	21	-0.8857***	8	24
1–5	0.4911***	29	2	-0.6730***	16	13	-0.6387***	16	21	-0.6442***	17	18	-0.7416***	16	21
3–5	0.6817***	43	2	-0.4513***	21	11	-0.4871***	21	15	-0.4765***	23	12	-0.4765***	23	12
5–7	0.6432***	45	2	-0.4510***	21	12	-0.4827***	22	15	-0.4679***	24	14	-0.4463***	24	14
5–10	0.8241***	62	2	-0.0242	44	6	-0.1934*	35	10	-0.0915	37	6	-0.0284	43	7
7–15	0.8981***	64	2	0.1760*	52	2	-0.0721	36	5	0.0394	40	5	0.1404	53	4
10–15	0.8752***	56	2	0.2599***	47	1	-0.0822	32	5	0.0188	34	3	0.1788	46	3
5+	0.9967***	62	1	0.3079***	56	1	-0.0757	34	3	0.0377	37	2	0.2368**	53	3
10+	1.2162***	64	1	0.6471***	64	1	-0.0524	30	2	0.0636	33	2	0.5181***	56	3
Broad Treasury	0.7625***	50	2	-0.1107	39	6	-0.2843**	33	12	-0.2035*	35	6	-0.1475	38	7
Term	1.5529***	79	0	1.1093***	74	1	0.0277	31	2	0.1466	35	2	0.9677***	73	3
Duration-adjusted	0.6133***	41	2	-0.3502***	14	9	-0.4376***	13	17	-0.3755***	14	10	-0.3724***	15	9

Table 6 cont'd.

Benchmark	SF-model Alpha	+	–	MF-model 1 Alpha	+	–	MF-model 2 Alpha	+	–	MF-model 3 Alpha	+	–	MF-model 4 Alpha	+	–
Panel C. Market phases with widening term spread, net returns (% p.a.)															
1–3	-3.4755***	0	94	-4.6153***	0	115	-2.6695***	0	99	-2.7465***	0	97	-3.6850***	0	113
1–5	-2.5552***	2	66	-3.4879***	2	101	-2.3731***	1	84	-2.4394***	0	86	-2.7735***	1	92
3–5	-1.7889***	10	42	-2.4580***	9	80	-1.9110***	9	67	-1.9028***	9	69	-1.9028***	9	69
5–7	-1.4616***	13	34	-1.9422***	12	66	-1.7147***	13	63	-1.6509***	13	62	-1.5498***	12	58
5–10	-0.3576**	27	9	-0.4646***	25	13	-0.7624***	24	19	-0.4883***	26	16	-0.3546**	25	9
7–15	0.0814	32	5	0.0536	30	5	-0.3677**	28	7	-0.0861	31	6	0.0819	32	4
10–15	0.2289*	28	4	0.2011	30	4	-0.3098**	27	8	-0.0884	28	4	0.1854	31	4
5+	0.3643**	39	3	0.4227***	40	3	-0.2410*	28	6	0.0287	27	4	0.4400***	41	3
10+	0.6959***	40	0	0.8508***	44	1	-0.1007	25	4	0.0893	25	3	0.8437***	43	1
Broad Treasury	-0.3923**	26	10	-0.4980***	24	13	-0.8147***	23	22	-0.5562***	23	17	-0.3984**	23	11
Term	1.2597***	51	0	1.4409***	56	1	0.0057	24	3	0.1316	25	1	1.4256***	54	0
Duration-adjusted	-1.1213***	0	38	-1.4514***	0	61	-1.3613***	0	53	-1.2094***	0	52	-1.1833***	0	54
Panel D. Market phases with contracting term spread, net returns (% p.a.)															
1–3	2.4371***	76	4	1.7895***	56	6	0.0591	0	18	0.2457*	4	16	0.8734***	21	9
1–5	1.8110***	63	7	0.9928***	29	10	0.0442	3	17	0.2078	3	16	0.3638***	5	14
3–5	1.5046***	57	10	0.5869*	11	14	-0.0810	3	21	0.0517	3	19	0.0517	3	19
5–7	1.1349***	43	12	0.0909	4	19	-0.1706*	1	24	-0.1192	3	25	-0.2069**	1	25
5–10	0.4934***	15	14	-0.5303***	2	27	-0.4469***	1	38	-0.5235***	1	38	-0.4799***	0	29
7–15	0.3043**	11	19	-0.6758***	0	28	-0.6930***	1	46	-0.7946***	1	42	-0.5407***	0	26
10–15	0.2144	10	18	-0.6367***	0	17	-0.8789***	0	49	-0.9610***	0	42	-0.5209***	0	19
5+	0.2446	11	24	-0.6303***	0	25	-0.8487***	0	51	-0.9241***	0	44	-0.4778***	0	26
10+	0.8056***	20	18	0.1177	2	16	-0.8694***	0	46	-0.8063***	0	38	0.1578	2	13
Broad Treasury	0.3855**	14	23	-0.6486***	1	36	-0.5761***	1	44	-0.7025***	1	45	-0.5272***	0	34
Term	1.0349***	21	10	0.5181***	8	6	-0.8677***	0	42	-0.7515***	0	35	0.4956***	4	8
Duration-adjusted	0.7220***	37	6	-0.3045**	5	12	-0.3187***	0	25	-0.3647***	2	22	-0.3097***	1	20

This table shows the performance of active US domestic government bond funds in the period from 01/1990 to 12/2014 where performance is measured using different models and against different benchmark indices. +, - represent the counts of funds with at the 5% level positively and negatively significant alphas. *,**,*** represent statistical significance on the 1%, 5% and 10% level, respectively.

Table 7
Duration bias of government bond fund groups

X ↓ *Y*→	1–3 (24 funds)	1–5 (10 funds)	3–5 (25 funds)	5–7 (37 funds)	5–10 (20 funds)	7–15 (7 funds)
Panel A. Overall period, MF-model 4 (% p.a.)						
1–3		-0.2330***	-0.4408***	-0.5715***	-1.0762***	-1.1361***
1–5	0.1506***		-0.2639***	-0.3577***	-1.0211***	-1.0425***
3–5	0.3069***	0.2119***		-0.0421***	-0.7063***	-0.7492***
5–7	0.3566***	0.2525***	0.0442***		-0.6674***	-0.7358***
5–10	0.5895***	0.4762***	0.4453***	0.4565***		-0.2303***
7–15	0.6772***	0.5341***	0.5992***	0.6350***	0.2866***	
10–15	0.7199***	0.5077**	0.6313***	0.6578***	0.3864***	0.0554
5+	0.7413***	0.5500**	0.7257***	0.7546***	0.4358***	0.1388**
10+	0.9115***	0.6850**	0.9273***	1.0641***	0.8942***	0.4791**
Broad Treasury	0.5520***	0.3454**	0.4015***	0.3330***	-0.1248***	-0.3095***
Term	1.0958***	0.9237***	1.3126***	1.5661***	1.5009***	0.9922***
Panel B. Market phases with widening term spread, MF-model 4 (% p.a.)						
1–3		-0.6729***	-1.6833***	-2.3637***	-4.9163***	-4.7884***
1–5	0.4684***		-0.8422***	-1.3287***	-3.6907***	-3.7475***
3–5	0.8688***	0.5379***		-0.3719***	-2.4672***	-2.6979***
5–7	1.0198***	0.7453***	0.2516***		-1.8704***	-2.2196***
5–10	1.4907***	1.4060***	1.2516***	1.2908***		-0.6183***
7–15	1.6452***	1.6100***	1.5759***	1.7488***	0.7062***	
10–15	1.6811***	1.5591***	1.6068***	1.8138***	0.9517***	0.2042
5+	1.7657***	1.8024***	1.8519***	2.1793***	1.2121***	0.4552***
10+	1.8807***	1.8568***	2.0804***	2.5648***	1.9022***	1.0631***
Broad Treasury	1.4663***	1.3657***	1.1861***	1.2599***	-0.0290	-0.6395***
Term	2.0453***	2.1364***	2.5625***	3.2403***	2.6414***	1.8012***

Table 7 contd.

X ↓	*Y*→	1–3	1–5	3–5	5–7	5–10	7–15
Panel C. Market phases with contracting term spread, MF-model 4 (% p.a.)							
1–3			0.1765***	0.6939***	1.1275***	2.3378***	1.8159***
1–5		-0.1739***		0.2868***	0.6014***	1.4400***	1.2106***
3–5		-0.2727***	-0.1160**		0.2798***	0.9012***	0.8944***
5–7		-0.3162***	-0.2510***	-0.1178***		0.4755***	0.5402***
5–10		-0.2774***	-0.4487***	-0.1874***	-0.2828***		0.0783*
7–15		-0.2096**	-0.5314**	-0.1991***	-0.3148***	-0.1285***	
10–15		-0.0798	-0.5510**	-0.1594**	-0.2677***	-0.1413***	-0.0130
5+		-0.0873	-0.5674**	-0.1292*	-0.2308***	0.0221	0.0809
10+		0.3454	-0.0747	0.2531**	0.4556**	1.0990***	0.7891*
Broad Treasury		-0.2314**	-0.6104***	-0.2465***	-0.3139***	0.0190	0.1255
Term		0.5499**	0.2184	0.5568***	0.8146***	1.5574***	1.1072*

This table shows duration biases of active US domestic government bond funds in the period from 01/1990 to 12/2014 measured by the average alpha differences between the indices in the rows and the respective duration-adjusted benchmarks of the columns. Panel A reports results using all returns during the sample period while Panels B and C split the sample into months with widening and contracting term-spreads. The term spread is defined as the monthly difference between the 10y and the 1y Treasury yields. All figures are denoted in% p.a. *,**,*** represent statistical significance on the 1%, 5% and 10% level, respectively.

Table 8
Corporate bond fund performance

	A. Overall period, net			B. Overall period, gross			C. Widening term spread			D. Contracting term spread		
Benchmark	MF4-Alpha	+	–	MF4-Alpha	+	–	MF4-Alpha	+	–	MF4-Alpha	+	–
1–3	-0.7519***	19	81	0.1379	54	26	-1.7895***	16	114	0.6068***	54	14
1–5	-0.5917***	22	73	0.2979***	67	16	-0.9874***	24	76	0.2039*	33	19
3–5	-0.3791***	30	57	0.5101***	84	12	-0.3390***	33	44	-0.0277	26	26
5–7	-0.3335***	32	53	0.5557***	84	10	-0.0318	44	34	-0.2165*	19	35
5–10	-0.0096	44	34	0.8796***	111	8	0.8234***	72	7	-0.2514*	23	34
7–15	0.0770	45	25	0.9661***	123	9	1.1210***	94	5	-0.2774**	23	34
10–15	0.0767	46	24	0.9658***	121	8	1.1398***	95	5	-0.2469	28	28
5+	0.1084	47	20	0.9975***	127	6	1.3086***	102	2	-0.2134	27	30
10+	0.3604***	60	14	1.2494***	138	6	1.4907***	104	2	0.6002***	53	19
Broad Treasury	-0.1676*	43	39	0.7215***	104	11	0.7515***	62	7	-0.3340**	23	41
Term	0.6810***	83	10	1.5699***	163	6	1.8321***	120	0	1.0030***	74	14
Duration-adjusted	-0.3877***	28	56	0.5015***	76	9	-0.1046	29	44	-0.1573	25	27

This table shows the performance of active US domestic corporate bond funds in the period from 01/1990 to 12/2014 where performance is measured against different benchmark indices. +, - represent the counts of funds with at the 5% level positively and negatively significant alphas. *,**,*** represent statistical significance on the 1%, 5% and 10% level, respectively.

Table 9
Duration bias of corporate bond fund groups

X ↓ *Y*→	1–3 (52 funds)	1–5 (19 funds)	3–5 (110 funds)	5–7 (72 funds)	5–10 (24 funds)
Panel A. Overall period, MF-model 4					
1–3		-0.2269***	-0.3292***	-0.4783***	-1.1601***
1–5	0.1366***		-0.1939***	-0.3105***	-0.8750***
3–5	0.2247***	0.1899***		-0.0526***	-0.4916***
5–7	0.2428***	0.2086***	0.0615***		-0.4283***
5–10	0.3257***	0.3852***	0.4350***	0.4183***	
7–15	0.3254***	0.4083**	0.5539***	0.5534***	0.0775
10–15	0.2977***	0.3972*	0.6055***	0.5688***	-0.0022
5+	0.3165***	0.4545*	0.6075***	0.6475***	0.0829
10+	0.3965***	0.7526***	0.9544***	0.9063***	0.3038
Broad	0.2421***	0.2442	0.3079***	0.3063***	-0.2881***
Term	0.4886***	0.9913***	1.2619***	1.2968***	0.8026***
Panel B. Market phases with widening term spread, MF-model 4					
1–3		-0.8220***	-1.4221***	-1.9958***	-3.9647***
1–5	0.3796***		-0.6222***	-1.1298***	-2.7467***
3–5	0.6047***	0.5361***		-0.3251***	-1.7160***
5–7	0.7166***	0.8105***	0.3310***		-1.2851***
5–10	0.9750***	1.4706***	1.1797***	1.0473***	
7–15	1.0406***	1.7011***	1.4643***	1.4158***	0.4610***
10–15	0.9874***	1.6859***	1.4822***	1.4716***	0.4945***
5+	1.0509***	1.8114***	1.6185***	1.6959***	0.8037***
10+	0.9890***	1.9036***	1.7464***	1.9836***	1.1667***
Broad	0.9180***	1.3750***	1.1027***	0.9918***	-0.0697**
Term	1.0432***	2.1326***	2.0046***	2.4639***	1.7583***
Panel C. Market phases with contracting term spread, MF-model 4					
1–3		0.3167***	0.7189***	0.8834***	1.1601***
1–5	-0.0617		0.2595***	0.4237***	0.6165***
3–5	-0.0967	-0.1602**		0.1377***	0.3547**
5–7	-0.2168***	-0.4210***	-0.1846***		0.1195
5–10	-0.3921***	-0.5535**	-0.1077	0.0642	
7–15	-0.5435***	-0.5654	-0.0530	0.0938	-0.0804
10–15	-0.7078**	-0.4413	0.0723	0.1650	-0.0657
5+	-0.5950***	-0.2834	0.1331	0.1424	-0.0599
10+	-0.2026	0.8369	1.1192***	0.8500***	0.8558***
Broad	-0.5294***	-0.4848	-0.0720	-0.0456	-0.1838**
Term	0.0484	1.3063*	1.5319***	1.2562***	1.3513***

This table shows duration biases of active US domestic corporate bond funds in the period from 01/1990 to 12/2014 measured by the average alpha differences between the indices in the rows and the respective duration-adjusted benchmarks of the columns. Panel A reports results using all returns during the sample period while Panels B and C split the sample into months with widening and contracting term-spreads. The term spread is defined as the monthly difference between the 10y and the 1y Treasury yields. All figures are denoted in% p.a. *,**,*** represent statistical significance on the 1%, 5% and 10% level, respectively.

