An empirical evaluation of fixed income fund performance: New evidence across alternative methods

Elizabeth Grace Littlejohn

Supervisors:

Professor Jonathan Fletcher

Professor Andrew Marshall

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Accounting and Finance

University of Strathclyde

Business School

United Kingdom

December 2020

Contents

List of Tables

List of Figures

Copyright statement

'The copyright if this thesis belongs to the author under the terms of the United Kingdom Copyrights Acts as qualified by the University of Strathclyde Regulation 3.49. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.'

Acknowledgements

Thanks to my supervisors Professor Jonathan Fletcher and Professor Andrew Marshall for their support throughout my PhD and helping to secure funding that allowed me to pursue it. This also extends to others within the department for their help over the years; Dr Patrick McColgan, Professor Krishna Paudyal, and Martin Kemmitt. I am very grateful to my examiners, Professor Qi Zhang and Dr Devraj Basu, for their helpful comments and insights.

I appreciate the feedback and experienced gained from various events; including the Scot Doc conferences, FEBs Malaga, BAFA Manchester and Glasgow. The Mitchell library deserves a special note of gratitude as my favourite place to work. It has been missed recently and I really hope that it survives the pandemic.

Shout out to Jamie, Ann-Marie, Matt, Nathan, Lisa, Louisa, Nicky, Emma, Aaron - very grateful to have such a hilarious group of friends, timeless antics! Glad I can now join in again. Also thanks to Charlene, Joanne, Habiba, Julie, Ayth, and James. Their constant queries of "are you finished yet?!" were always a source of motivation! Thankful to Dr Alistair Byrne and Dr Apostolos Katsaris for their encouragement to undertake my PhD and guidance along the way.

As always, I am grateful to my mum and brother for their constant inspiration and support. A particularly special note of thanks to my late Gran - many of my adventures would not have been possible without her and I am forever grateful for everything she did to help me with all my achievements.

Elizabeth Littlejohn, FRM

December 2020

Abstract

With a wealth of research directed at fund performance evaluation, that which is specific to bond funds is relatively minimal. This thesis aims to shed further light on this topic using a new sample of UK data. To most effectively judge the merits of active management the specification of a reliable benchmark model is paramount. The first empirical chapter involves testing a range of single and multi-factor models to determine which of these can be considered the most suitable for evaluation of the funds in question. The Gibbons, Ross, and Shanken (1989) test of mean-variance efficiency is first applied as an absolute test, with further alpha test statistics used for a relative evaluation of the candidate models. Overall, the results show that a five-factor model (named Maturity 5) that includes adjustment for both term and credit is the most reliable. This is primarily indicated by the lowest absolute alphas and corresponding goodness of fit statistics, i.e., low standard errors and high (adj)R^2*. A multi-period analysis is conducted to determine the extent to which this varies over time. The results are consistent, with the same model performing the best in each case.*

The second empirical chapter builds upon these findings and uses Maturity 5 to evaluate the sample of UK bond funds. The results over the whole period from January 1999 until July 2016 are consistent with much of the academic literature; the funds are found to underperform on a risk-adjusted basis by approximately a magnitude of costs. However, during the recent subsample from September 2009 the performance is neutral. The use of dummy variables and Wald tests indicates that the Government, Corporate, and Diversified funds perform significantly better here. Having identified during testing

that minimal bias is likely to be induced by the Maturity 5 model, it can be inferred that the performance is not just a result of exposure to passive portfolios (as represented by the test assets). Instead, there appears to be some ability beyond this being employed by the active funds. The measure of Treynor and Mazuy (1966) has been used to identify if market timing makes a positive contribution to performance. Evidence is minimal when the Barclays Sterling Aggregate Index is used as a proxy, however, 25% of Corporate bond funds exhibit positive ability relative to the category-specific benchmark as assigned by Morningstar.

The third and final empirical chapter seeks to shed some light as to whether the active managers are lucky with respect to alpha generation or indeed exhibit true outperformance. A bootstrap procedure is first applied to the individual funds to do so. The method used here is known as entire-cases resampling (Fama and French 2010), whereby the time-ordering is maintained across the sample. This differs from the approach of KTWW (2006), which is used prevalently throughout the fund performance literature. To date, the entire-cases method has not yet been applied to a sample of bond funds. The initial results in this chapter support superior performance in the low-rate environment; this being evident across all funds from the 97h percentile, and Corporate from the 95th. To add further robustness, two variations of the false discovery rate method have been used to adjust for luck. The "classical" approach of BSW (2010) also finds positive alpha, isolated to the post-crisis period. Lower expenses characterise these funds, along with a higher number of observations per fund, and the average alpha is approx. 1.68% p.a. Recent literature has proposed many refinements to the methods used to address multiple hypothesis testing issues. The Ferson and Chen (2019) approach expands upon the entire-cases method already used in this chapter, allowing for not only a lucky distribution to be considered, but also simulates those defined as both "good" and "bad", incorporating power and confusion parameters. The results here are the most positive yet. At 10% 10% level, the proportion of outperforming funds is 11% and 33% across the whole and recent periods respectively. This is again driven primarily by the Corporate category of funds.

Declaration

I certify that the thesis I have presented for examination for the PhD in Accounting and Finance degree of the University of Strathclyde is solely my own work other than where I have clearly indicated that it is the work of others.

I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

Chapter 1

Introduction

This chapter introduces the thesis; outlining the motivations, research questions to be addressed in the three empirical chapters, and the key contributions of the study. A summary of the conclusions are presented and areas of interest for further research are highlighted.

1.1 Introduction

The performance of investment vehicles has long been a research area of interest, for both academics and practitioners alike. Given the evolving range of products and investment environment, it ensures that there continues to be a rich field of interesting topics that warrant further study. The UK has seen the total value of assets under management increase to approximately £9.1 trillion as at the end of 2017, representing around 35% of all assets under management across Europe. Outside of the US, it is evident that the UK maintains a dominant presence as a key global investment centre.

Since the recent financial crisis, opportunities and challenges have been created as a result of various factors; from regulators seeking further transparency regarding value creation, challenging investment conditions due to macro intervention, and increased competition as the number and availability of products grows. Fixed income funds have been the subject of a number of investigations and studies, predominantly in relation to their trading activity and potential threats to financial stability. For example, The International Monetary Fund (IMF) highlighted concern with the extent of derivative use by fixed income funds, as much of which was going unreported . A high proportion of their trading is conducted over-the-counter (OTC), and thus obtaining transparency and estimations of leverage employed is often difficult. Furthermore, an interesting paper by Feroli et al. 2014, "Market Tantrums and Monetary Policy", stresses the importance of considering the performance of actively managed fixed income funds in a wider context. With interest rates at historic lows at the time of writing, the paper speculates as to the repercussions in event of a rate rise. Given that rising rates imply falling bonds prices and potentially undesirable impact on fund performance, the concern regards the sell-offs that may ensue. As liquidity conditions in bond markets having been highlighted as fragile in recent years, it is proposed that central banks take careful consideration of the implications of their actions on such funds and the potential for systemic risk (Goldstein et al. 2017).

With respect to the challenges presented by increased competition, exchange-traded funds (ETFs) and index-tracking mutual funds present low-cost alternatives to their more traditional active counterparts. For example, a comparison of the Total Expense Ratios (TERs) of a fixed income mutual fund relative to that of an ETF highlights the justification for such fee pressure; with TERs in the range of 0.85% to as low as 0.1% for each of the vehicles respectively. In terms of applicability to fixed-income products, the provision of ETFs was slower on the uptake than was the case with equities. One reason for this relates to difficulties with regards to fixed-income index replication. Nonetheless, demand has been increasing resulting in additional competition for active bond fund managers. Further to this, studies have highlighted the implications of this new product development in relation to market dynamics, such as price discovery and again liquidity concerns (Dannhauser 2017). Given that ETFs are very liquid with their shares easily accessible and traded on exchanges, it has been thought that there is a disconnect between this and the liquidity of the underlying securities. Thus, despite their ability to offer cost-effective alternatives to a wide range of investors, the potential for market disruption is considered notable.

In the academic literature, the merit of active management has been an extensively addressed topic. Studies seek to identify if such funds appear to have the ability to return value to investors and what skills the managers are employing to do so. If the assertions of The Efficient Markets Hypothesis (Fama 1970) hold, then fund managers cannot consistently beat the market. In this context, it has been proposed that active management is a value decreasing activity and investors would be best advised to allocate capital to low-cost, passive investment vehicles. Central to this, however, is the designation of an appropriate benchmark model with which to judge such merits of active management on a risk-adjusted basis (Aragon and Ferson 2006). There is a wealth of literature regarding the specification and testing of equity factor models for this purpose (C. R. Harvey and Liu 2019). However, that which relates to bonds is minimal in this respect. Although commonalities exist between the risk factors to which stocks and bonds are exposed, bond-specific factor models are required for fund performance evaluation.

With interest rate risk identified as the dominant risk factor (Litterman and Scheinkman 1991), and the relatively predictable nature of bond cash flows, the performance of actively managed bond portfolios could perhaps be perceived as a relatively simple process; at least when compared to the heterogeneity that exists with equity funds. (Campbell, Lo, and MacKinley (1997)). This proves not to be the case in many respects. For example, nonlinearities inherent in the price-yield relationship potentially impose biases when assessing fund manager skill. Furthermore, there is no consensus definition of a bond market portfolio, likely adding greater subjectivity across studies in this area.

It is evident that although the underlying theory is well established and tested, the evaluation of active management continues to be a dynamic research topic. This thesis does not attempt to address all the issues as discussed; however, it is clear from a review of existing literature framed in the context of various industry developments, that there is much scope for further research in the fund performance area; particularly in relation to fixed income funds.

1.2 Why this topic warrants further research

A cornerstone of fund performance academic literature is the concept of mean-variance efficiency, as developed in the seminal work by (Markowitz 1952). In the context of portfolio management, this proposes that managers should be motivated to optimise their portfolio within the available investment universe, aiming to maximise return for a given level of risk. Central to the assessment of whether this is the case, is the application of an appropriate model to do so . Here, a reliable model should be mean-variance efficient and comparable to the funds on a risk-adjusted basis. In this respect, the selection of factor(s) in the model must represent an otherwise equivalent strategy to the funds in question(Aragon and Ferson 2006). Many interesting research questions grounded in this theory have already been addressed; however, this thesis identifies that there are still significant gaps in the fund performance literature that warrant significant further study.

With the merits of active management continuing to face increased scrutiny, and results of studies often concluding that investors would be best advised to select low-cost passive vehicles, it highlights the importance of the accuracy with which the performance is being evaluated. To date, there is no bond fund performance academic literature specific to the UK market. Therefore, it makes for a new setting in which to evaluate candidate models regarding their reliability for bond fund performance evaluation. Furthermore, the data available in this thesis allows for analysis over some interesting macroeconomic conditions, such as the global financial crisis of 2007 to 2009 and the historically low interest rate environment that has been prevalent since. This thesis, to the best of my knowledge, will be the first to study the performance of UK bond funds in such a setting, having specified a new set of candidate models to do so.

Regarding the selection of a methodology for this purpose, a benefit of the linear factor model approach Jensen's alpha (1968) and the comparison to passive benchmarks is that it is relatively easy to interpret. For example, although the indices used as factors are themselves not directly investable, they may be related to ETFs/Index trackers to evaluate potential investment alternatives. This is particularly relevant given the prevalence and rising interest in these vehicles that are accessible at a much lower cost than their active counterparts. The dominance of interest rate risk presents the opportunity for application of term structure models for fund performance evaluation (Ferson et al. 2006) This method is far less common for this purpose. Nonetheless, an application to the UK market may present extensions of this initial study regarding bond fund performance.

Aside from determination of skill and identifying the prevalence of alpha, the results of fund performance studies may have wide-reaching implications in with regards to related topics of interest. The basis upon which investors allocate their capital is central to the flows of funds literature; for example, they may chase alpha, or various other characteristics. Bond funds in this context have been receiving increased coverage in the academic literature, as potential systemic risks are highlighted due to the increased flows but reduced liquidity. Some interesting points to note highlight concerns raised with pricing transparency and hence performance manipulation (Cici et al. 2011) that may encourage accumulation of flows. Thus, it is evident that bond funds have become an area of increasing attention and further research regarding their activities is clearly warranted.

A review of the existing academic literature provides minimal evidence to overturn the general consensus that active management most often fails to deliver outperformance on a risk-adjusted basis. Many studies find that funds underperform benchmark models by approximately a magnitude of costs, most commonly using the method of Jensen's alpha (1968). To date, there is limited academic literature specifically directed to evaluating bond funds in this respect. Furthermore, the majority that does so is focused upon the US market. (Ferson (2006a), Ferson (2006), Blake et al. (1993), Elton et al (1995)) The UK features as part of a wider European sample (Silva et al. 2003), however, the coverage is limited as only 45 UK bond funds are included.

As discussed, the UK asset management industry has been subject to many notable changes. These are likely to have important implications at present and are projected to do so for years to come, for example due to the Pension Freedoms reform and the ongoing Brexit negotiations. Given this, it is of even more relevance to ensure that the investment opportunities are considered separately from the rest of Europe. Thus, the UK market is highlighted as one in which there much scope to assess various research questions, aside from the more common US focus. By doing so, it will allow for insights to be gained from a new sample of bond mutual funds whilst using traditional methods of performance evaluation. A study of the merits of active bond management in the UK may help to inspire further thought as to how such performance is derived, as this continues to be an area attracting further attention in both academia and practice.

1.3 Research aims and key findings

The overall aim of this thesis is to evaluate the performance of a new sample of UK bond funds using data obtained from the Morningstar Direct database. To do so, candidate models are first required to be specified and tested to determine which is most appropriate to apply in the performance evaluation. As noted, this is the first study to use such a sample and therefore asset pricing models have not already been specified for this purpose. The sample of UK bond funds is comprised of various style-specific categories which include; Corporate, Government, Diversified, High Yield, and Flexible bond funds. These will be the subject of the three empirical chapters. Further details are outlined as follows with the key findings of each chapter summarised.