Table 10
Explaining alpha performance with term structure changes

	Broad Treasury index				Duration-adjusted benchmark			
	Level	Slope	Curvature	Combined	Level	Slope	Curvature	Combined
I. Government bond funds								
A. SF-model								
dL : DG	-0.0012***			-0.0011***	0.0013			0.0013
	(-3.31)			(-3.15)	(1.06)			(0.99)
dS : DG		0.0124***		0.0086***		-0.0056		0.0085
		(3.91)		(7.17)		(-0.73)		(1.22)
dC : DG			0.0016	-0.0001***			-0.0023	-0.0018
			(1.44)	(-3.04)			(-1.12)	(-0.97)
Adj. R^2	0.02	0.02	0.00	0.02	0.00	0.00	0.00	0.00
N	20,723	20,723	20,723	20,723	20,723	20,723	20,723	20,723
B. MF4-model								
dL : DG	-0.0014***			-0.0014***	0.0003			0.0003
	(-3.80)			(-3.67)	(0.29)			(0.22)
dS : DG		0.0162***		0.0124***		-0.0031		0.0098
		(4.56)		(6.46)		(-0.35)		(1.34)
dC : DG			0.0021*	0.0000			-0.0032	-0.0036*
			(1.86)	(0.01)			(-1.32)	(-1.82)
Adj. R^2	0.03	0.03	0.01	0.04	0.00	0.00	0.00	0.00
N	20,723	20,723	20,723	20,723	20,723	20,723	20,723	20,723
II. Corporate bond funds								
A. SF-model								
dL : DG	0.0034***			0.0033***	0.0034**			0.0034**
	(2.89)			(2.89)	(2.26)			(2.27)
dS : DG		-0.0255**		0.0147		-0.0278		-0.0175
		(-2.13)		(1.15)		(-1.40)		(-0.84)
dC : DG			-0.0085**	-0.0067**			-0.0016	0.0003
			(-2.52)	(-2.25)			(-0.35)	(0.07)
Adj. R^2	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.00
N	36,844	36,844	36,844	36,832	36,844	36,844	36,844	36,832
B. MF4-model								
dL : DG	-0.0015***			-0.0015***	-0.0001			-0.0001
	(-5.29)			(-5.37)	(-0.11)			(-0.22)
dS : DG		0.0113***		0.0061**		-0.0069		-0.0143**
		(3.85)		(2.20)		(-0.94)		(-2.31)
dC : DG			0.0008	-0.0009			-0.0005	0.0001
			(0.79)	(-1.10)			(-0.29)	(0.1)
Adj. R^2	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00
N	36,844	36,844	36,844	36,832	36,844	36,844	36,844	36,832

This table shows pooled panel regressions of bond fund alphas using the broad Treasury index and the duration-adjusted benchmark on interactions between term structure changes and the duration gap between the index and the fund for active US domestic bond funds in the period from 01/1990 to 12/2014. All variables are calculated in overlapping 12-month rolling windows. For the combined explanation models, we use democratic, simultaneous procedure by Klein and Chow (2013). Term structure parameters level, slope, and curvature are defined following Wilkens (1994). *,**,*** represent statistical significance on the 1%, 5% and 10% level, respectively. Standard errors are calculated using two-dimensionally clustered standard errors (Petersen, 2009).

Table 11
Performance persistence

I. Government bond funds (127 funds)

	Broad index		Term factor		Duration-adjusted (Average)		Duration-adjusted (Monthly)	
A. Fama and McBeth (1973) regressions								
	Coef.	R^2	Coef.	R^2	Coef.	R^2	Coef.	R^2
MF1-model	0.3562**	0.22	0.3361**	0.20	0.2406**	0.18	0.1967*	0.14
MF2-model	0.2790**	0.23	0.3915***	0.26	0.1733	0.21	0.1734	0.20
MF3-model	0.2333**	0.19	0.3478***	0.21	0.1024	0.17	0.1149	0.15
MF4-model	0.1553	0.20	0.1725*	0.15	0.0806	0.15	0.0754	0.14
B. Spearman rank correlations								
MF1-model	0.1915**		0.2203***		0.1228*		0.0826	
MF2-model	0.1994**		0.2253***		0.1133		0.1213*	
MF3-model	0.1239		0.1609**		0.0594		0.0808	
MF4-model	0.1323		0.1556**		0.0618		0.0619	
II. Corporate bond funds (291 funds)								
A. Fama and McBeth (1973) regressions								
	Coef.	R^2	Coef.	R^2	Coef.	R^2	Coef.	R^2
MF1-model	0.1926***	0.10	0.1996***	0.09	0.1681***	0.09	0.1707***	0.07
MF2-model	0.1889***	0.08	0.1742***	0.08	0.1536**	0.08	0.1535***	0.06
MF3-model	0.2123***	0.10	0.2117***	0.10	0.1537***	0.07	0.1596***	0.06
MF4-model	0.1664***	0.07	0.1986***	0.08	0.1500***	0.06	0.1420***	0.05
B. Spearman rank correlations								
MF1-model	0.1731***		0.1485***		0.1454***		0.1380***	
MF2-model	0.1717***		0.1438***		0.1345***		0.1343***	
MF3-model	0.2027***		0.1799***		0.1376***		0.1296***	
MF4-model	0.1748***		0.1646***		0.1410***		0.1181***	

This table shows the results of Fama and MacBeth (1973) regressions of contemporary alpha on lagged alpha (Panels A) and respective Spearman rank correlations (Panels B) for active US domestic government (I) and corporate bond funds (II). Fund alphas are calculated for non-overlapping 12-month rolling windows in the sample period from 01/1990 to 12/2014. *,**,*** represent statistical significance on the 1%, 5% and 10% level, respectively.

Table 12
Prospectus benchmarks

Panel A. Number of government bond funds (105 funds)

Maturity range from prospectus benchmark	Duration-adjusted benchmark group								
	1–3y	1–5y	3–5y	5–7y	5–10y	7–15y	10–15y	10y+	Total
3 months	2								2
Below 1y	2								2
1–3y	16		1						17
Short-term			1						1
1–5y	1	6		1					8
Intermediate, broad or unspecified	3	1	19	27	16	5			71
5–10y				1					1
Long-term								1	1
10y+								2	2
Matching	20	7	19	27	16	0	0	3	91
Total	24	7	21	29	16	5	0	3	105

Panel B. Number of corporate bond funds (268 funds)

Maturity range from prospectus benchmark	Duration-adjusted benchmark group									
	1–3y	1–5y	3–5y	5–7y	5–10y	7–15y	10–15y	5y+	10y+	Total
3 months	7		1							8
Below 1y	7									7
1–3y	22			1						23
Short-term	2									2
1–5y	9	7	2							18
Intermediate, broad or unspecified	3	11	104	57	19	3	4	1		202
5–10y				1						1
Long-term						1		1	4	6
10y+	1									1
Matching	38	7	104	57	19	1	0	1	4	231
Total	51	18	107	59	19	4	4	2	4	268

This table shows the number of active US domestic government (Panel A) and corporate bond funds (Panel B) sorted into groups according to their duration-adjusted benchmark (columns) and having prospectus benchmarks with specific maturity ranges (rows). The shaded areas indicate an approximate match between duration-adjusted and self-stated benchmarks according to maturity range.

5 Article IV: Option-based benchmark indices – A review of performance and (in)appropriate measures

Markus Natter[a]

A revised version of this paper will be published in the Journal of Futures Markets.
(doi: 10.1002/fut.21865)

VHB-Jourqual 3: B

Abstract. This paper reviews the performance and profitability of different option strategy benchmark indices provided by the CBOE. Using different approaches to measure performance, this study shows that performance measurement of these indices is highly complex and sensitive to the model choice. I am the first to control for time-varying delta exposure and develop a novel linear option-factor model to catch the inherent option exposure adequately. Splitting the sample, I find that outperformance found by previous studies is mostly driven by limited data sets. Moreover, the profitability of option strategies for private investors is evaluated on the basis of investment products, which are easily investable at low cost.

JEL Classification: G11, G20, G23

Keywords: performance, options, hedging, option strategy benchmark indices.

5.1 Introduction and literature overview

In 2002, the CBOE introduced the first option-based benchmark strategy index – the BuyWrite® Index (BXM). Since then, the palette of these indices has experienced dramatic growth. More than twenty strategy indices with different underlying indices and different properties are available on CBOE's homepage.[25] Scientific researchers and practitioners both analyze the performance of these hypothetical portfolios, especially for option-writing indices.[26] This study reviews the performance of strategy benchmark indices using common and novel approaches to measuring profitability. To the best of my knowledge, I am the first to measure the performance of a large number of benchmark indices instead of exclusively one strategy. Moreover, this paper uses a long time horizon as well as different time windows for each strategy index, and introduces novel approaches to measure the performance of portfolios containing options.

So far, there are many studies attesting superior performance for CBOE's strategy benchmark indices (e.g., Whaley, 2002 and Ungar and Moran, 2009). The first index published was the BXM BuyWrite Index, which is a simple passive covered call strategy that is long the S&P 500 and sells one-month call options on the underlying index. The original purpose of the BXM was to provide a sufficient benchmark for investors whose portfolios contain options.[27] Whaley (2002) describes the construction of this index and finds more than 20 basis points outperformance on a monthly and risk-adjusted basis compared to the S&P 500.[28] At first glance, this result seems surprising since the BXM strategy theoretically solely invests in the S&P 500 and the risk-free rate, because the replicating portfolio of a short call option consists

[25] http://www.cboe.com/products/strategy-benchmark-indexes.

[26] A collection of contributions on option strategy benchmark indices can be found at http://www.cboe.com/products/strategy-benchmark-indexes/bibliography.

[27] Natter et al. (2016) construct a risk-factor derived from BXM's return.

[28] The annualized outperformance is therefore 2.76% p.a.

of a short position in the respective underlying and a long position in a zero bond. Following a covered call strategy, therefore means being long the underlying and simultaneously selling some part of the same underlying, whereas the remaining amount of money is invested in a riskless bond. An investment in an index mixed with the risk-free rate should not generate any risk-adjusted outperformance. Nonetheless, there is a vast stream of literature showing the profitability of option writing strategies in the past. Among others, prominent representatives are the studies of Pounds (1978) as well as Bookstaber and Clarke (1981, 1984 and 1985). Several studies also find outperformance for CBOE's covered call strategies. Feldman and Roy (2005), for example, report an annualized Jensen's alpha of almost 3% p.a. for the BXM over a 16-year period. Kapadia and Szado (2007) analyze a similar strategy with the Russell 2000 as the underlying index and report an annualized alpha of 2%. Ungar and Moran (2009) measure the performance of the CBOE PutWrite index whose payoff is the analog to a covered call strategy and find an outperformance of over 6% p.a. on a risk-adjusted basis.

This paper contributes to the literature on performance measurement of option-based strategies as follows. (i) So far, most of the recently introduced indices have not been analyzed in detail. (ii) This paper applies and discusses more accurate methods to measure the performance of portfolios containing options by allowing time-varying delta exposure in linear models.[29] (iii) The volatility as the most important determinant of option values is considered in performance measurement via different novel option-factor approaches. (iv) By splitting my sample into two separate parts and show that outperformance documented by previous studies is driven by the first half (1990 to 2003) of observations.[30] In addition (v), I examine the profitability of option strategies in different crisis scenarios. While Natter et al. (2016) find superior performance

[29] Israelov and Nielsen (2015) and Israelov and Klein (2016) develop an approach to decompose the return of option strategies adequately and consider time-varying delta exposure as well. However, their approach requires the knowledge of the exact portfolio composition and data on deltas of actually traded options. The models in this paper shall be translatable to any portfolio return time-series.
[30] Constantinides et al. (2009) find overpricing in S&P 500 options in this period.

among actively managed equity funds investing in options; I examine whether this result is transferable to passively managed investments and evaluate benefits of direct investments for private investors (vi). Overall, I review and validate previous studies and critically assess their results.

The remainder of the paper is organized as follows. Section 5.2 introduces the data. Section 5.3 describes the performance models and Section 5.4 reports the results of the empirical analysis. In Section 5.5, I highlight the possibility to gain exposure towards option strategy benchmark indices. Section 5.6 concludes.

5.2 Data and index description

CBOE's strategy benchmark indexes analyzed in this paper can be distinguished into six subgroups[31]:

BuyWrite Indexes: A BuyWrite Index or covered call strategy is a passive hypothetical strategy that is long a specific underlying and writes call options on that underlying. Hence, these call options consequently are considered to be covered. The indices of this subgroup differ in their underlyings as well as in the characteristics of the written call options.

PutWrite Indexes: A Putwrite Index writes put options on different underlyings, whereas proceeds of the received option premia are invested into riskless T-bills. Hence, the put options are cash-secured and therefore also covered. The payoff of this investment strategy is similar to that of a covered call strategy.

Combo Index: The CBOE S&P 500 Covered Combo Index (CMBO) combines BuyWrite and PutWrite strategies. The index sells one-month out-of-the-money call options and is simultaneously short out-of-the-money cash-secured put options. The result is a payoff that is

[31] A detailed description of the construction of benchmark indices can be found on CBOE's homepage.

similar to those of the individual strategies but steeper at the beginning compared to Buywrite and Putwrite Indexes.

Butterfly and Condor Indexes: The Iron Butterfly Index (BFLY) sells at-the-money call and put options and buys out-of-the-money call and put options. The Iron Condor Index strategy (CNDR) is both short and long out-of-the-money calls and puts with different deltas. The short puts and calls exhibit deltas of about -0.2 and 0.2 and the long options have deltas of -0.05 and 0.05, respectively.

Collar Indexes: The portfolio of a collar index strategy consists of a long position in the underlying S&P 500 index and a protective out-of-the-money put as well as a written out-of-the-money call. The indexes differ in the characteristics of the options, i.e. the portfolios can, for example, be set up at no cost.

Put Protection Index: The Put Protection Index (PPUT) is a passive option benchmark index that follows a simple protective put strategy as proposed by Merton, Scholes and Gladstein (1978 and 1982). The strategy is long the S&P 500 index and buys 5% out-of-the-money put options to limit the potential downside risk of the long position in the equity index.[32]

Historical monthly option benchmark index prices are available with CBOE's website[33]. Since historical data begin and end at different dates, I limit the sample period to February 1990 to the end of 2016 to obtain a comprehensive but comparable sample. Consistent with the literature on performance measurement of option-based benchmark indices, I compare the option strategies to the underlying index and investment products on benchmark indices (see, e.g., Whaley, 2002, Ungar and Moran, 2009). Since I consider only option strategies on the S&P 500, the S&P 500 total return index (SP500TR) serves as my primary benchmark. The data on

[32] For a graphical illustration of the static payoff profiles for all option strategy benchmark indices see Appendix A.
[33] https://www.cboe.com/micro/buywrite/

the S&P 500 total return index is also provided by CBOE, while data on retail investor products stems from Morningstar Direct.