1.3.1 Model testing

Consistent with the application of Jensen's alpha methodology as mentioned in section 1.1, five candidate asset pricing model are specified; two single-factor, and three multi-factor models. These are constructed based upon models prevalent throughout the bond fund performance evaluation literature and are representative of risk factors known to explain the cross-section of bond returns, i.e. term and credit risk (Litterman & Scheinkman, (1991) Gebhardt et al., (2005), Fama & French, (1993)) The objective of this first empirical chapter is to identify which of these can be considered most reliable for the subsequent performance evaluation. The Gibbons, Ross, and Shanken (1989) test is a classic method by which to evaluate the mean-variance efficiency of benchmark models. However, it must be noted that this is an absolute test of model validity, but it cannot be used for a relative evaluation of the candidate models. For this purpose, additional alpha statistics are required. Example of such application are used in papers by Fama and French (2015, 2015b) whereby they augment their three-factor (1993) model with additional factors to represent profitability and investment. Furthermore, given that determination of alpha significance is a key research objective of performance evaluation studies, it makes sense to focus on the alpha statistics more so than the GRS test to determine the extent of biases that may be expected. In existing literature on bond fund performance evaluation, tests of model suitability of this type are minimal. Ferson et al (2006a) do conduct tests of their candidate models in the stochastic discount factor (SDF) framework. This is implemented by means of joint estimation using Generalised Method of Moments (GMM). Other statistics of model fit have been used at times, for example Ayadi & Kryzanowski (2011) use the Schwartz Information Criterion statistic, and many others use the adjusted R^2 to make relative judgements. A shortcoming of these, however, is that they don't give an indication as to the extent of alpha bias that may be induced. Additionally, it has been proposed that bond fund performance evaluation is not as sensitive to the selection of model as is the case with equities, given the dominance of the systematic as opposed to idiosyncratic risk factors. Thus, a comparison of model reliability provides a further test of this perception.

The test assets are a selection of indices provided by Markit iBoxx, representing passive portfolios. These cover a range of maturities across both credit and government exposures. As such, they should provide a reasonably complete cross-section of the investment universe available to the actively managed funds under evaluation. The hypothesis to be tested is whether the alphas across the test assets when evaluated using each candidate model are jointly equal to zero. Testing is conducted over three periods in relation to the macro-economic conditions prevalent at such times. The whole period available is from April 2002 until July 2016, from which two sub-samples have been created; April 2002 until August 2009 and September 2009 until July 2016 .

Overall, a five-factor model representing maturity buckets of term and credit (named the Maturity 5 model) is consistently found to be the most reliable in terms of minimising bias, as it is associated with the lowest absolute alphas and tightest fit. This is evident from the high (adj)R^2 and low standard error of alpha. The results, however, highlight the somewhat counter-intuitive nature of the GRS test, if forming judgments based on this statistic alone. For example, during the recent sub-period from 2009 to 2016, all the candidate models fail the GRS test, despite exhibiting preferential characteristics for fund performance evaluation, i.e. tiny alphas and tightness of fit. This highlights the importance of extending the analysis beyond a single decision criteria, such as a p-value, when evaluating a research question. Considering the full range of information available, the Maturity 5 model is concluded to be well-specified, with minimal potential for alpha bias to be induced.

1.3.2 Performance evaluation of UK bond mutual funds

The second empirical chapter seeks to identify if there is evidence of superior performance relative to the benchmark model, Maturity 5. The results indicate that over the whole period, from January 1999 until July 2016, the actively managed UK bond mutual funds underperform when net returns are used. The category of All funds shows a significant alpha of -0.06%. In terms of style-specific performance, the High Yield category exhibits the least negative results, with an insignificant alpha of -0.01%. On the other hand, the worst performing are the Flexible funds, exhibiting significant monthly underperformance of -0.09%. Here, the findings are consistent with much of the academic literature; when evaluated at an aggregate level, the active funds fail to outperform the passive benchmark after risk adjustment and costs. However, although the alphas realised are statistically significant, their magnitude is tiny and economic significance minimal. For example, this is by comparison to findings of equity fund performance studies. Additional regressions confirm that expenses are a significant determinant of the negative performance observed.(Ferson et al., (2006), Comer et al., (2013)).

Additionally, the sample has been split into sub-periods. It appears that the results are more encouraging using the recent data since 2009. Wald tests have been used here to confirm that the performance here is significantly different relative to the whole period (1999 to 2016) considered. The findings show that the Corporate, Diversified, and Government funds all realise improvements. With evidence of improved performance identified in the post-crisis period, it would be desirable to isolate the driving factors behind this. It is unfortunately the case that this is a complicated task when it comes to bond funds. This is due the interaction of term and credit effects the complexities involved in splitting them. However, it is possible to achieve more general insights with regards to positioning relative to the market. To do so, the timing measure of Treynor and Mazuy (1966) has been applied.

The Barclays Sterling Aggregate index has been used as a bond market proxy, with additional tests conducted using the style-category model as specified by Morningstar. Due to the vast cross-section of bond markets it is notoriously difficult to capture the most relevant characteristics of the strategy. By additionally using a style-based model it provides an interesting comparison and perhaps a more representative investment universe. The results show that there is minimal evidence of timing ability relative to the Aggregate index. This is generally consistent with existing literature, albeit that which is focused on the US market. However, a different picture emerges on a style-specific basis. Here, 25% of the Corporate and 4% of the Government fund show evidence of positive timing ability. This appears to be most pronounced when both the global financial crisis and low interest rate environment are considered. Timing of credit spreads is a strategy potentially undertaken by fund managers on a tactical basis. It is possible that this is the case, whereby they are optimally adjusting their positioning. However, research tends to find that exposure to the credit risk premium persists; potentially bringing with it negative consequences due to the dilution of diversification potential with equity markets. Further granularity regarding the fund managers' actions during this time would make for an interesting are for future research.

Overall, the findings from this chapter suggest that active UK bond fund managers struggle to return value to investors on a net basis. There is evidence of time-variation, with the low yield environment since 2009 appearing to be conducive to better performance across the Corporate, Diversified, and Government categories. It has been noted that such times have been challenging for fund managers; however, the results here suggest there is potential to identify positive aspects of performance and one which warrants further investigation. To add robustness to the results observed, various methods are applied in Chapter 6. This helps identify to what extent outperformance is due to luck or can indeed be attributed to truly superior ability.

1.3.3 False discoveries in fixed income

A limitation of using portfolios of funds to assess the performance in aggregate is that the extreme performance is likely to be masked. This is unfortunate, as it is likely to be these funds which are of most interest to investors, i.e. the very worst and the very top performers. The analysis may be conducted by applying OLS to individual funds; however, this does still not provide conviction as to whether the results are spurious or are indeed due to skilled management. A bootstrap simulation can be applied to help provide some further insights in this respect. The bootstrap has continued to gain popularity, in part due to advancements in computational power that allow for relative ease of implementation and the flexibility that it affords to address a wide variety of statistical applications. The imposition of zero alpha on the simulated distribution allows for hypothetical performance to be derived in the absence of superior performance. By doing so, this can be compared to what was observed from the active funds in the sample. Various bootstrapping techniques may be employed when using regression analysis. For example, the residual resampling with replacement method of Kosowski et al. (2006) and entire-cases resampling of Eugene Fama and Kenneth French (2010) are two commonly used methods. However, aside from this thesis, the entire-cases resampling technique is yet to be applied to a sample of bond funds.