5.3 Methodology

The baseline model for the performance analysis is the CAPM regression and the resulting Jensen Alpha (Jensen, 1968).

$$r_{i,t} - r_f = \alpha_i + \beta_i(r_{SPTR,t} - r_{f,t}) + \varepsilon_{i,t} \tag{1}$$

where $r_{i,t}$ is the return of option strategy *i* and r_f is the risk-free rate. Since all of these indices are constructed from the S&P 500 total return index (SPTR), this index is the corresponding market index in the performance regression. α_i is Jensen's Alpha and β_i is the systematic risk or can be interpreted as the delta of the option strategy, respectively, i.e., β_i measures the sensitivity of option strategies towards the underlying.

Among others, Lhabitant (2000) and Goetzmann et al. (2007) show that applying risk-adjusted performance measures on option-based investments may lead to potential biases in these figures since portfolios containing options exhibit asymmetric return distributions with fatter tails than the normal. Leland (1999) shows that dynamic strategies consisting of stocks and bonds may bias performance measures, which are grounded in a mean-variance framework. Since the replicating portfolio of an option is a self-financing dynamic strategy consisting of stocks and bonds by definition (Black and Scholes, 1973), even performance measures for buy and hold strategies involving options can be misleading.

Israelov and Nielsen (2015) develop an approach, which accounts for the time-varying character of delta exposure into and is therefore able to decompose the option index return in single constituents adequately. Their proceeding is very powerful and attributes the performance of an investment in connection with options correctly. However, contrary to this paper, their approach requires deltas for the actual options included in the respective portfolio. The models presented here are supposed to be applicable to all portfolios containing options without knowing the exact portfolio holdings positions. To discover the necessity for allowing time-varying systematic risk, consider a simple plain vanilla S&P 500 index call option. The replicating portfolio consists of a long position in delta shares of the S&P 500 and a short position in a zero bond. When the price of the underlying S&P 500 drops dramatically, the share of stocks in the replicating portfolio diminishes as well. It is, therefore, likely that assuming constant betas over time might lead to biased estimators for both performance and systematic risk respectively for the delta of the strategies. Delta indicates the sensitivity of an option or an option strategy to changes in prices of the underlying. Since the underlying of all option strategy benchmark indices is the S&P 500, the beta obtained from a CAPM regression can be interpreted as the option strategy's delta. This leads to the necessity of considering time-varying beta factors in performance models.

Jagannathan and Korajczyk (1986) criticize known timing approaches used in mutual fund research and show both theoretically and empirically that these models may detect spurious timing abilities if portfolios contain options. Conversely, this means that these approaches are able to detect option exposure in portfolios and hence, I employ timing models to measure performance of option strategies. One approach to allow for time-varying betas is similar to the timing model introduced by Treynor and Mazuy (1966) where beta is a linear function of the market return.

$$r_{i,t} - r_f = \alpha_{TM,i} + \beta_{TM,i,t}(r_{SPTR,t} - r_{f,t}) + \varepsilon_{i,t} \qquad (2)$$

with
$$\beta_{TM,i,t} = b_i + \gamma_i(r_{SPTR,t} - r_{f,t}) \qquad (2.1)$$

The slope coefficients of this regression have implications similar to the greeks of the Black/Scholes/Merton model. $\beta_{TM,i,t}$ can be interpreted as the index' overall delta. γ_i indicates the gamma of the strategies' delta, i.e., the sensitivity of delta to changes in the underlying S&P 500.

Another approach including time-varying betas is the Henriksson and Merton (1981) model. This approach is different to the Treynor and Mazuy procedure since beta can only adopt two distinct values. More specifically, the time-varying systematic risk is a function of an option on the market, where $\mathbb{1}$ is an indicator function that is one if the market excess return is positive and zero otherwise.

$$r_{i,t} - r_f = \alpha_{HM,i} + \beta_{HM,i,t}(r_{SPTR,t} - r_{f,t}) + \varepsilon_{i,t} \tag{3}$$

with
$$\beta_{HM,i,t} = b_{\ i} + \gamma_i \mathbb{1}_{\{r_{SPTR,t} - r_{f,t} > 0\}} \tag{3.1}$$

Options added to portfolios generate asymmetric return distributions with higher moments different from those of the normal distribution. Hence, several new approaches have been proposed by researchers to control for this problem and avoid biases in performance measures.

Leland (1999) argues that dynamic strategies generating skewed return distributions can bias performance measures. In particular, left skewed returns lead to positively biased performance measures. Contrary to that, right skewed returns lead to negatively biased two-dimensional performance figures. To control for skewness, kurtosis and any higher moments of the return distribution, I compute Leland's Alpha for all benchmark indices (see Leland, 1999).

$$\alpha_{L,i} = E(r_i) - B_{L,i}\left[E(r_{SPTR}) - r_f\right] - r_f \tag{4}$$

where
$$B_{L,i} \frac{cov\left(r_i; -(1 + r_{SPTR})\right)^{-b}}{cov\left(r_{SPTR}; -(1 + r_{SPTR})\right)^{-b}} \tag{4.1}$$

with
$$b = \frac{ln\left(E(1 + r_{SPTR})\right) - ln\left(1 + r_f\right)}{Var\left(ln(1 + r_{SPTR})\right)} \tag{4.2}$$

Following Whaley (2002), I additionally use his methodology to determine the risk-adjusted performance of option-based strategy benchmark indexes. Whaley's alpha exclusively considers downside risk to calculate systematic risk.

$$min(r_{i,t} - r_{f,t}, 0) = \alpha_{W,i} + \beta_{W,i} min(r_{SPTR,t} - r_{f,t}, 0) + \varepsilon_{i,t} \qquad (5)$$

In complete markets, option strategies should not be more profitable on a risk-adjusted basis than investments in the underlying. However, there is extant literature that options are richly priced (e.g. Constantinides et al., 2009 and Chambers et al., 2014). To disentangle falsely attributed performance not only due to non-linearities but also resulting from overpriced options, I employ the following option-factor model approach inspired by the hedge fund related work of Agarwal and Naik (2004).[34] I augment the CAPM regression equation with a straddle-factor[35] to capture non-linear risk exposure that comes from non-linear payoff profiles of options. The option strategy factor is the return time-series of buying a one-month at-the-money straddle and holding it until expiration. The straddle-factor is then computed as the discrete return of this buy and hold strategy in excess of the risk-free rate.

I calculate option prices using the Black/Scholes/Merton formula using the implied volatility of actually traded S&P 500 index options derived from the Volatility Index (VIX) to obtain a sufficient proxy for option prices.

$$r_{i,t} - r_f = \alpha_i + \beta_{i,t}(r_{SPX,t} - r_{f,t}) + \beta_{straddle,i} straddle_t + \varepsilon_{i,t} \qquad (6)$$

with $$straddle_t = \frac{max(SPX_t - SPX_{t-1}; 0) + max(SPX_{t-1} - SPX_t; 0)}{call\_price_{t-1} + put\_price_{t-1}} - 1 - r_f \qquad (6.1)$$

[34] Fung and Hsieh (2004) employ a look-back-straddle strategy as a trend-following factor for hedge funds.
[35] An illustration of the option-factor's payoff can be found in Appendix A: Exhibit 8

The straddle-factor model might catch biases arising from asymmetric return distributions as well as from potentially overpriced options.

Many of CBOE's benchmark indices invest in options to hedge portfolio risk or earn premia. The income received from writing options should protect the portfolio from sharp declines in crisis periods. I run an additional performance regression with a crisis dummy variable that equals one if a crisis took place in a given month and zero otherwise. I determine different crisis definitions, for which a detailed description can be found in Section 5.4.5.

$$\begin{aligned} r_{i,t} - r_f = \ & \alpha_i + \beta_{M,i}(r_{SPTR,t} - r_{f,t}) + \ \beta_{C,i} crisis_t \\ & + \beta_{C,M,i}\left(crisis_t * \left(r_{SPTR,t} - r_f\right)\right) + \varepsilon_{i,t} \end{aligned} \tag{7}$$

5.4 Empirical results

5.4.1 Descriptive Statistics

In a first step, I calculate summary statistics for each group of option strategy indices. Table 1 shows the results.

The discrete returns of the S&P 500 total return index (SPTR) are slightly skewed to the left and a kurtosis in the amount of 4.2162 reflects fatter tails than the normal. The figures of this index serve as standard of comparison for option strategy indices. Option-writing indices, namely BXM, PUT and CMBO, have average returns similar to the underlying SPTR. However, risk in terms of volatility or semi-deviation, respectively, is considerably lower for option strategies. This leads to a finding in line with the majority of previous studies, namely that these indices outperform on a risk-adjusted basis. The Sharpe Ratio (Sharpe, 1966) is higher for option writing indices than for the underlying index. Other multidimensional performance measures, e.g. the Sortino Ratio (Sortino and Price, 1994) or the Stutzer Index (Stutzer, 2000), deliver similar results. The payoffs for BXM and PUT exhibit capped upside potentials and so the skewness is more than twice as large in absolute terms as the SPTR's (-1.2670 and -1.8886). In contrast to option-writing indices, the PPUT's returns are skewed to the right, since the inherent long put option limits the downside potential of this investment. This is in line with the work of Leland (1999) pointing out that options create asymmetric return distributions that are highly skewed to the left or right, respectively. Overall, it is interesting that all indices, which outperform the S&P 500 on a risk-adjusted basis, exhibit left skewed return distributions with fatter tails than the normal. This is first evidence for the necessity to control for higher moments measuring the performance of option strategies.

[Insert Table 1 here.]

5.4.2 Linear performance models and time-varying betas

Most previous studies regarding option strategy benchmark indices measure the performance using linear performance models. The first problem arising from this approach is that options' delta and risk alter over time. Options can be replicated via stocks and bonds, whereas the replicating portfolio changes every infinitesimal small time step (Black and Scholes, 1973) and thus, beta and delta vary over time. I am the first to consider time-varying delta exposure in a linear regression model via approaches as proposed by Treynor and Mazuy (1966) as well as Henriksson and Merton (1981). I estimate the models for every index over the whole time period and compare them with a simple CAPM or Jensen's alpha approach, respectively. Table 2 shows the results.

[Insert Table 2 here.]

The majority of option indices (BXM, PUT, CMBO, BFLY, and CNDR) outperform the market up to almost three percentage points p.a., although this result is only statistically significant for the PUT and CNDR. This is in line with Whaley (2002), and Ungar and Moran (2009). Since I analyze a longer time period, the results may slightly differ from previous studies. Option strategies should neither exhibit superior nor inferior performance compared to their underlying index after risk-adjustment. The time-varying beta models draw an interesting picture. The gamma of both models is highly significant for all indices and both model specifications, i.e. beta depends remarkably on the return of the SPTR.

[Insert Figure 4 here.]

For further analyses of this phenomenon, the BXM and the PPUT serve to demonstrate the rolling character of time-varying delta and risk exposure of option strategy benchmark indices (Figure 4). Since the payoffs of these two option strategy indices are contrarian, they are especially suitable for displaying the time-varying character of their delta exposure. The BXM sells call options in every month, thus, if the underlying rises and exhibits a positive return, the amount of stocks in the replicating portfolio of the short call option gets more negative and consequently, the BXM's beta drops. The contrary effect can be observed with the PPUT, which is long put options in each month. Long put options are replicated selling stocks short and investing the proceeds in a riskless bond. All else being equal, as the underlying's return increases, the amount of shares sold short diminishes and the beta of the PPUT also increases. Based on the movements shown in Figure 4, I suppose that linear time-varying beta approaches are able to approximate the changing delta exposure character of the strategy indices.

The interpretation of the performance derived from time-varying beta models is non-trivial. Bunnenberg et al. (2017) point out that the constant of these regressions should not be understood as the overall performance. In connection with mutual fund performance, the alpha reflects the selection performance of a fund manager. The overall performance of an investment consists of the selection performance plus the timing performance. Passive option strategy indices should exhibit neither selection nor timing. However, it is possible to compute the total performance for any investment analytically and therefore; I follow Bunnenberg et al. (2017)

to determine option indices' total performance. For both timing model approaches the total performance is not clearly distinguishable from the total performance measured by Jensen's alpha. While the change in alphas is 54 basis points for the Henriksson and Merton total performance of the BFLY compared to its Jensen's alpha, all other figures show smaller alterations without losing their original sign and economic significance. Modeling time-varying delta exposure seems to be necessary but does not enhance or change the measured performance figures remarkably. A possible reason for these similar results could be that the beta of option strategies is not a linear function of the underlying's return. Israelov and Nielsen (2015) show theoretically that the functional relation proposed by Treynor and Mazuy (1966) and Henriksson and Merton (1984) is not feasible for portfolios containing options.

5.4.3 Controlling for higher moments

In Section 5.3 I show that controlling for higher moments is theoretically essential.Therefore, I employ Leland's approach for all option indices in my sample over the whole time period. Moreover, Whaley (2002) published the first study examining the BXM BuyWrite Index and proposes an approach that only considers downside risk in returns. Consequently, I estimate Equation (5) for the option strategy benchmark indices. Table 3 displays the results.

[Insert Table 3 here.]

Since Leland's alpha is not obtained from a linear regression, there is no test on significance reported in Table 3. An interesting finding is that the alphas obtained by both Leland's and Whaley's models get even more pronounced instead of changing the direction due to left skewed returns. For the BXM, Leland's alpha rises from 1.42% (Jensen's alpha) up to 2.26%, although the skewness is highly negative (-1.2670). The estimated parameters for the PPUT are in line with Leland's theory, since this strategy generates less negatively skewed returns (-

0.3179) compared to the S&P 500 and hence, the alpha increases or becomes less negative, respectively. The same result is observable for the CLL whose skewness is even more positive.

Alphas measured via Whaley's model are now throughout statistically significnant. Alpha for the BFLY changes dramatically in sign from 2.73% to -10.36%, whereas all other figures point in the same direction. This result shows the sensibility of measured performance and the dependence of the chosen model. However, I observe that controlling for higher moments in performance models does not change results significantly similar to the outcomes for time-varying beta models.