The key results from the bootstrap show that there is evidence of superior performance, even when net returns are used. From 1999-2016 this is evident across all funds in the 98th and 99th percentiles. Likelihood statistics have also been calculated to add further conviction to the extent of superior performance in each percentile. These indicate that approximately 76%, and over 90%, of the actual alphas were greater than the simulated in these percentiles respectively. In terms of style-specific performance, the

Corporate bond funds show the most potential to add value, with significant superior performance apparent from the 97th percentile. On the contrary, evidence of ability is scarce for the Government and Diversified funds, as it is not until the 99th percentile for these categories that this shows for the majority. The strongest evidence of positive alpha is apparent from 2009-2016. Here, outperformance is evident from the 96th percentile when all funds are considered. This is primarily driven by the Corporate and Diversified categories. The results are robust to altering the minimum return requirement and alternative model specification.

Despite the improvements afforded by the bootstrap, it is still prone to the multiple hypothesis testing issues that plague empirical work. This is the case when conclusions are derived from statistical techniques applicable to one hypothesis, yet are applied across multiple. When doing so the decision rules regarding significance, and ultimately outperformance, are biased. A relatively straightforward way minimise such issues is to adjust for False Discovery Rates - i.e the classification of funds as skilled when this is in fact not the case. In this final empirical chapter two methods are applied; the classical approach of Barras et al. (2010), and the recent version of Ferson and Chen (2019). There is yet, to the best of my knowledge, a study that applies either of these methods to a sample of bond funds. The findings from the BSW (2010) approach suggest that superior performance is isolated to the post-crisis period. Here, approximately 6% of all funds are classified as good. In the left tail 19% are classified as bad, which is a notable reduction compared to the whole period whereby the proportion was 42%. The Corporate category performs well, with 10% exhibiting positive performance and only 14% negative in the post-crisis period. Consistent with the findings in Chapter 5, expenses are shown to be associated with poor performance. The extent of superior results are even more widespread when the Ferson and Chen (2019) method is applied. In this case, given a 10% test size, 11% and 33% of all funds are classified as good during the whole and recent periods respectively. These proportions are even more pronounced for the Corporate funds, whereby these numbers increase to 17% and 48% for the corresponding periods

With Corporate funds appearing to offer the most potential for superior performance, further analysis on this category may help to identify the driving factors behind this. These funds are shown to exhibit a higher correlation with the High Yield index in the post-crisis period. As such, this could be indicative of reaching-for-yield tendencies during the low-yield environment. As a basis of comparison, relative to a passive proxy, the funds show this to a greater extent. This is understandable given that funds are likely afforded more flexibility, relative to the rules-based composition of benchmarks. Literature has identified that premium received for doing so is fully explained by the additional risk assumed. Reaching for yield may increase the returns on an absolute basis, but whether this can be defined as alpha is debatable. Furthermore, the diversifying properties of such investments in a wider portfolio context (e.g. 60/40) may be diluted due to the increased correlation with equities. This is in fact evident. The average correlation since 2009 is 0.43, whereas prior this it was 0.08. Overall, the results in this chapter provide conviction in terms of the superior performance post crisis, however, the extent to which this may be defined as true alpha remains a generally open and interesting question.

1.3.4 Implications and main contributions

Given the scrutiny active management faces, it highlights the need to be drawing conclusions with as much precision as possible. Hence, model testing is required to be a key component of such studies, as examined in the first empirical chapter. This thesis is the first to conduct model testing in the bond fund performance evaluation literature using a sample of UK data. To address further research questions in future work, the models may be easily extended to accommodate various additional aspects of performance evaluation; for example, conditioning information. Another relevant consideration and justification for conducting such an analysis, is the sheer magnitude and vast cross-section of bond markets. For example, for the one issuer, there may exist various bonds all with different characteristics. The complexities inherent in bond market replication give rise to scope for off-benchmark positions.

The second empirical chapter is concerned with the initial assessment of active management. An interesting finding here is the significant difference in performance observed between the isolated time-periods. This is a key consideration for investors, from both a tactical and strategic perspective. By analysing specific periods, it helps to provide further insights as to the investment opportunities at such times. Perhaps most notably, it helps to provide insights as to whether managers can perform during times of market stress and generally adverse market conditions when utility is highest for investors (Kosowski 2006) With regards to the significance of fees, the findings from this thesis provide further evidence that management costs are a notable detriment to performance. Overall, the contributions from this chapter encourage further analysis of time-varying performance. In terms of "value" being returned to investors, if this is to be defined as net alpha then lower costs funds should be selected based on the results here. Managers may engage in timing credit spreads to capture premiums. The availability of Corporate bonds at the time under analysis evolved due to various factors. For example, "Fallen Angels" are a results of credit downgrades from Investment Grade to High Yield. It has been identified that these are likely to be more accessible to funds than is the case for benchmarks due to rules-based composition restrictions. Furthermore, there is evidence that issuance of Sterling credit declined during this time, whereas High Yield issuance increased. If investment mandates allow so, industry-wide changes such as this may allow the Corporate funds access to a wider scope of credit opportunities than was otherwise available.

Additional robustness tests prove to be valuable here. The application of the bootstrap method is an initial means to do so. This thesis, to the best of my knowledge, is the first to apply the entire-cases bootstrapping approach to a sample of bond funds. Previous studies have adopted the more commonly applied residual-only approach (KTWW (2006)) which focused on a sample of Canadian bond funds (Ayadi and Kryzanowski 2011) With interesting results identified whereby improvements are realised in the post-crisis period, adjusting for luck using two False Discovery Rate (FDR) methods adds further conviction. Again, these have yet to be applied to bond funds. The findings make a valuable contribution to an expanding field of research that seeks to address the multiple hypothesis testing issues that plague empirical work with regards to the accuracy of inferences that may be drawn. The literature thus far has focused on equity and hedge funds, however, there is notable scope for further research regarding the optimal empirical methods specific to bond funds. The findings from this thesis suggest there is much to be gained from continued work in this area.

The main contributions to the literature from this thesis can be summarised as follows: (1) The findings provide an extension to the UK literature by focusing on bond funds, as equity funds tend to dominate (2) To the best of my knowledge, there have not as yet been any studies that have specified and tested candidate models for the purpose of UK bond fund performance evaluation (3) Application of the entire-cases resampling methodology (Fama & French (2010)) yields results that provide support for the merits of active bond fund management, most notably so using recent data since 2009 (4) This improved performance is further confirmed by the False Discovery Rate (FDR) methodologies, providing more convincing evidence that positive alpha maybe achieved on a risk-adjusted basis. Overall, the findings here provide a sound basis from which many further research questions may be explored. This is particularly relevant given the continued scrutiny and pressure faced by active management regarding its merits and fee structure.

1.4 Suggestions for future research

With superior performance having been identified in this thesis, further insights with regards to how this is generated would be desirable. It appears that the Corporate bond fund category has the most potential to generate positive alpha and return value to investors. At present, it is acknowledged that limitations exist with current methodologies that make skill identification more problematic with bond funds relative to equities. This is due to certain characteristics of bond funds, such as non-linearities (Chen et al., (2010)), and the style-based factor models are not available with regards to bonds as they are with equities. By isolating the underlying sources of return in the context of style-factor definition, this allows for an additional way to provide further insights regarding security selection within the term and credit components and identify what is driving the performance observed. (Ang (2009), (Houweling & Zundert, 2017), (Brooks & Richardson, 2018)) Investment vehicles adopting such a strategy seek to consistently outperform the benchmark by systematically implementing the fundamental drivers of fixed income returns; value, momentum, carry, defensive. Israel et al. (2018) attribute the performance of US mutual and hedge funds to these factors. An important finding here is that the managed funds are under-exposed to characteristics that generate meaningfully positive risk-adjusted returns. Furthermore, the mutual funds are found to have a high exposure to carry which adds potentially undesirable, greater implicit market risk to investors' portfolios. Thus, it is highlighted that alternative ways to assess the selection abilities of bond fund managers presents an area of growing interest and one in which many further insights can hopefully be derived.