5.4.4 Option-factor models

As introduced in Section 5.3, I employ a novel option-strategy model to measure the performance of option strategies. Table 4 displays the outcome of the option-factor model regressions. The first thing to mention is that the augmented straddle option-factor seems to catch option exposure in option indices adequately (Panel A). Indices systematically selling options show negative loadings on the return of the straddle strategy and in contrast, the coefficient for the PPUT index, which is long put options, is significantly positive. Since my straddle-factor is constructed from both call and put options, I assume that this factor is able to detect the correct overall exposure in a portfolio investing (partly) in options with different long and short positions.[36]

[Insert Table 4 here.]

Since the additional factor is almost orthogonal to the SPTR[37], the market beta does not change in magnitude and thus, multicollinearity seems not to be an issue in my test setting. However, performance in terms of alpha changes dramatically in sign for all indices. For the BXM,

[36] Agarwal and Naik (2004) only use separate call and put option-factors.
[37] The linear correlation between the excess return of the SPTR and the straddle-factor is -0.02.

CMBO and BFLY the underperformance of up to almost 4.5 percentage points per year is not only statistically, but also economically significant. The change in sign deserves more attention. According to Leland (1999) left skewed returns generate positively biased alphas and vice versa. If the option-factor model adequately catches non-linearities, the alphas should be pulled into the opposite direction. Again, the BXM and the PPUT serve to illustrate this mechanism: Jensen's alpha for the BXM is positive and the performance measured with the augmented straddle-factor is negative. In contrast, the alpha of the PPUT, which exhibits returns that are skewed to the right, changes from a negative to a positive value. The problem of this model specification is that I am not able to distinguish between biases that come from asymmetric return distributions and such arising from overpriced options. This might be a reason why the alphas of some indices become statistically significant, i.e. they could be under- or overestimated.

Another main driver of the profitability of option strategies is the overall market price level of traded options. It seems plausible that in times when the price of options is generally high, selling options should be more lucrative. The straddle-factor indirectly reflects the price level due to the implied volatility that is used to compute actual traded option prices. For a deeper understanding of the results, I augment Equation (4) with the return of the VIX in excess of the risk-free rate. Every alpha is not statistically indistinguishable from zero anymore. The VIX-factor shows weak statistical significance, especially for some indices, which exhibited a statistically significant alpha in the previous regression setting.

However, one should interpret the results displayed in Panel B of Table 4 carefully. For an adequate interpretation of alpha as risk-adjusted performance as a return term, the risk-factors must be directly investable. This is true for the straddle-factor since it is a simple and repeated passive buy-and-hold strategy, which can be invested with relatively low cost. On the contrary, a direct investment into the Volatility Index (VIX) is not possible. There are indeed some investment vehicles, which allow exposure that is highly correlated with the VIX. Futures and options on the VIX are traded at CBOE and investment banks offer ETFs and ETNs on these futures.[38] All in all, there is no such investment product, which exactly reflects the development of CBOE's VIX and hence, the alpha of this performance model cannot be directly interpreted as performance denoted in percentage points.

5.4.5 Crises analysis

As shown in summary statistics, option strategy benchmark indices exhibit lower risk in terms of volatility. The indices are said to have an inherent protection against market declines. Indices that sell options have a cushion in the amount of the option premium that prevents the portfolios from potential losses. Indices that buy options, especially puts (for example the PPUT), should also be limited in losses. Schulte and Stamos (2015) find abnormally high returns in the recent financial crisis (2008-2010) in the long run. Therefore, it seems reasonable to test whether option strategies prevent the investor from experiencing losses in crises periods.

I determine five different crises scenarios: crisis scenario (i) is defined following Chalmers et al. (2013). Scenario (ii) defines all times as crisis when the implied volatility is extremely high, i.e. the VIX exceeds the 75^{th} percentile. Defining scenario (iii), I obtain business cycle data that indicates recessions in the U.S. from NBER and St. Louis Fed, respectively.[39] The first three scenarios reflect acknowledged crises definitions. Finally, I specify two different crises scenarios with respect to the S&P500 total returns. Scenario (iv) considers all points in time as crisis period when the return of the SPTR falls below the 25^{th} percentile, whereas scenario (v) takes all times into account where the SPTR's return was negative. For every crisis scenario, I run the model given in Equation (6). Table 5 shows the result of this regression model.

[38] An example for an ETN on the VIX is offered by Barclays Capital iPath®: iPath® S&P 500 VIX Short-Term Futures™ ETN (VXX) http://www.ipathetn.com/US/16/en/details.app?instrumentId=259118. This product demands a fee of 0.89% p.a., which is not included in the VIX-factor used in the performance model above.
[39] Data is freely available at St. Louis Fed's homepage: https://fred.stlouisfed.org/series/USREC

[Insert Table 5 here.]

Since Equation (6) is not a performance model regression in the classical meaning, the interpretation of the parameters is not straightforward. Alpha itself shall not be seen as performance in terms of return denoted in percentage points, as parts of the factors in the model are not investable. Hence, I focus on the analysis of the crisis dummy variable and the interaction with the SPTR's return. For scenarios (i), (ii) as well as (iii), the crisis dummy is consistently indistinguishable from zero. As this coefficient may be interpreted as additional performance effect during crisis periods, it seems that option strategy indices do not provide a shelter from losses in turbulent times in the classical meaning. However, the last two scenarios that are determined solely by the SPTR's returns exceeding a certain threshold, exhibit crisis dummy coefficients that are throughout statistically significant. For the BXM, PUT, CMBO, BFLY as well as CNDR, the estimated parameters are positive. According to this result, I conclude that these option strategies are indeed able to protect an investor from experiencing drawdowns due to crashes in the underlying S&P 500. On the other hand, one can observe significantly negative coefficients for the CLL and PPUT, which means that these indices tend to perform weaker in crises. The results attained from this analysis could be caused by the fact that indices selling options earn premia that serve as a cushion against strong market declines. On the contrary, indices that are mainly long options suffer from paying options premia and therefore underperform in crises. A possible reason for this phenomenon are unfairly priced options so that the realized payoffs are not congruent with the discounted expected payoffs (e.g. Constantinides et al., 2009).

Another important result to look at is the coefficient for the interaction term between the SPTR return and the crisis dummy. This estimate is mostly statistically significant in all scenarios and the sign pattern is exactly contrarious to the pattern of the crisis dummy. The interaction term denotes the additional loading of the market return in crisis periods, i.e. the additional systematic risk the index is exposed to in every scenario. For the first five indices, the throughout positive coefficients indicate significantly higher market exposure during turbulent times. Significantly negative correlations for the CLL and PPUT reveal a reduction in systematic risk when scenario dummy variables take on the value one. This finding is consistent with previous results in Section 5.4.2 where I show both analytically and empirically the dependency of systematic risk on the return of the underlying S&P 500.[40] Since crisis periods go along with stock price declines or negative returns, respectively, the sign and statistical significance of the interaction term coefficients is coherent.

5.4.6 Different time periods

The first study attesting an outperformance for one of the option strategy benchmark indices is Whaley (2002) who analyzes the sample period from 1988 to 2001. The outperformance of 2.76% on an annualized basis is both statistically and economically significant. Ungar and Moran (2009) have a longer time frame available (1986 to 2008) and their estimate for BXM's outperformance is 1.92% p.a. It seems that the outperformance diminishes with longer windows analyzed and thus, I naively divide my sample into two equally long sub-samples. The first sample period spans from 1990 to 2003 and the second covers the time period from 2004 to the end of 2016. I repeat the calculations for the descriptive statistics and for Jensen's alpha, respectively, for each sample separately and display the results in Table 6.

[Insert Table 6 here.]

At first, I compare the descriptive statistics for both time frames. An interesting outcome is that on the one hand, the average returns drop dramatically for the entire set of option indices in the second half of the sample. On the other hand, risk in terms of volatility does not change

[40] One should note that the time-varying beta model by Henriksson and Merton (1981) is nested in scenario (v) but in addition, scenario (v) includes the dummy variable indicating negative returns.

remarkably and thus, risk-adjusted performance extremely diminishes. For example, the BXM exhibits an average return of 11.9% from 1990 to 2003, whereas from 2004 on, the mean return only amounts to 5.7%. Since the volatility changes merely by two basis points, the Sharpe Ratio falls from 0.22 to 0.13 and is considerably smaller than that of the S&P 500. It seems that positive outperformance reported in previous studies was driven by the specific sample period.

Parameter estimates of performance regression models for the early sample period are in line with previous studies on option benchmark indices. Table 6 shows significant outperformance at the five percent level for indices that are mostly short options and peaks in BFLY's Jensen's alpha with more than 6 percentage points on an annualized basis. In panel B of Table 6, however, there is no outperformance observable, as all alphas are close to zero and no longer statistically significant at conventional levels, except for the CLL. The initial presumption of vanishing performance in the second half seems to be confirmed.

Systematic risk in terms of CAPM beta also changes in the second half of my sample, namely it increases for all indices, i.e. option strategies tend to converge to the movements of the underlying S&P 500. While the increase in beta sums up to more than 13 percentage points for the PUT (0.50 to 0.63), betas in the last three columns do not change significantly but fall slightly for these last three indices.

The question arising from the results obtained from Table 6 is why the results differ in such a manner. One possible explanation could be more efficient option pricing. In the first sample period options might be richly priced and consequently, premiums earned from selling options are considered as outperformance in excess of the S&P 500. Constantinides et al. (2009) actually find evidence for overpricing of S&P 500 index options for their analyzed time horizon from 1986 to 2006. To analyze this conjecture in more detail, I estimate rolling Jensen's alphas and betas with 60-months overlapping windows. If performance diminishes over time due to for example more efficient option pricing, the estimated rolling alphas should reveal a dependency on time elapsed. On the other hand, performance might also be driven by varying betas or specific return regimes. Therefore, I run the following OLS regression and the results are displayed in Table 7.

$$\alpha_{i,t} = \beta_{0,i} + \beta_{SPTR,i} r_{SPTR,t} + \beta_{r_f,i} r_f + \beta_{year,i} year_t + \varepsilon_{i,t} \tag{8}$$

The return of the S&P500 is negatively correlated for the first six option strategies, whereas only the BFLY reveals statistical significance. This means that, all else being equal, if the return of the underlying is negative the performance of these strategies tends to increase. The contrary effect can be observed with the protective put index, whose estimated parameter is statistically significant and positive. The different signs are not surprising, since the PPUT and the remaining strategies are contrary. Loadings on the rolling 60-months beta factor are entirely negative and statistically distinguishable from zero for four indices. Alpha is therefore also correlated with different values of systematic risk. A very interesting outcome is the coefficient for year, which is highly significant for all option strategies. Strategies possessing mainly short option exposure, which can be derived from the loadings on my option straddle-factor in Table 4, seem to lose performance over time. In contrast to this, the sign of the coefficient for the PPUT is positive, which means that the performance increases ceteris paribus over time.

[Insert Table 7 here.]

Figure 5 illustrates the development of 60-months rolling alphas over time. The performance diminishing effect is also graphically observable for most of the strategies and explains at least partly the results for the split sample analyses. The fact that alpha seems to vanish over time for option writing strategies (e.g. BXM) and underperformance for long option strategies, like the PPUT, is less pronounced, is further evidence that overpricing in options as documented by Constantinides et al. (2009) diminishes with proceeding time.

[Insert Figure 5 here.]

5.5 Analysis of investable products

Although Cici and Palacios (2015) do not find any effect of options on mutual fund portfolios, Natter et al. (2016) find superior performance among U.S. domestic equity mutual funds. The outperformance is mainly driven by short option positions taken by mutual fund managers (e.g. found with covered call writing). The question is whether this phenomenon is also observable among passively managed investment vehicles that bear short option exposure.

I analyze two different investment opportunities enabling private investors to take exposure towards the oldest option writing strategy, the BXM. The products are actively traded and easily investable for private market participants at low cost. Specifically, the BXM is represented by an Exchange Traded Fund, the PowerShares S&P 500 BuyWrite ETF (BXM ETF), and an Exchange Traded Note, namely the iPath® S&P 500 BuyWrite ETN (BXM ETN).[41] Data on these products is available as early as June 2007 for the BXM ETN and January 2008 for the BXM ETF.

The purpose of this analysis is to test the investments' eligibility to track the option strategy benchmark index and their performance compared to the underlying. Figure 6 shows the development of $ 1 invested in the respective investment vehicle or in the underlying index, respectively. Additionally, the performance of buying the S&P 500 total return index is also included via the SPDR S&P 500 ETF Trust (SPDR ETF). As the products start at different dates, the results are not directly comparable. If a private investor invests $ 1 in the BXM ETN in June 2007, she would end up with $ 1.39, whereas a hypothetical direct investment in the BXM would have yielded $ 1.46. A conservative and direct long position in the SPDR ETF

[41] Exposure against the PUT can be obtained by buying WisdomTree S&P 500 PutWriteStrategy ETF (PUT ETF). However, Data is only available from March 2016 onwards and would result in an insufficient sample size and thus, I focus only on investments regarding the BXM.

returned a much higher dollar amount of $ 1.80. The right graph of Figure 6 shows the dollar development for a time frame starting from January 2008 and an ETF on the BXM. The resulting final dollar amounts are $ 1.31 for the BXM ETF, $ 1.43 for the hypothetical BXM investment, and $ 1.86 for buying the dividend reinvesting SPDR ETF on the S&P 500. From 2007 to 2013 the S&P 500 performed worse than the option index investments since the latter have a cushion for losses in contrast to the equity index. From 2013 onwards, however, the outperformance of the S&P 500 exposure is clearly superior due to capped upside potential resulting from short option positions of the strategy benchmark indices.

[Insert Figure 6 here.]

One dimensional performance measures reveal initial drawbacks of option strategies but do not take risk considerations into account. Therefore, I estimate Jensen's alpha in three different model approaches for both the BXM ETF and the BXM ETN, respectively. In scenario (i), the market index used is the SPTR, whereas in scenario (ii) the investment's return is regressed on the replicated BXM due to the fact that it should be the adequate benchmark.[42] As fund companies demand cost for providing ETFs, in scenario (iii) the SPDR ETF serves as benchmark as it also includes costs. For all three scenarios, I use both net and gross returns. A summary of the results can be found in Table 8.

[Insert Table 8 here.]

Compared to the SPTR both the BXM ETF as well as the BXM ETN reveal an underperformance up to 1.80 percentage points p.a. However, these measures are statistically indistinguishable from zero. The BXM is by definition the adequate benchmark for these investment products and Panel B displays the outcome of the performance regressions. From

[42] Using a linear performance model in this case should be unproblematic since both the investment as well as the benchmark are supposed to generate the same non-linearities in returns.

the first three columns, one can see that investors do not earn the same return as the underlying BXM, as alphas are negative and statistically significant. The gross performance of the exchange-traded note is indeed positive but only hardly significant at the ten percent level, and, more important; an investor is not able to earn the gross performance of an ETN. Although the total expense ratio is 0.75%, the ETF's performance is slightly worse than the ETN's performance. On the one hand, this is due to the fact that the time periods slightly differ and on the other hand, the replication is incongruent, what can be derived from the different loadings on the BXM (1.0015 vs. 1.0210) as well as from the tracking error[43] (0.18% p.a. for the ETF vs. 0.17% p.a. for the ETN).