Bond funds may be considered as an attractive investment opportunity with a view to gaining diversification benefits (Boney et al., (2009), Chen et al., (2010)). However, correlations between equity and credit markets are relatively high, particularly so during crisis periods. Therefore, opportunities for diversification may be limited unless a more targeted approach is applied. For example, Houweling and Van Zundert (2017) and Brooks and Richardson (2018) find that combinations of the fundamental factors identified to drive corporate bond returns are diversifying in this sense. Low or negative correlations are found between traditional fixed income and equity market returns. Overall, the use of style-based premia allows for greater granularity regarding managements' active decisions and may help to add further dimensions to the skills literature. This applies to both the security selection decisions of managers and optimising diversification benefits in a wider portfolio context.

Chapter 2

Research Positioning

This chapter presents a discussion of academic literature in conjunction with notable industry events and developments, that help to frame a detailed analysis of bond fund performance as an important area for further research.

2.1 Introduction

A review of significant developments within the asset management industry identifies interesting perspectives that may inspire additions to the existing academic literature. For example, Exchange Traded Funds (ETFs) not only provide low-cost passive alternatives to their active counterparts, but it has been identified that they have an impact on the valuation of underlying securities (Dannheuser 2017). Changes in the overall dynamic of the asset management industry have been cited in studies (Fama & French 2010) as a determinant of the time-varying performance observed. This may be attributed to increased competition and range of vehicles available to an extensive investor base. Given that the proliferation of low-cost funds and their corresponding impact is a relatively recent development, it makes for interesting opportunities to assess the extent to which the active performance observed is consistent with prior or is indeed indicative of industry evolution.

The low-yield environment since the global financial crisis has encouraged new perspectives on how to generate value. This relates to both defensive strategies, and those taking a more aggressive approach with regards to deriving alpha. With the level of interest rates at historic lows, the scope for price appreciation is minimised as they cannot fall much further. For defensive strategies, whereby the aim is to minimise interest rate risk, a more flexible approach than usual may be required given the increased uncertainty with regards to interest rate policies. In terms of the role of fixed income in a wider portfolio context, it may also be the case that the defensive role is not fulfilled to as great an extent. For example, higher correlations between credit and equity exposure is likely to reduce the benefits from diversification during times of market crisis. Thus, it is evident that bond fund managers have faced a challenging environment. As such, many interesting areas that warrant further research are presented with regards to their performance in these unique circumstances of late.

The purpose of this chapter is to highlight the positioning of this thesis in terms of academic contributions, and recent developments primarily within the UK asset management industry. Points that will be addressed include: 1 - Why is the UK market of specific interest? 2- What key developments have helped to position this research? and 3 - To what extent does existing academic literature already address similar themes? Section 1 provides an overview of the UK asset management industry, with a focus on key issues and trends that relate most specifically to fixed income products and investors. Section 2 reviews associated academic literature. Section 3 introduces the challenges faced from new product development. Section 4 discusses the risk factors to which bonds are exposed and the implications for model construction, and section 5 concludes.

2.1.1 Flows of funds, liquidity, and new product development

An understanding of the asset allocation behaviour of investors is of interest to managers, policy makers, and academics alike. There is an extensive literature regarding flows of funds, with studies aiming to identify the drivers of investors' allocation decisions and the impacts on the markets and performance that may result. The Investment Association (IA) conducts an annual survey of Asset Management in the UK . The survey is a useful resource for analysing trends within the UK asset management industry, and predictions of those to come.

The IA note that the asset mix has historically been dominated by equity funds. For example, they represented 87% of funds under management in 1995, however, only around 55% in 2014. Various factors may be responsible for such a shift - with increased allocation to fixed-income being one of them. It is interesting to observe from the survey the time-variation in flows of funds. For example, both equity and fixed income funds received high inflows in 2009. This is consistent with the notion that investors prefer to allocate to experienced managers rather than rely on their own abilities during times of volatility and market stress.

It has been a long-debated issue in the academic literature as to whether active managers have enough skill to justify their fees and add a source of value for investors, over and above the passive alternatives. A distinct lack of consistent outperformance is cited. Morningstar statistics show that in 2016 (as at end of May 2016), $213bn has been lost from active funds whereas $240bn has been gained by passive funds on a global basis. An upward trend in passive AUM has been observed since 2008. The IA's 2014 survey recognises for the first time that there has been a development of strategies that may not be defined specifically as active or passive, such as "Enhanced Index", or "Smart Beta". However, the growth in passive AUM is likely to be understated in this survey as it does not fully incorporate the ETF market. Therefore, the shift to passive is likely to be more pronounced. The major influence upon the active vs passive debate is costs.

In relation to new product development, Dannhauser, (2017) studies the impact of innovation in the corporate bond ETF market on the underlying securities. Regulators such as the Financial Stability Oversight Committee have previously voiced concerns regarding the potential systemic risk they may present with respect to the liquidity discrepancies. ETFs are very liquid and easy to trade; however, the actual underlying securities are not. Overall, the author finds that there is evidence of a positive valuation effect on the market due to innovation of the Corporate bond ETFs; citing benefits such as previously unavailable intra-day pricing and an enhanced investment universe of index products.

The price discovery properties of ETFs have also been highlighted in recent research. The value of the underlying securities is the primary determinant of the price. However, it is also influenced by demand, flows, liquidity, and market volatility. The relationship between the price of the ETF and the value of the underlying securities may give rise to arbitrage opportunities. If significantly above the value of the underlying, then it could be profitable for arbitrageurs to trade the purchased with the ETF provider, exchanging them for new ETF shares. With the price being high at that point, the opportunity in the market to sell them for a small profit arises. In this respect the converse also applies. Thus, such actions mean that the ETF price is not likely to stray too far from NAV, in either direction for any length of time. As a result, various benefits of ETF price discovery have been identified. These include the observation that the observed premium/discount may reflect price discovery and not actually mispricing, ETFs are quicker to capture and show value changes in investor sentiment, and finally that the price discovery allows investors a way to identify and potentially capitalise on relative value opportunities. These could be accessed in the markets in terms of ETFs vs over-the-counter (OTC) valuations, or perhaps otherwise through associated derivatives.

Thus, it is evident that investors are responsive to the continued development of new products and shift their capital accordingly. Fund flows play a key role with respect to gaining insights as to market sentiments and form the basis of much academic literature. Notable papers in this area with regards to fixed income are discussed throughout the remainder of this section.

In light of volatile equity market conditions during the crisis, fixed income funds received their highest volume of flows in 2009 as investors sought relatively safe haven assets. However, the experimental monetary policy that has since ensued has contributed to a challenging environment for fixed income fund managers, with continued uncertainty regarding the direction of interest rates. Additionally, regulations that have been enacted to stabilise the banking system have been thought to have had an impact on fixed income markets. Investment banks were the primary market makers in the instruments, however, due to capital adequacy requirements imposed they have since retreated from such activity, therefore potentially contributing to liquidity issues . With all these factors combined, it has raised alarms as to the potential for further systemic risks within the asset management industry, particularly in relation to fixed income funds. For example, in their discussion paper "Market Tantrums and Monetary Policy", Feroli, Kashyap, Schoenholtz, & Shin, (2014) proposed that policy makers must take account of the asset management industry before making interest rate adjustments; highlighting the increased volume of assets in these funds and reduced liquidity.