The last Panel comprises the performance measures for the comparison with the SPDR ETF. The similarity to the figures in Panel A is not surprising since the SPDR ETF has a high correlation of 0.9977 with the SPTR and is very cheaply investable with an expense ratio of 9 basis points per year. All in all, I conclude from this analysis that potential benefits of option-writing indices – as far as there are any – are consumed by the costs investment providers charge. This is not coercively conflicting with the findings of Natter et al. (2016), as the outperformance of actively managed funds engaging in options the authors measure is compared to their non-using peers.

5.6 Concluding remarks

This study aims to review the benefits of option strategy benchmark indices in terms of performance and risk. I show that the performance measurement of such indices is indeed difficult and sensitive to the method used. Several problems should be addressed when analyzing the advantages and drawbacks of portfolios containing options. One finding is the

[43] The tracking error is calculated following Cremers and Petajisto (2009) as the standard deviation of the residuals obtained from performance regressions using monthly net returns.

time-varying delta exposure towards the underlying S&P 500 index, which might lead to severe biases if neglected. However, approaches derived from timing literature insufficiently control for this phenomenon. A solution is the methodology as proposed by Israelov and Nielsen (2015), at which the evaluator has to know the exact portfolio composition and additionally the detailed options' characteristics (delta). Analytically correct approaches as proposed by Leland (1999) and Whaley (2002) are also not able to appropriately disclose performance, as resulting performance measures do not differ remarkably from standard methods. The novel approach I develop in this study shows significant loadings on my option straddle-factor though it is not clear if the constants of these regressions reflect the proper performance. Option strategies are said to be protections from sharp market declines, for example during times of crises and, for some strategies, I am able to confirm this empirically. Another finding, which is possible due to the long sample period, is that eventual outperformance is mostly driven by the period in the first half of my sample to the end of 2003.

Further, I analyze the benefits of the BXM for private investors by measuring the performance of two directly investable products reflecting the development of this option benchmark. The conclusion, which can be drawn from this examination, is that fees charged by the fund companies lead to an underperformance compared to the underlying option index.

Appendix A:

Exhibit 1:

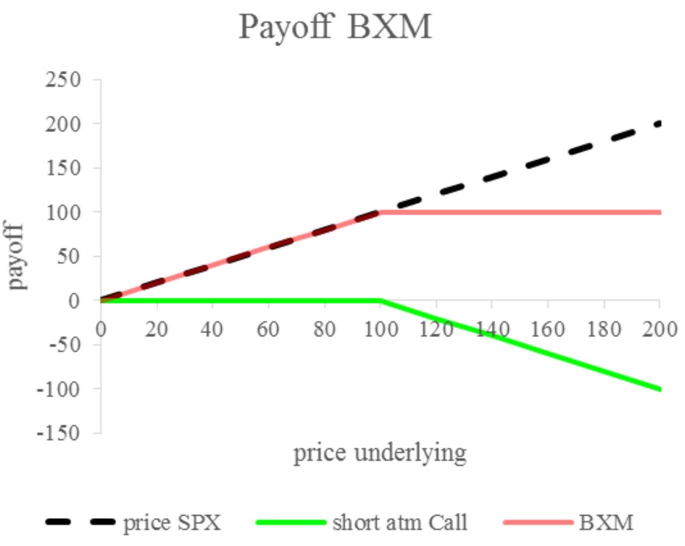


Exhibit 2:

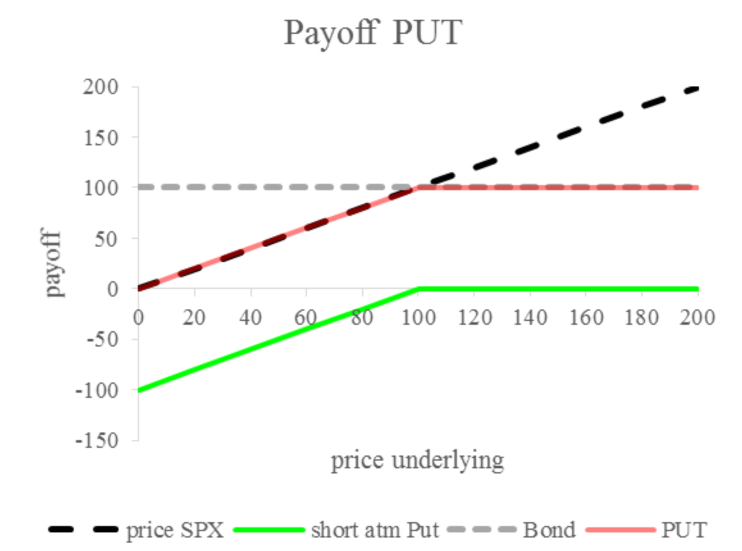


Exhibit 3:

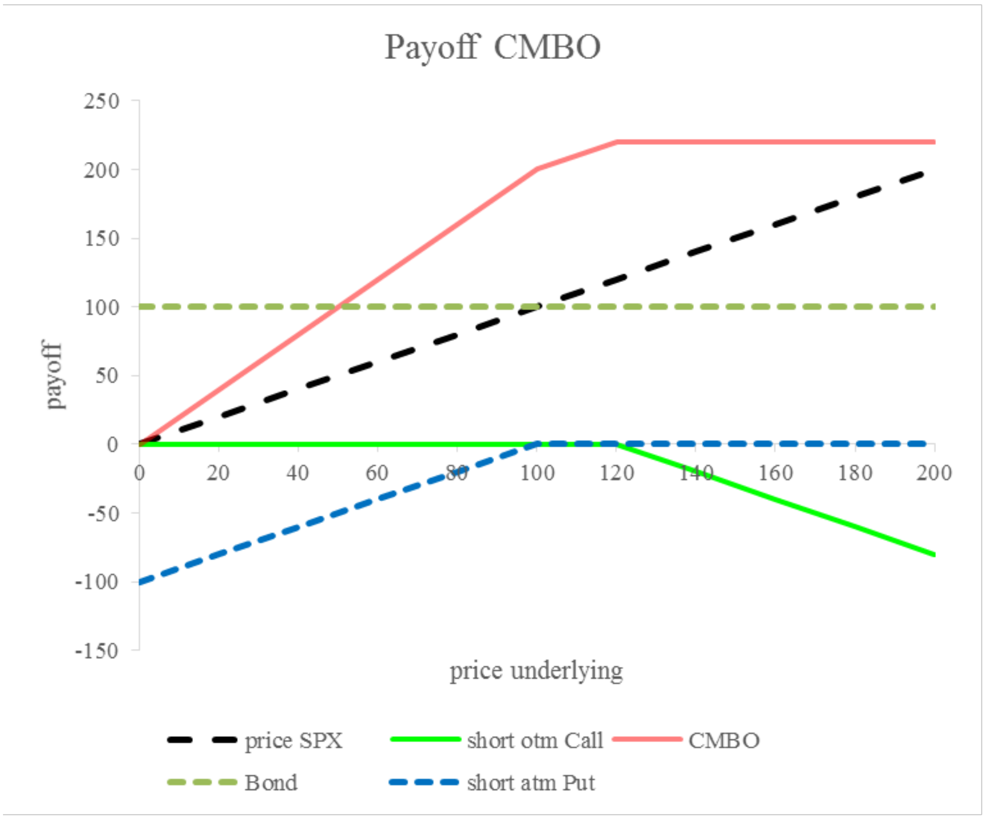


Exhibit 4:

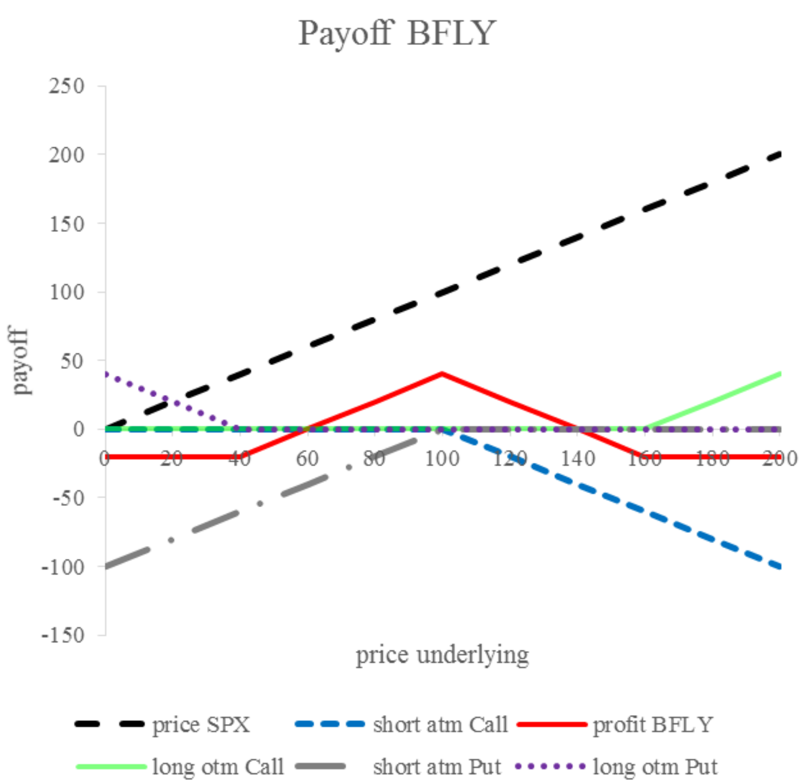


Exhibit 5:

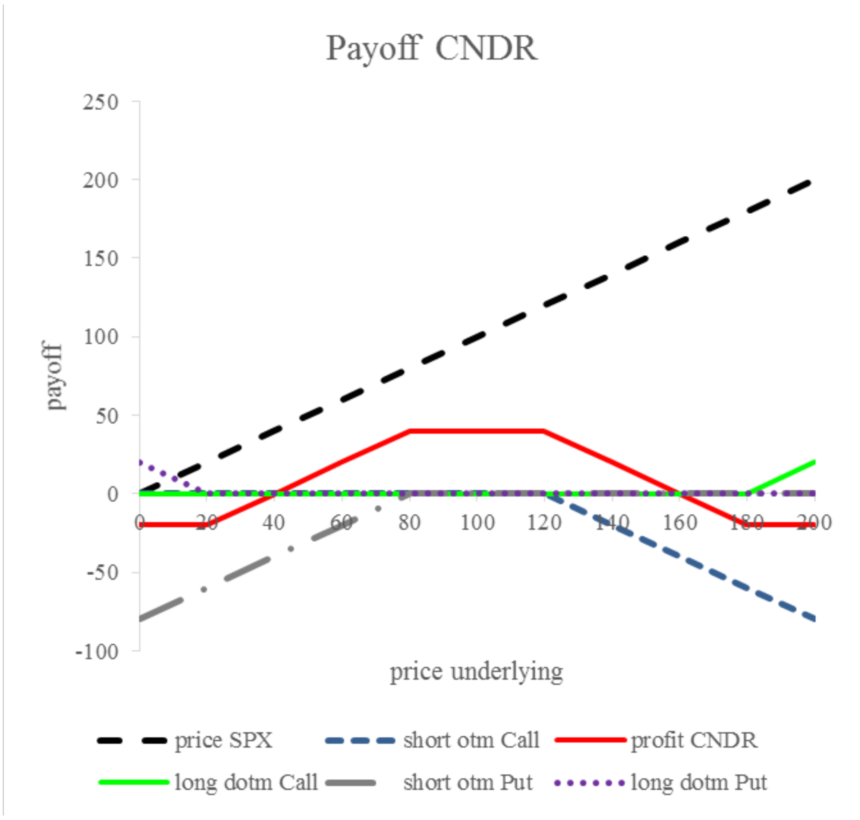


Exhibit 6:

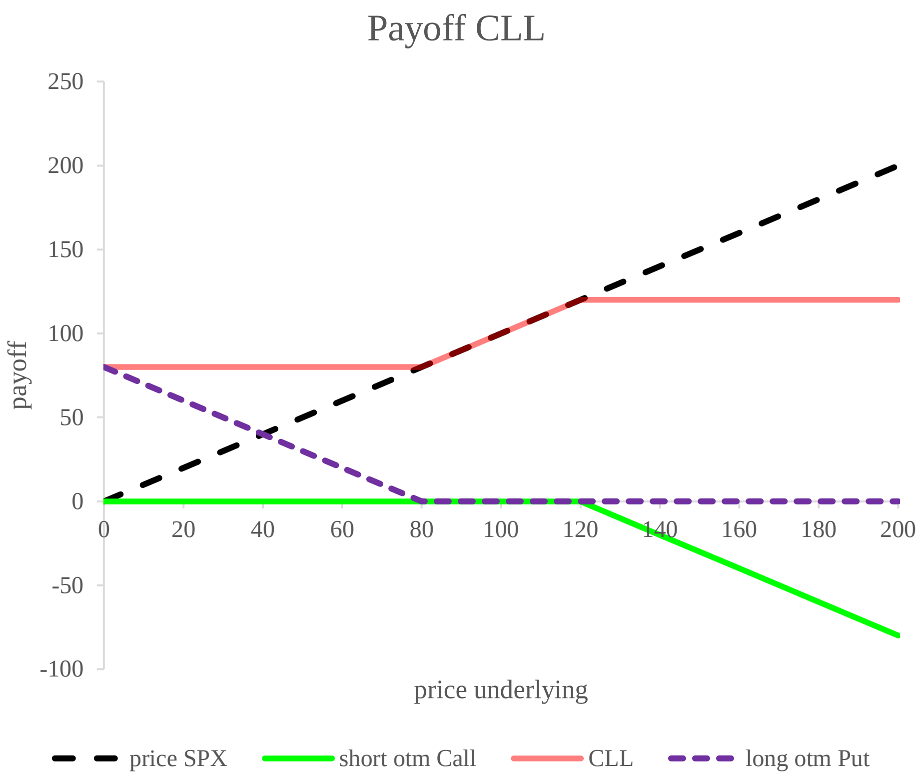


Exhibit 7:

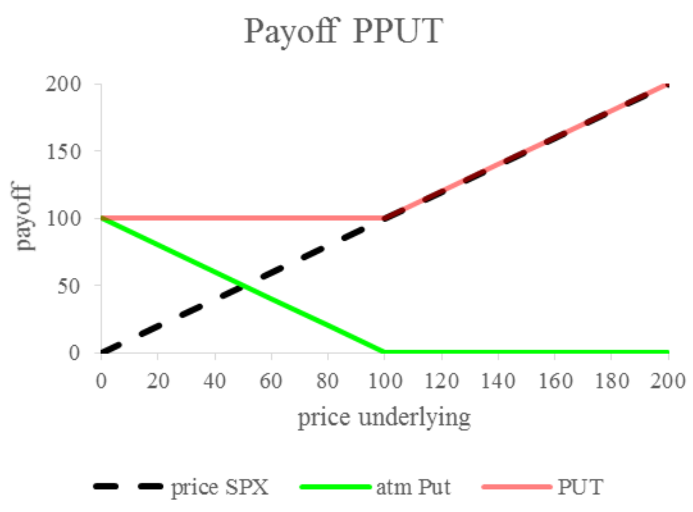
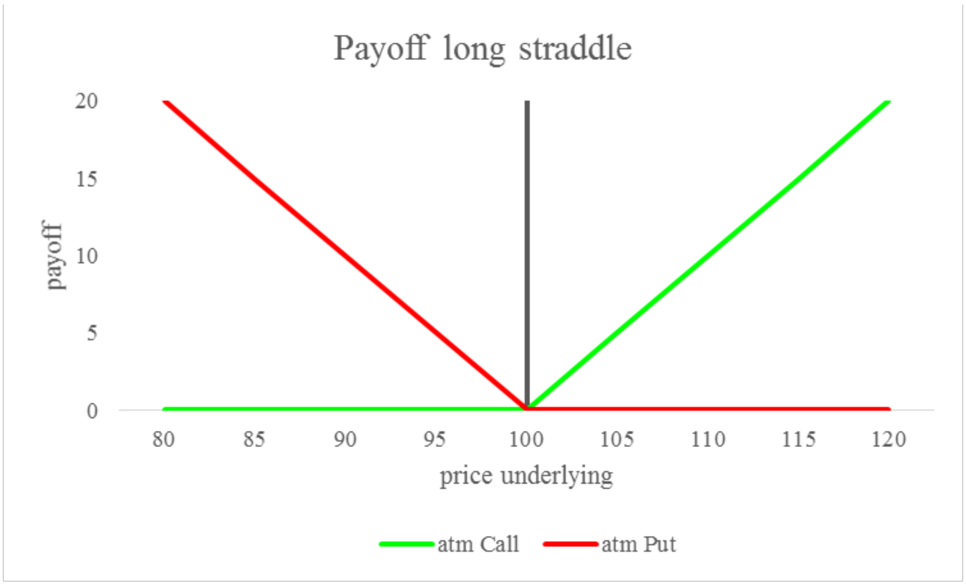


Exhibit 8: Payoff long straddle



References

Agarwal, V. and Naik, N. Y. (2004) Risks and portfolio decisions involving hedge funds. Review of Financial Studies 17 (1): 63–98.

Black, F. and Scholes, M. (1973) The pricing of options and corporate liabilities. Journal of Political Economy 81 (3): 637–654.

Bookstaber, R. and Clarke, R. G. (1981) Options can alter portfolio return distributions. Journal of Portfolio Management 7 (3): 63–70.

Bookstaber, R. and Clarke, R. G. (1984) Option portfolio strategies: Measurement and evaluation. Journal of Business 57 (4): 469–492.

Bookstaber, R. and Clarke, R. G. (1985) Problems in evaluating the performance of portfolios with options. Financial Analysts Journal 41 (1): 48–62.

Bookstaber, R. and Clarke, R. G. (1985) Problems in evaluating the performance of portfolios with options. Financial Analysts Journal 41 (1): 48–62.

Bunnenberg, S., Rohleder, M., Scholz, H. and Wilkens, M. (2017) Jensen's alpha and the market timing puzzle. Working paper, University of Augsburg, University of Hannover, University of Nuremberg. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1253923.

Chalmers, J., Kaul, A. and Phillips, B. (2013) The wisdom of crowds: Mutual fund investors' aggregate asset allocation decisions. Journal of Banking and Finance 37: 3318–3333.

Chambers, D. R., Foy, M., Liebner, J. and Lu, Q. (2014) Index option returns: Still puzzling. Review of Financial Studies 27 (6): 1915–1928.

Cici, G. and Palacios, L.-F. (2015) On the use of options by mutual funds: Do they know what they are doing? Journal of Banking and Finance 50: 157–168.

Constantinides, G. M., Jackwerth, J. C. and Perrakis, S. (2009) Mispricing of S&P 500 index options. Review of Financial Studies 22 (3): 1247–1277.

Cremers, K.J. M. and Petajisto, A. (2009) How active is your fund manager? A new measure that predicts performance. Review of Financial Studies 22 (9): 3329–3365.

Feldman, B.E. and Roy, Dhruv (2005) Passive options-based investment strategies – The case of CBOE S&P 500 buy write index. Journal of Investing 14 (2): 66–83.

Fung, W. and Hsieh, D. A. (2004) Hedge fund benchmarks: A risk-based approach. Financial Analysts Journal 60 (5): 65–80.

Goetzmann, W., Ingersoll, J., Spiegel, M. and Welch, I. (2007) Portfolio performance manipulation and manipulation-proof performance measures. Review of Financial Studies 20 (5): 1503–1546.

Henriksson, R. D. and Merton, R. C. (1981) On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. Journal of Business 54 (4): 513–533.

Israelov, R. and Klein, M. (2016) Risk and return of equity index collar strategies. Journal of Alternative Investments 19 (1): 41–54

Israelov, R. and Nielsen, L. N. (2015) Covered calls uncovered. Financial Analysts Journal 71 (6): 44–57.

Jagannathan, R. and Korajczyk, R. A. (1986) Assessing the market timing performance of managed portfolios. Journal of Business 59 (2) 1: 217–235.

Jensen, M. C. (1968) The performance of mutual funds in the period 1945-1964. Journal of Finance 23 (2): 389–416.

Kapadia, N. and Szado, E. (2007) The risk and return characteristics of the buy-write strategy on the Russell 2000 Index. Journal of Alternative Investments 9 (4): 39–56.

Leland, H. E. (1999) Beyond Mean-Variance: Performance measurement in a nonsymmetrical world. Financial Analysts Journal 55 (1): 27–36.

Lhabitant, F. S. (2000) Derivatives in portfolio management: Why beating the market is easy. Derivatives Quaterly 7 (2): 39–45.

Merton, R. C., Scholes, M. and Gladstein, M. (1978) The returns and risk of alternative call option portfolio investment strategies. Journal of Business 51 (2): 183–242.

Merton, R. C., Scholes, M. and Gladstein, M. (1982) The returns and risks of alternative put-option portfolio investment strategies. Journal of Business 55 (1): 1–55.

Natter, M., Rohleder, M., Schulte, D. and Wilkens, M. (2016) The benefits of option use by mutual funds. Journal of Financial Intermediation 26: 142–168.

Pounds, H. M. (1978) Covered call option writing: Strategies and results. Journal of Portfolio Management 4 (2): 31–42.

Schulte, D. and Stamos, M. Z. (2015) The performance of equity index option strategy returns during the financial crisis. Working paper, Tecta Invest GmbH, Allianz Deutschland.

Sharpe, W. F. (1966) Mutual fund performance. Journal of Business 39 (1): 119–138.

Sortino, F. A. Price, L. N. (1994) Performance measurement in a downside risk framework. Journal of Investing 3 (3): 59 –64.

Stutzer, M. (2000) A portfolio performance index. Financial Analysts Journal 56 (3): 52–61.

Treynor, J. and Mazuy, K. (1966) Can mutual funds outguess the market? Harvard Business Review 44 (4): 131 –136.

Ungar, J. and Moran, M. (2009) The cash-secured putwrite strategy and performance of related benchmark indexes. Journal of Alternative Investments 11 (4): 43–56.

Whaley, R. E. (2002) Return and risk of CBOE buy write monthly index. Journal of Derivatives 10 (2): 35–42.

White, H. (1980) A Heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. Econometrica 48 (4): 817–838.

Figures and Tables

Table 1

Summary Statistics Indices

	SP500 TR	BXM	PUT	CMBO	BFLY	CNDR	CLL	PPUT
Mean (p.a.)	0.1034	0.0888	0.1003	0.0936	0.0628	0.0649	0.0642	0.0680
Vola (p.a.)	0.1443	0.1021	0.0979	0.1060	0.1086	0.0697	0.1033	0.1172
Semi Vola (p.a.)	0.1605	0.1308	0.1446	0.1343	0.1014	0.1093	0.1034	0.1218
cum. Return	11.0888	8.3794	11.8883	9.5456	3.6131	4.3358	3.8553	4.1576
Skewness	-0.5928	-1.2670	-1.8886	-1.2496	0.0605	-2.0832	-0.1708	-0.3179
Kurtosis	4.2162	7.8561	12.0088	7.4806	2.6015	8.6469	2.7314	3.2375
Sharpe Ratio	0.1511	0.1724	0.2139	0.1790	0.0932	0.1542	0.1014	0.0985
Sortino Ratio	0.1358	0.1343	0.1444	0.1411	0.0992	0.0974	0.1011	0.0949
Stutzer Index	0.1490	0.1664	0.2004	0.1727	0.0934	0.1470	0.1013	0.0982
Omega Ratio	1.6475	2.1058	2.9390	2.0762	1.0974	2.8452	1.2587	1.2908
Max. Drawdown	-0.5095	-0.3581	-0.3266	-0.3813	-0.3375	-0.1366	-0.3547	-0.3892

This table shows summary statistics on the S&P 500 total return index as well as on option strategy benchmark indices. Mean, volatility and semi-volatility are denoted in absolute values on an annualized basis. The sample period spans from 1990 to 2016.

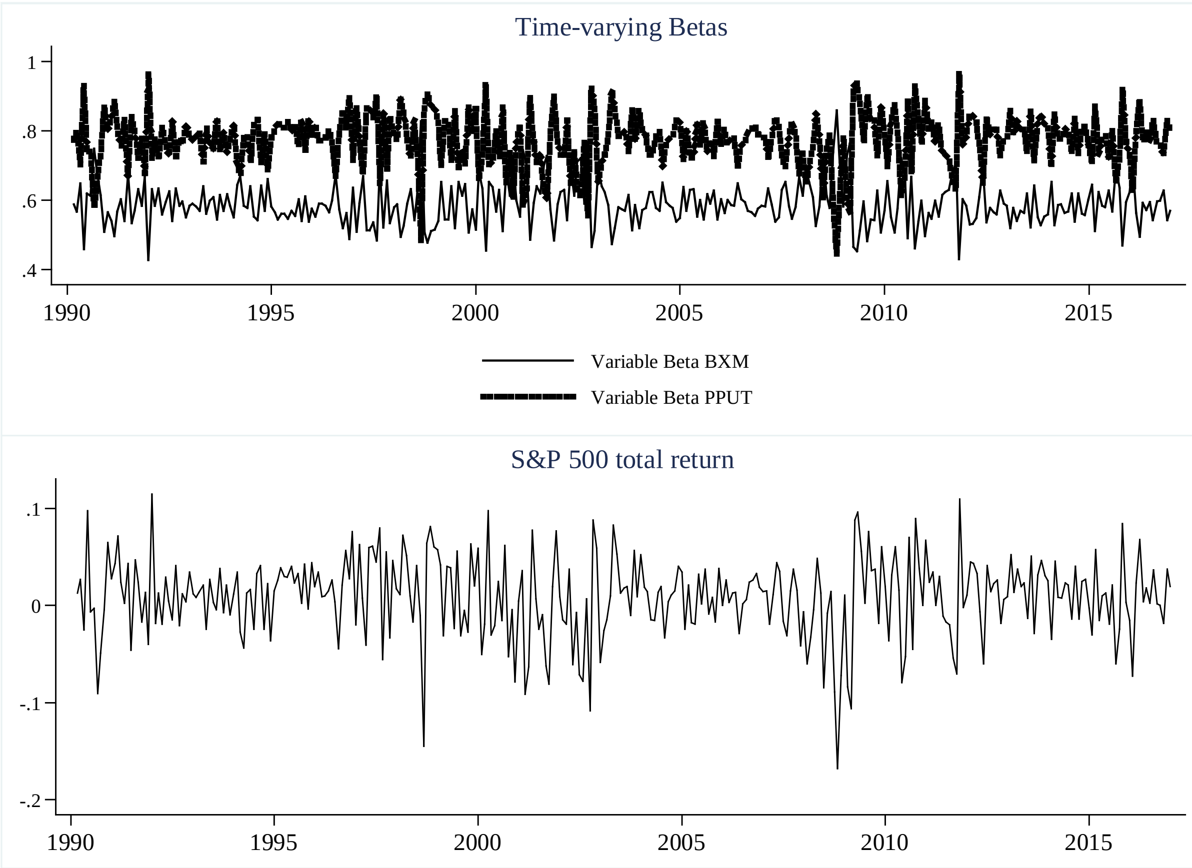
Table 2

Performance – Time-Varying Beta Models

	SP500 TR	BXM	PUT	CMBO	BFLY	CNDR	CLL	PPUT
CAPM Alpha	0.0000	0.0142	0.0301**	0.0157*	0.0273	0.0256*	-0.0131	-0.0158
CAPM Beta	1.0000	0.6180***	0.5592***	0.6616***	0.1003**	0.1502***	0.6531***	0.7389***
TM Alpha	0.0000	0.0482***	0.0709***	0.0482***	0.0687***	0.0685***	-0.0430***	-0.0577***
TM Beta	1.0000	0.5994***	0.5369***	0.6439***	0.0777	0.1268***	0.6694***	0.7617***
TM Gamma	0.0000	-1.5382***	-1.8436***	-1.4682***	-1.8735***	-1.9389***	1.3512***	1.8925**
TM Total Perf	0.0000	0.0162	0.0326	0.0176	0.0297	0.0281	-0.0149	-0.0183
HM Alpha	0.0000	0.0811***	0.1034***	0.0781***	0.1295***	0.1171***	-0.0637***	-0.0814***
HM Beta	1.0000	0.7819***	0.7388***	0.8145***	0.3507***	0.3745***	0.5293***	0.5780***
HM Gamma	0.0000	-0.3392***	-0.3717***	-0.3164***	-0.5181***	-0.4642***	0.2564***	0.3329**
HM Total Perf	0.0000	0.0177	0.0340	0.0190	0.0327	0.0304	-0.0158	-0.0193

This table shows performance measures for the approaches following Jensen (1968), Treynor and Mazuy (1966) as well as Henriksson and Merton (1981) based on monthly discrete returns. The sample period spans from February 1990 to December 2016. The market index used in all performance regressions is the excess return of the S&P 500 total return index. Performance in terms of alpha is denoted in absolute values p.a. Estimation of standard errors is heteroscedasticity consistent according to White (1980). ***, **, * denote significance of the estimated parameter at the 1%, 5%, and 10% level, respectively.