In relation to such concerns, Goldstein et al. (2015) focus on US corporate bond funds over the period 1992-2014, whereby a flow of funds analysis can perhaps provide further insights regarding potential systemic risks. They note that there has been a significant increase in the AUM of these funds in recent years and propose they may be a relatively ignored source of fragility in the financial system. The flow relation observed for the corporate bond funds is concave; they are more sensitive to poor performance than they are to positive performance, i.e., the funds experience a greater outflow as a result of bad performance, than they receive due to good performance. The potential for fragility is highlighted due to this relationship being exaggerated in the presence of illiquid assets.

Cici, Gibson, & Merrick (2011) investigate the transparency regarding the marking of corporate bonds by mutual funds. Marking refers to pricing of bonds – options for doing so include model and/or market. There are various ways a mutual fund could mark its holdings; thus, allowing the managers significant discretion and scope for manipulation. The study looks at the time-series of bond dispersion. As such, the observations can be related to the crisis and the impact of increased regulation. Return smoothing would impact the risk and return profile of the funds. As such the key performance metrics that are used by investors will be biased, and therefore may lead to their allocation decisions being misinformed. The findings from this study show that it is the funds that underperform their benchmarks are more likely to smooth their returns. The converse also applies. It is noted that much of the existing literature focuses on hedge funds, however, more attention is needed for mutual funds, given that much less attention has been focused on the latter from a regulatory standpoint. Also, there is much more granular data available for mutual funds. As such, this paper uses actual portfolio holdings, allowing them to construct a holdings-based custom benchmark to identify the marking dispersion.

From a regulatory and policy making perspective, flows of mutual funds have received attention in relation to the predictive content they may provide.Chordia et al. 2005 show that fund flows are an important area to consider when making linkages between the micro and macro structures of financial markets. Liquidity dynamics are a highlighted area of concern. They find that bond fund flows are useful in terms of the predictive content they can provide, able to significantly forecast increases in bond spreads. Thus, they conclude that further understanding of flows maybe helpful in terms of policy decisions and stability.

The factor structure of mutual fund flows is not such a commonly addressed research area as is the factor structure of the returns (Edwin J Elton et al. 1999). Ferson and Kim 2012 use statistical factor analysis to identify commonalities between the flows of US bond, equity, and money market funds over the period 1981-2009. The factor extraction method used (Connor and Korajczyk (1988)) allows for missing data, i.e. they can accommodate funds with short time series. Flow factors are extracted separately for the funds, allowing for identification of variation between sectors. Their analysis shows that the first few factors capture more than 40% of the variance for equity and bond funds, slightly less for money market funds. They find there is substantial variation across individual mutual funds in the sensitivity of flows to common factors. The flow betas are modelled as functions of the characteristics of a fund, including age, size, and expense ratio. Strong evidence for time-varying flow-betas is identified. The authors address this by using a rolling estimation. An interesting finding is the predictive power of the flows. They find that the lagged flows bear a predictive relation to future values of several variables representing economic conditions. This suggests that the investors are not simply chasing past performance but are also looking to the expected future economic conditions. It is identified that the flows have predictive power for future economic growth and interest rates.

With fixed income fund flows identified as an important indicator in many respects, it highlights the importance of research with respect to the underlying methods and performance metrics that are informing these allocation decisions. On a related note, various characteristics and economic conditions have been found to be influential in determining the flows of funds. Chen. and Qin (2014) find that corporate bond fund flows are sensitive to both fund performance and macro conditions. In contrast to Goldstein et al. (2015), this study finds that the reaction of investors to poor performance is to the same extent as that of positive performance, as opposed to more concave as otherwise identified. Flows to high-quality bond funds are more sensitive to prior performance than flows to high-yield bond funds are. By using data at the share class level, as opposed to per fund, it allows for further analysis as to the types of investors behind the flows. For example, distinctions can be made between retail and institutional. Furthermore, different levels of fees and restrictions often apply across the range of share classes available; some may charge high load fees and impose lock up periods, perhaps deterring investors.Fama and Macbeth (1973) cross-sectional regressions are used to examine the relation between flows and recent fund performance. The results obtained from using the lagged returns suggests that investor flows chase past performance. They find that a 1% increase in fund return in the past year is associated with 0.41% increase in fund flows in current month. There is some evidence that shows that fund flows are negatively associated with volatility, as represented by VIX, and negatively with expense ratios.

In terms of return chasing, Zhao (2005) assesses the behaviour of US bond fund investors over the period 1992-2001, determining their reaction to returns calculated by different methods. The results indicate that they chase risk-adjusted leaders, as opposed to raw return leaders. Furthermore, an inverse relation between bond fund flows and equity market performance is observed; when the markets are down, bond funds will receive higher levels of inflows. This is consistent with the data published by the IMA citing that 2009 and 2010 saw the highest volume of flows to bond funds. Two flow measures are noted – the dollar amount of flows, and the commonly used measure that is calculated based on the growth rate of assets net of Holding Period Returns (HPRs). In this study, the former is the preferred choice as it allows for a better proxy as to where investors are in fact putting their money. Various determinants of flows are considered – fund size, performance, expenses, fund age and turnover, average maturity, total AUM in fund family, and a bond fund family dummy. The Risk-3 measure of Blake et al. (1993) is used for risk-adjusted performance, incorporating government, credit, mortgage backed security, and high yield factors . The Sharpe ratio is the other measure of risk-adjusted performance they use as a flow determinant. They find that investors in High Yield bond funds do not chase absolute performance and are more likely to rely on brokers and financial advisers.

A summary of the academic literature regarding fund flows highlights that bond fund performance is a key area of interest for investors and policymakers, with much scope for further academic research. Investors are identified as being reactive to industry offerings and are becoming increasingly demanding with respect to the performance and transparency with which this is disclosed. It is evident that fixed income funds have been attracting increased flows, much of which has been relative to the changing macroeconomic conditions of recent years. Further research as to the performance of such funds and the methods used to estimate the value returned to investors is clearly warranted. As noted, the performance metrics are a key influence on flows of funds and the repercussion of this that may ensue. Therefore, the models used to derive this are an interesting area for further research. The following section will introduce aspects of the basis upon which asset pricing models are constructed for the purpose of performance evaluation., which are also related to new product development.

2.2 Bond risks and sources of return

With the purpose of this thesis being to conduct a performance evaluation of a sample of bond mutual funds, appropriate asset pricing models are first required to be specified. The following discussion highlights the significant risks to which bonds are exposed and will thus form the basis of these models. Performance evaluation studies most often use a variation of a style factor model, incorporating exposures to underlying sources of return generation; for example, value, growth, size, momentum, as captured by the Fama and French (1993) three-factor model and Carhart 1997 four-factor model. However, this is not the case with fixed income. It has been identified that there is a growing interest with the applicability of factors in credit; thus, this is an area of interesting research and one in which is there is much scope for future work. This will be briefly outlined in section 2.4.3. Additionally, comparisons are made with the factor modes most commonly used for performance evaluation studies, although these relate to equities.