Figure 4: Time-varying betas estimated following Treynor / Mazuy (1966)



This figure shows time-varying betas for the BXM and PPUT, respectively, estimated via the Treynor and Mazuy (1966) approach. The time period spans from 1990 to 2016.

Table 3

Performance – Considering higher moments

	SP500 TR	BXM	PUT	CMBO	BFLY	CNDR	CLL	PPUT
CAPM Alpha	0.0000	0.0142	0.0301**	0.0157*	0.0273	0.0256*	-0.0131	-0.0158
CAPM Beta	1.0000	0.6180***	0.5592***	0.6616***	0.1003**	0.1502***	0.6531***	0.7389***
Leland's Alpha	0.0000	0.0226	0.0397	0.0230	0.0496	0.0466	-0.0015	-0.0056
Leland's Beta	1.0000	0.6396	0.5858	0.6819	0.1277	0.1768	0.6346	0.7116
Whaley's Alpha	0.0000	0.0156***	0.0208***	0.0154***	-0.1036***	-0.0277***	-0.0350***	-0.0393***
Whaley's Beta	1.0000	0.7027***	0.6465***	0.7348***	0.2146***	0.2692***	0.5843***	0.6520***

This table shows performance measures for the approaches following Jensen (1968), Leland (1999) as well as Whaley (2002) based on monthly discrete returns. The sample period spans from February 1990 to December 2016. The market index used in all performance calculations is the excess return of the S&P 500 total return index. Performance in terms of alpha is denoted in absolute values p.a. Estimation of standard errors is heteroscedasticity consistent according to White (1980). ***, **, * denote significance of the estimated parameter at the 1%, 5%, and 10% level, respectively.

Table 4

Performance – Straddle-factor model

	SP500 TR	BXM	PUT	CMBO	BFLY	CNDR	CLL	PPUT
CAPM Alpha	0.0000	0.0142	0.0301**	0.0157*	0.0273	0.0256*	-0.0131	-0.0158
CAPM Beta	1.0000	0.6180***	0.5592***	0.6616***	0.1003**	0.1502***	0.6531***	0.7389***
Panel A:								
Alpha (OF)	0.0000	-0.0264**	-0.0130	-0.0219*	-0.0442*	-0.0246	0.0137	0.0159
Beta (OF)	1.0000	0.6153***	0.5563***	0.6592***	0.0956**	0.1469***	0.6549***	0.7409***
Straddle (OF)	0.0000	-0.0116***	-0.0123***	-0.0107***	-0.0204***	-0.0144***	0.0077***	0.0091***
Panel B:								
Alpha (OF2)	0.0000	-0.0196	-0.0066	-0.0141	-0.0295	-0.0142	0.0135	0.0178
Beta (OF2)	1.0000	0.5878***	0.5307***	0.6278***	0.0366	0.1050**	0.6558***	0.7332***
Straddle (OF2)	0.0000	-0.0107***	-0.0114***	-0.0097***	-0.0184***	-0.0129***	0.0076***	0.0093***
VIX (OF2)	0.0000	-0.0089	-0.0083	-0.0102*	-0.0192*	-0.0136*	0.0003	-0.0025

This table shows performance measures for the approaches following Jensen (1968) and for my novel option-factor approach in the spirit of Agarwal and Naik (2004) based on monthly discrete returns. The sample period spans from February 1990 to December 2016. The market index used in all performance regressions is the excess return of the S&P 500 total return index. The straddle-factor is the return of a simple buy-and-hold strategy in excess of the risk-free rate: in month t open a long call and put position and hold it one month until expiration and repeat this procedure every month. Panel A displays the results for a CAPM regression augmented with the straddle-factor, whereas Panel B contains outcomes for the same model with the return of the VIX as an additional risk-factor. Performance in terms of alpha is denoted in absolute values p.a. Estimation of standard errors is heteroscedasticity consistent according to White (1980). ***, **, * denote significance of the estimated parameter at the 1%, 5%, and 10% level, respectively.

Table 5
Crisis analysis

		BXM	PUT	CMBO	BFLY	CNDR	CLL	PPUT
Scenario (i)	Alpha	0.0209**	0.0366***	0.0239***	0.0295	0.0281*	-0.0247***	-0.0250***
	Beta	0.5714***	0.5002***	0.6133***	0.0871	0.1126**	0.7193***	0.8208***
	Crisis	-0.0138	-0.0030	-0.0232	-0.0059	0.0100	0.0338	0.0051
	Interaction	0.1436**	0.1857**	0.1470**	0.0402	0.1211	-0.2004***	-0.2577**
Scenario (ii)	Alpha	0.0344***	0.0538***	0.0327***	0.0576**	0.0612***	-0.0373***	-0.0388***
	Beta	0.4761***	0.4041***	0.5412***	-0.0868	-0.0129	0.7757***	0.8898***
	Crisis	0.0003	-0.0062	0.0008	-0.0151	-0.0527	0.0287	0.0058
	Interaction	0.2424***	0.2624***	0.2060***	0.3136***	0.2574***	-0.1977***	-0.2554***
Scenario (iii)	Alpha	0.0228**	0.0402***	0.0232***	0.0369	0.0310**	-0.0253***	-0.0261***
	Beta	0.5750***	0.5040***	0.6245***	0.0522	0.1114***	0.7127***	0.8048***
	Crisis	-0.0310	-0.0289	-0.0277	-0.0335	-0.0039	0.0445	0.0180
	Interaction	0.1560**	0.2041*	0.1346*	0.1750*	0.1490*	-0.2158**	-0.2496*
Scenario (iv)	Alpha	0.0507***	0.0636***	0.0508***	0.0894***	0.0829***	-0.0277***	-0.0415***
	Beta	0.4938***	0.4339***	0.5451***	-0.0994	-0.0312	0.7290***	0.8461***
	Crisis	0.1304***	0.1838***	0.1086***	0.1565*	0.1282***	-0.1995***	-0.2042***
	Interaction	0.4159***	0.4874***	0.3724***	0.5997***	0.5267***	-0.4096***	-0.4782***
Scenario (v)	Alpha	0.0419***	0.0547***	0.0492***	0.0792**	0.0810***	-0.0101	-0.0245***
	Beta	0.5090***	0.4497***	0.5470***	-0.0825	-0.0285	0.6950***	0.8145***
	Crisis	0.1109***	0.1384***	0.0816***	0.1425**	0.1021***	-0.1527***	-0.1620***
	Interaction	0.3697***	0.4102***	0.3388***	0.5579***	0.4922***	-0.2998***	-0.3783***

This table shows performance measures for different crisis scenarios following Jensen (1968). The sample period spans from February 1990 to December 2016. The market index used in all performance regressions is the excess return of the S&P 500 total return index. Scenario (i) is defined as in Chalmers et al. (2013), scenario (ii) defines times when the implied volatility measured by the VIX exceeds the 75^{th} percentile, scenario (iii) reflects business cycles obtained from NBER and St. Louis Fed. The crisis dummy in scenario (iv) is one if the SPTR's return is below the 25^{th} percentile and in scenario (v), the binary crises variable takes on the value one if the SPTR's return is negative. Performance in terms of alpha is denoted in absolute values p.a. Estimation of standard errors is heteroscedasticity consistent according to White (1980). ***, **, * denote significance of the estimated parameter at the 1%, 5%, and 10% level, respectively.

Table 6
Summary Statistics Indices Split Sample
Panel A: 1990 – 2003

	SP500 TR	BXM	PUT	CMBO	BFLY	CNDR	CLL	PPUT
Mean (p.a.)	0.1211	0.1188	0.1248	0.1220	0.1094	0.0945	0.0914	0.0817
Vola (p.a.)	0.1495	0.1018	0.0929	0.1046	0.1037	0.0697	0.1099	0.1254
Semi Vola (p.a.)	0.1607	0.1289	0.1410	0.1303	0.0930	0.1152	0.1065	0.1283
cum. Return	3.5854	3.8232	4.2991	4.0199	3.2309	2.5796	2.2694	1.7884
Skewness	-0.4768	-1.2173	-1.8530	-1.1819	-0.0453	-2.1371	-0.0876	-0.2187
Kurtosis	3.5349	6.6433	9.9957	6.6338	2.2082	8.8091	2.5797	2.9031
Sharpe Ratio	0.1511	0.2157	0.2553	0.2187	0.1853	0.2141	0.1274	0.0894
Sortino Ratio	0.1404	0.1700	0.1676	0.1752	0.2063	0.1291	0.1314	0.0873
Stutzer Index	0.1498	0.2071	0.2377	0.2101	0.1858	0.2005	0.1276	0.0894
Omega Ratio	1.5692	2.2115	3.2821	2.1509	1.1974	3.3947	1.2877	1.2568
Max. Drawdown	-0.4473	-0.3019	-0.2900	-0.3210	-0.1193	-0.1332	-0.2070	-0.3453
CAPM Alpha	0.0000	0.0296**	0.0430**	0.0301**	0.0615**	0.0394**	-0.0034	-0.0225
CAPM Beta	1.0000	0.5928***	0.4979***	0.6267***	0.0642	0.1545***	0.6640***	0.7846***
Alpha (OF)	0.0000	-0.0073	0.0029	-0.0036	0.0014	-0.0070	0.0246*	0.0041
Beta (OF)	1.0000	0.5987***	0.5043***	0.6320***	0.0737	0.1619***	0.6595***	0.7803***
Straddle (OF)	0.0000	-0.0121***	-0.0132***	-0.0111***	-0.0198***	-0.0153***	0.0092***	0.0088***
TM Alpha	0.0000	0.0728***	0.0912***	0.0715***	0.1286***	0.0961***	-0.0383***	-0.0542***
TM Beta	1.0000	0.5789***	0.4824***	0.6134***	0.0426	0.1363***	0.6752***	0.7947***
TM Gamma	0.0000	-1.8601***	-2.0732***	-1.7796***	-2.8876***	-2.4377***	1.4997***	1.3628***
HM Alpha	0.0000	0.1054***	0.1274***	0.1017***	0.1852***	0.1501***	-0.0693***	-0.0763***
HM Beta	1.0000	0.7742***	0.6998***	0.7979***	0.3600***	0.4193***	0.5066***	0.6560***
HM Gamma	0.0000	-0.3652***	-0.4065***	-0.3447***	-0.5956***	-0.5330***	0.3170***	0.2589**

Table 6 continued

Panel B: 2004 – 2016

	SP500 TR	BXM	PUT	CMBO	BFLY	CNDR	CLL	PPUT
Mean (p.a.)	0.0845	0.0566	0.0740	0.0631	0.0129	0.0332	0.0350	0.0532
Vola (p.a.)	0.1388	0.1020	0.1028	0.1072	0.1122	0.0689	0.0954	0.1081
Semi Vola (p.a.)	0.1602	0.1326	0.1478	0.1382	0.1089	0.1042	0.1000	0.1142
cum. Return	1.6364	0.9446	1.4322	1.1007	0.0903	0.4906	0.4851	0.8497
Skewness	-0.7466	-1.3252	-1.8903	-1.3142	0.1890	-2.0546	-0.3419	-0.4719
Kurtosis	5.1307	9.1387	13.1610	8.2667	2.9512	8.6021	2.8344	3.6895
Sharpe Ratio	0.1507	0.1261	0.1742	0.1375	0.0025	0.0894	0.0697	0.1101
Sortino Ratio	0.1309	0.0972	0.1213	0.1069	0.0026	0.0589	0.0666	0.1045
Stutzer Index	0.1483	0.1229	0.1653	0.1338	0.0025	0.0871	0.0697	0.1095
Omega Ratio	1.7368	2.0000	2.6279	2.0000	1.0000	2.3913	1.2286	1.3284
Max. Drawdown	-0.5095	-0.3581	-0.3266	-0.3813	-0.3375	-0.1366	-0.3547	-0.3892
CAPM Alpha	0.0000	-0.0025	0.0160	0.0000	-0.0095	0.0108	-0.0234 *	-0.0082
CAPM Beta	1.0000	0.6487***	0.6346***	0.7045***	0.1441**	0.1445**	0.6394***	0.6825***
Alpha (OF)	0.0000	-0.0477***	-0.0270	-0.0417***	-0.1002***	-0.0456**	-0.0033	0.0281
Beta (OF)	1.0000	0.6359***	0.6224***	0.6927***	0.1185*	0.1286***	0.6451***	0.6928***
Straddle (OF)	0.0000	-0.0114***	-0.0108***	-0.0105***	-0.0228***	-0.0141***	0.0050***	0.0091***
TM Alpha	0.0000	0.0236	0.0483***	0.0242*	0.0108	0.0433**	-0.0482***	-0.0567***
TM Beta	1.0000	0.6266***	0.6072***	0.6840***	0.1269	0.1169**	0.6605***	0.7236***
TM Gamma	0.0000	-1.2417**	-1.5424**	-1.1544***	-0.9667	-1.5533***	1.1832	2.3117**
HM Alpha	0.0000	0.0575***	0.0770***	0.0542***	0.0768	0.0879***	-0.0571**	-0.0836***
HM Beta	1.0000	0.7988***	0.7873***	0.8402***	0.3601***	0.3376***	0.5550***	0.4937***
HM Gamma	0.0000	-0.3219***	-0.3276*	-0.2911***	-0.4634*	-0.4143***	0.1812	0.4050*

This table shows summary statistics as well as performance measures following Jensen (1968), my novel option-factor approach in the spirit of Agarwal and Naik (2004) and the time-varying beta approaches by Treynor and Mazuy (1966) as well as Henriksson and Merton (1981) based on monthly discrete returns. The first time window (Panel A) begins in February 1990 and ends in December 2003. Panel B displays the results from 2004 to 2016. The market index used in all performance regressions is the excess return of the S&P 500 total return index. Performance in terms of alpha is denoted in absolute values p.a. Estimation of standard errors is heteroscedasticity consistent according to White (1980). ***, **, * denote significance of the estimated parameter at the 1%, 5%, and 10% level, respectively.