2.2.1 Term risk

Litterman & Scheinkman, (1991) conduct Principal Components Analysis (PCA) of a sample of zero-coupon bonds. They identify that the first three components, level, slope, and curvature of the yield curve explain over 96% of the variation in bond returns, with the level factor dominating. Their initial analysis of zero-coupon bonds can also be extended to coupon bearing securities. In terms of performance evaluation, the level factor as measured by duration is considered analogous to beta in equity markets. With the influence of changes in the yield curve clearly identified, central to the performance of bond funds the ability of the manager to manage this interest rate risk. As such, the variation of returns to fixed income are predominantly driven by systematic factors. This is on the contrary to equity investments, whereby they have a large idiosyncratic risk component of around 80-90%, which is thus diversifiable (Ang 2014). The term premium captures the difference between the yield from holding a long relative to a short-term government bond. The short rate captures the level of interest rates, and as such the long rate captures the expected deviation from this. Therefore, this isolates the effect of interest rate changes on the return earned. Fund managers will be seeking to capitalise on predicted interest rate movements by means of duration management, which may involve significant use of derivatives.

2.2.2 Credit risk

Despite the dominance of the Term factor, there is still scope to access and capitalise on return generated from the credit risk premium . Mixed evidence exists as to the extent of this, and it is an interesting area gaining further attention in academic literature.

The default premium can be defined as the excess return from a corporate bond over a government bond of the same maturity. The literature has provided little historical evidence of a credit risk premium (Fama & French (1993), Ilmanen (2011)). In their analysis of the common factors in stocks and bonds, Fama & French (1993) show that it is evident that the returns on both the government and corporate bonds are well explained by the Term and Default factors. However, the average risk premiums are very small, almost zero. From this it can be inferred that there is not a significant excess premium to be gained from holding corporate over government bonds. In terms of return variation to be explained, the mean returns of the bond test assets are all less than 0.15% per month. This is less than half of what the stock test assets return on a monthly basis, as they range from 0.32% per month to 1.05%. The term premium is 0.06% per month and the default premium is 0.02%. Although these are very small, the factors are about as volatile as size and value and are therefore able to capture a large proportion of common variation in returns.

Thus, not much convincing evidence is presented as to the value added by corporate bonds. However, the findings of Asvanunt & Richardson, (2015) attribute this to the way by which the credit excess returns are calculated; usually without a correct adjustment for term risk. Their findings show that a positive credit risk premium does seem to be prevalent over their long time period, having incorporated duration adjustments into the calculation. On an annualized basis, Investment grade corporate bonds return 137bps in excess of Treasuries from January 1936 until December 2014, with a Sharpe ratio of 0.37. From August 1988 until December 2014, a higher premium is found for high yield corporate bonds; 248bps, Sharpe 0.26. The difference between the calculations of the credit risk premium in this paper relative to other studies that have previously documented a mutual evidence of a credit risk premium is the duration adjustment. Corporate and government bond returns will have different cash flow maturity profiles, and as such corporate bond have lower interest rate duration. Due to the term premium being captured within this, when isolating the credit risk premium, it appears to be understated.

To put their findings of the credit risk premium in context, it far exceeds what Fama & French, (1993) found as the reward for exposure to their default factor which was only 2bps per month. However, as the sample periods differ, the results are not directly comparable on that basis. Therefore, as a robustness test Asvanunt & Richardson, (2015) compare the credit risk premium over their sample period based on the unadjusted method Fama and French (1993) relative to their modified version – the results are 7bps and 137bps respectively, clearly highlighting the variation between the different measures. The duration adjustment is only required to be calculated for their older sample period, as index providers such as Barclays Capital provide the recent data which already includes this modification.

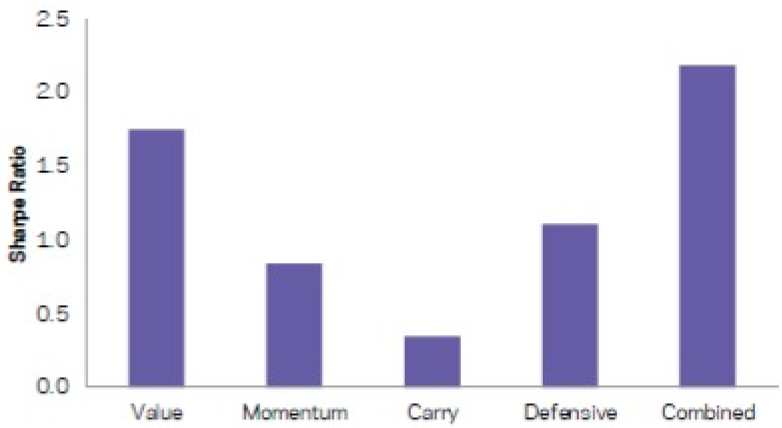
Throughout the academic literature, the models used for bond fund performance evaluation are most often comprised of a selection of indices that proxy for the exposure to both term and credit risk. This is the case when the method of Jensen's alpha is applied, whereby the models assume a linear functional form and the factors have been selected to represent a passive alternative on a risk-adjusted basis. It is on this basis upon which the models used throughout this thesis will be based (Chapter 4). However, it is worth noting some interesting perspectives with regards to skilful selection within each of these exposures and how the model construction for fixed income differs to equity factor construction.

2.2.3 Style-based premima

With term and credit factors known to be the drivers of bond returns, from a portfolio performance perspective it makes for an interesting consideration regarding how managers may access and capitalise on the associated premiums. Style-based sources of risk premia are well known in the equity area; for example, size, value, and momentum (Fama and French (1993), Carhart (1997)) but not so commonly studied in bonds. However, it has been acknowledged that they are as relevant in fixed income markets as they are in equities. Recent research has begun to examine these underlying sources of return in more detail. For example, with the overall returns to fixed income driven by rate and spread components it is of interest to identify managers' security selection abilities within each of these exposures, as this drives the generation of excess returns over passive fixed income indices (Brooks & Richardson, (2018)).

Investment vehicles adopting such a strategy seek to consistently outperform the benchmark by systematically implementing the fundamental drivers of fixed income returns; as defined in the fixed income space as Value (Correia et al. 2012), Momentum (Jostova et al. 2013), Carry (Koijen et al. 2018) and Defensive (Ilmanen 2011). Brooks and Richarson (2018) find that for each of these style factors, there is evidence of positive risk-adjusted returns. A combination of styles across both government and corporate bonds shows even greater return than if the styles are concentrated within either type of instrument, as depicted in Figure 2.1 Furthermore, they find that correlation between the multi-style composite and traditional indices is low e.g. Barclays Aggregate, S&P500.