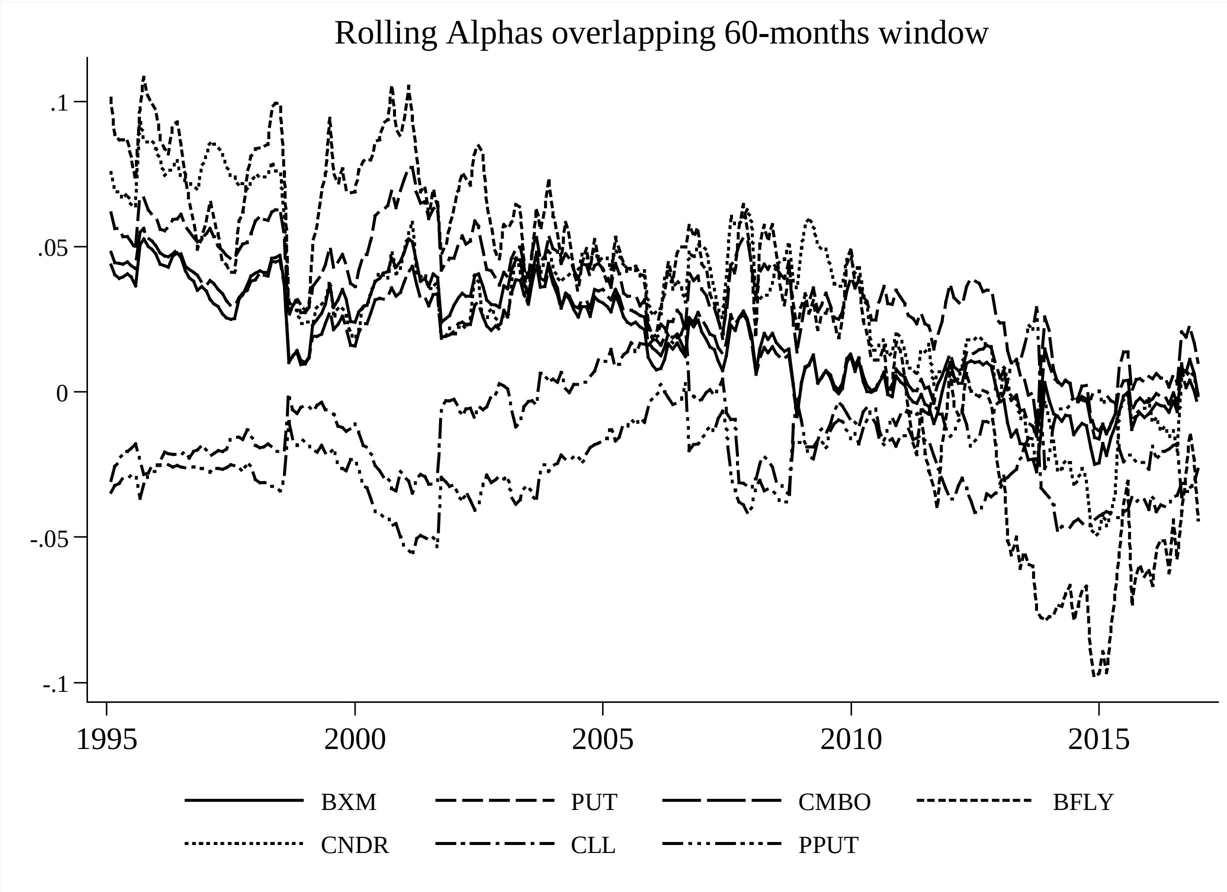
Table 7

Rolling Alphas 60 months

	BXM	PUT	CLL	BFLY	CMBO	CNDR	PPUT
S&P500	-0.0241	-0.0307	-0.0028	-0.0914**	-0.0008	-0.0142	0.0446**
Beta	-0.0111	-0.0090	-0.1403***	-0.0140	-0.0215***	-0.0331***	-0.0535***
year	-0.0026***	-0.0024***	-0.0018***	-0.0069***	-0.0021***	-0.0036***	0.0007***
R^2	0.76	0.67	0.56	0.77	0.75	0.65	0.43
N	264	264	264	264	264	264	264

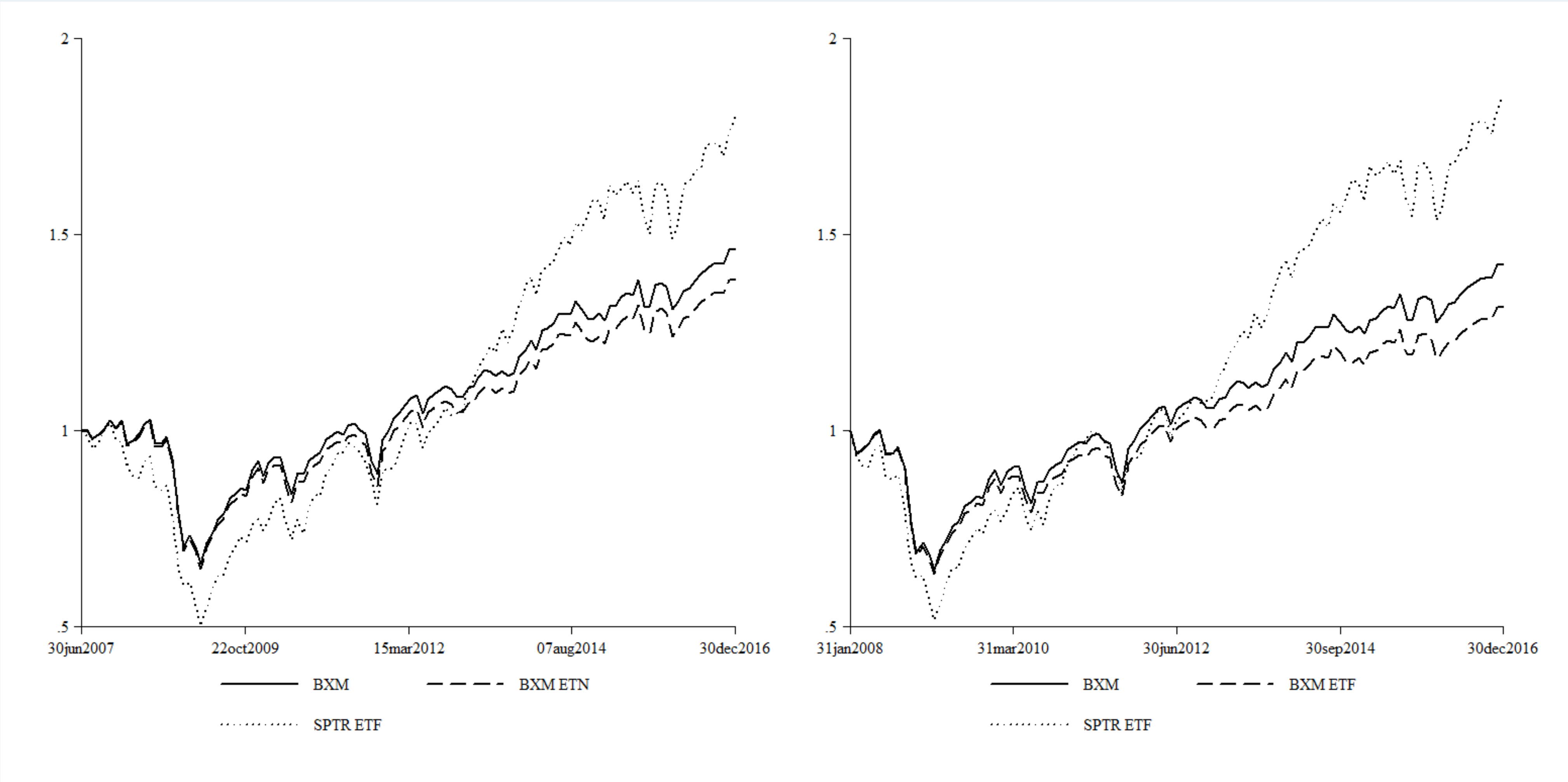
This table shows results for regressions of rolling alphas for all option indices on the return of the SPTR, the rolling beta as well as a year variable. The sample period spans from January 1995 to December 2016. Rolling performance and risk measures are estimated using the Jensen's (1966) approach. The dependent variable alpha is denoted on an annualized basis. Estimation of standard errors is heteroscedasticity consistent according to White (1980). ***, **, * denote significance of the estimated parameter at the 1%, 5%, and 10% level, respectively.

Figure 5**:** Development of rolling 60-months alphas over time



This figure shows the development of 60-months overlapping and rolling alphas for the time period from January 1995 to December 2016.

Figure 6: Development of $1 for different investments



This figure shows the development for $1 invested in the ipath® S&P 500 BuyWrite ETN (BXM ETN), the PowerShares S&P 500 BuyWrite ETF (BXM ETF), the BXM and the SPDR S&P 500 ETF Trust (SPDR ETF). The sample period for the BXM ETN spans from June 2007 to the end of December 2016 and the dataset for the BXM ETF starts in January 2008 ending in 2016.

Table 8

Investments' performance

	BXM ETF		BXM ETN	
	net	gross	net	gross
Panel A: S&P 500 TR				
Alpha	-0.0178	-0.0103	-0.0104	-0.0029
Beta	0.6694***	0.6698***	0.6784***	0.6788***
R^2	0.7987	0.7987	0.7976	0.7976
Panel B: BXM				
Alpha	-0.0091***	-0.0016**	-0.0063***	0.0012*
Beta	1.0015***	1.0021***	1.0210***	1.0216***
R^2	0.9998	0.9998	0.9998	0.9998
Panel C: SPDR ETF				
Alpha	-0.0175	-0.0100	-0.0100	-0.0025
Beta	0.6748***	0.6752***	0.6837***	0.6841***
R^2	0.7960	0.7960	0.7948	0.7948
Observations	108	108	115	115

This table shows performance measures following Jensen (1968) for the PowerShares S&P 500 BuyWrite ETF (BXM ETF) as well as the ipath® S&P 500 BuyWrite ETN (BXM ETN). The sample period for the BXM ETF spans from January 2008 to the end of 2016 and the dataset for the BXM ETN starts in June 2007 ending in December 2016. In Panel A: the market index used in performance regressions is the excess return of the S&P 500 total return index, in Panel B, the BXM serves as benchmark and in Panel C, the return of the SPDR S&P 500 ETF Trust (SPDR ETF) in excess of the risk-free rate is the market proxy. Performance in terms of alpha is denoted in absolute values p.a. Estimation of standard errors is heteroscedasticity consistent according to White (1980). ***, **, * denote significance of the estimated parameter at the 1%, 5%, and 10% level, respectively.

6 Conclusion

This dissertation addresses important issues in mutual fund research related to different sources of non-linearities in fund returns. Specifically, such sources are inherent in instruments with non-linear payoff structures like derivatives and bonds. So far, only a few studies analyze how derivatives and other investment practices relate to mutual fund performance and risk. The first article therefore examines the benefits of option use on equity fund performance and risk. The main finding is that option users outperform nonusers at lower risk. The second article similarly examines the impact of complex investments on bond fund performance and risk; however, it comes to a different conclusion. Specifically, bond funds may employ interest rate futures to speculate on interest rate changes but they are not successful.

Another source of non-linearities are bond returns itself. The third paper uncovers a systematic duration bias that occurs in previous performance measurement of bond mutual funds because state-of-the-art models ignore this non-linearity. Finally, combining questions regarding complex investments and non-linearity adjusted performance measurement, the fourth paper shows the complexity of portfolios containing options and revisits appropriate performance models. The main finding is that performance of option containing portfolios is highly sensitive to model choice and the analyzed time period.

The findings in this thesis are novel and highly relevant for different groups of market participants and regulators. First, politicians and regulators, respectively, can use these insights for future legislation. The findings of this dissertation show the necessity of passing individual regulations for funds with different investment objectives rather than imposing them with standardized regulations in general. The reason for this claim is that complex investments and leverage generating instruments do not occur equally among different investment styles. They are much more common among bond funds than among equity funds. Moreover, most of the leverage generating instruments are not harmful for the fund universe but beneficial as it comes to options among equity funds. A standardized regulation might harm not only equity but also bond funds, as they are reliant on these investments due to the highly competitive market and a discriminatory yield environment. Thus, individual regulation laws will appropriately consider the differing characteristics of diverse investment objective mutual funds and the differing effects of complex investments on their performance and risk.

Second, the results of this thesis are relevant for investment decisions of both retail and institutional investors. Options in equity mutual funds can be beneficial for investors as option users generate an outperformance at lower systematic risk. Contrary to regulators' concerns these funds do not employ these derivatives to speculate but rather to generate income that is directly passed to investors via higher risk-adjusted performance. When it comes to bond funds, however, investors should be aware that some complex investments are used to speculate and that these practices may diminish risk-adjusted performance and increase interest rate risk. Therefore, the outcomes facilitate investors' decisions to invest in both equity as well as bond mutual funds.

The duration bias illustrated in the third article is also relevant for regulators and investors at the same time. The evaluation of bond funds' performance is severely affected by the choice of the benchmark. The approach proposed in the third paper helps assessing the performance of bond funds accurately. Hence, investors are saved from systematically overestimating performance and are consequently able to make right investment decisions.

Lastly, investors that engage in option strategies can benefit from the findings of the fourth article. Assessing the performance of portfolios with option-like components is highly complex. Outperformance found by previous research is dependent on many factors such as time horizon and model choice. It can be shown that costs charged by investment providers consume eventual benefits generated by these strategies.

Beyond the results of this thesis, there are still research gaps to be closed. The insights presented in this dissertation enable other researchers to expand knowledge on complex investments among mutual funds. A relatively young field in mutual fund research is the examination of funds' portfolio holdings, as started by Daniel et al. (1997). Further research might be able to detect long and short positions or the detailed type of complex investment. For example, researchers can deal with performance and risk of the derivative portfolio and the share of actuals in the same portfolio separately. Morningstar Direct's database allows conclusions about the actual exposure arising from derivative securities that can help to judge hazards derivative securities may pose. On December 11 2015, the SEC released a proposal[44] for further limiting mutual funds' derivatives use and at the same time, Deli et al. (2015) published a white paper on this matter.[45]

Furthermore, researchers can build on the findings of the last two papers. Results obtained from previous studies must be revisited with respect to biased performance measures. The approach to overcome the duration bias presented here can be translated to almost every question about bond fund performance to ascertain that this bias is not the main driver of the results.

[44] https://www.sec.gov/rules/proposed/2015/ic-31933.pdf
[45] https://www.sec.gov/dera/staff-papers/white-papers/derivatives12-2015.pdf

Bibliography

Daniel, K., Grinblatt, M., Titman, S. and Wermers, R. (1997) Measuring mutual fund performance with characteristic-based benchmarks. Journal of Finance 52 (3): 1035–1058.

Deli, D, Hanouna, P., Stahel, C. W., Tang, Y. and Yost, W. (2015) Use of derivatives by registered investment companies. SEC White Paper, https://www.sec.gov/dera/staff-papers/white-papers/derivatives12-2015.pdf Accessed 29 March 2017.