Figure 2.1: Style in credit



In terms of application to bond fund performance, Israel & Richardson (2015) first use a cross-sectional analysis to determine what risk factors are driving corporate risk-adjusted returns. They find that carry, defensive, momentum, and value, can explain almost 15% of the variation of such. An important finding in relation to multi-asset investing is that the positive risk-adjusted returns to these characteristics are diversifying with respect to both market risk premia and equity characteristic returns. The authors use these characteristics to conduct an attribution analysis of both mutual and hedge funds. The results indicate that the mutual funds give the positive exposure to credit markets, as would be expected, and to momentum, carry and defensive. This sample of mutual funds does not give exposure to value. Credit hedge funds, however, provide a mirror image of these exposures. The paper makes three main contributions to the literature: first they explore the role of characteristics to explain the cross-section of returns that are unique to credit; show that the return predictability is economically meaningful; conduct a detailed analysis of the time-series and cross-sectional determinants of actively managed credit hedge and mutual funds. An important finding from the analysis of the funds' performance is that individual actively managed credit funds are under-exposed to characteristics that generate meaningfully positive risk-adjusted returns. The overall findings from their empirical analysis show that mutual fund active returns load on carry, which is the least well compensated of the characteristics examined. However, it is the easiest to implement, which may explain its prevalence. The high exposure to this characteristic adds, potentially undesirable, greater implicit market risk to investors' portfolios.

The bond specific factors as outlined above are highlighted as an interesting are for future research with regards to evaluating the skills of active bond fund managers. There has been a growing interest in credit research, as illustrated by Figure 2.2.

Figure 2.2: Credit factor research

In terms of the models used for performance evaluation of bond funds relative to equity funds, this is where a key difference lies. The factors used in the Fama and French

(1993, 2015) and Carhart (1997) models for example, are publicly available for researchers to use at will. This is not the case in the bond fund literature, as factor development of this kind is very much in the early stages and there is no consensus definition of the bond market portfolio. Studies have attempted to apply the size, value, profitability, and investment factors of Fama & French, (2015) to both the US and European Corporate bond markets to determine the explanatory power Bektić et al. 2016 They find that these factors show both economic and statistical significance in the US High Yield market, however, evidence is mixed for the Investment Grade markets in both the US and Europe. High Yield bonds have been known to have similar characteristics to equities. For example, Domian and Reichenstein 2008 find that they behave approximately like 65% in equities, 30% IG bonds, and 5% cash. Furthermore, they exhibit a small-cap tilt. As such, it is likely that the factors as constructed for equities will bear significance in the High Yield market. However, the evidence is otherwise supportive of bond-specific factor construction. (Houweling & Zundert (2017)) Thus, it highlights the notable scope in the bond fund space for the testing and application of a variety of models.

Overall, this section has linked the risk factors to which bonds are exposed to the basis of model construction for performance evaluation. A key point here is with regards to bond funds, the factors are most often defined as indices to represent the exposure to term and credit and are likely to vary more significantly between studies. Equities on the other hand, are evaluated often on a style-premia basis and the factors are more consistent here. Nonetheless, the findings from recent research suggest style-premia in the bond space as an interesting area for continued research.

2.3 Conclusions

Having reviewed significant industry developments, predominantly in recent years, it is evident that there has been a response to the changing needs of investors and the impact of increased regulation and oversight. With a wider range of investment products available, such as ETFs, the competition among managers is likely to grow. The pressure on active management to deliver value has been one of the most notable areas of discussion for many years. This has been exacerbated by the rise of many low-cost passive alternatives that are easily accessible to a wide investor base.

As noted, the low-yield environment has presented challenging circumstance for achieving high returns. It may be perceived that in this case fixed income would make for an unattractive allocation, but benefits can still be gained in the wider portfolio context as bonds could make effective diversifiers. From the perspective of return generation, however, it may be required that a more creative/flexible approach is adopted to seek attractive risk-adjusted returns from credit (Houweling & Zundert, (2017) ,Brooks & Richardson (2018)).

New product development has implications for active management from various perspectives. One of which is the impact on the underlying market dynamics. For example, ETFs have been studied with respect to their influence on price discovery, valuation, and liquidity of securities (Dannheuser, (2017)) Thus, it is one among many interesting points to consider as to why performance evaluation of active funds may be impacted, and results vary relative to prior studies.

It is evident from a review of recent literature on sources of return generation in fixed income, that this is an interesting area for further study. In terms of bond fund performance evaluation, it helps to provide further insights as to the skills of the manager that may be implemented to achieve positive excess returns.

Chapter 3

Literature Review

This chapter reviews existing literature regarding the evaluation of bond mutual funds. It first presents some underlying theory and discussion of the methods used for evaluating funds. This literature review is relatively general, with the subsequent empirical chapters each reviewing the existing work that is more specific to each.

3.1 Introduction

This chapter will present a review of theories underlying the performance evaluation of actively managed funds and the existing literature in this area. It will be relatively general in nature, with each of the following three empirical chapters providing further discussion of related studies and methods that are more specific to their content.

The focus of the fund performance literature is the identification of manger skill and whether value is added for investors. There are various methods that may be used to do so. Furthermore, the definition of value is likely to vary, contingent upon the investor and their objectives. Nonetheless, value returned to investors is most commonly referred to as net alpha, as per the performance evaluation literature reviewed herein. The purpose of this study is to evaluate the extent to which positive alpha is derived, using a new sample of UK bond mutual funds. Existing literature that includes the UK does so as part of a wider European study. Thus, there is minimal evidence regarding the performance of UK bond mutual funds from an academic perspective.

Since the global financial crisis there has been a significant increase in the regulation directed towards encouraging greater transparency from actively managed funds and the value they provide investors. An area of concern relates to the magnitude of fees charged by the active managers. Some studies provide insights in this respect by using returns both before and after costs. Overall, the majority find that investors would be advised to select low-cost, passive vehicles, rather than the actively managed funds. However, given the various methodologies that may be used, in addition to advancements in terms of data availability, studies are continuing to shed further light on sources of potential value. For example, when portfolio holdings data is used, studies do identify that there is evidence of skill and potential for outperformance.

The chapter will first discuss the underlying theory and commonly used methods of performance evaluation that are considered most relevant to this thesis. Next, the literature on bond fund performance will be discussed; covering various markets, outlining alternative methodologies, and different ways by which skill and value added may be identified.

3.2 Performance evaluation methods

Given the criticisms of active management due to the apparent failure of the funds to generate consistent superior performance, it presents a debate as to the rationality of investors. Recent models, such as that of Berk and R. Green 2004 propose that continuing to allocate to funds despite minimal evidence of the value they add is not necessarily irrational behaviour. Furthermore, there is a wealth of literature directed at the decomposition of fund manager skill. For example, timing and security selection abilities, and perhaps the identification of allocation to profitable factors; aspects of active management that may be inaccessible to retail investors otherwise. Various methodologies may be used, and different definitions as to what constitutes value, have enriched the debates on the merits of active management.

Persistent alpha generation would be indicative that the Efficient Markets Hypothesis can be rejected. The contrary, however, does not necessarily hold. Failure to identify consistent superior performance does not confirm that markets are efficient.(Ang and Goetzmann 2009) Furthermore, the "value" added by active management for investors is not always necessarily in relation to outperforming a benchmark. Nonetheless, net alpha remains to be the commonly defined measure of this. As noted, much of the performance evaluation literature is focused on equity funds. A full discussion of the related performance evaluation literature is out-with the focus of this thesis. However, key theoretical contributions are important to review which draws attention to the adaptations required to apply these to fixed income funds.

Central to the assessment of fund performance is the specification of an appropriate benchmark model with which to evaluate the funds on a risk adjusted basis. Underlying many of which, and forming the basis of modern portfolio theory, is the Capital Asset Pricing Model (CAPM), developed by Sharpe 1964 and Lintner 1965. The model predicts a linear relationship, depicted by the Security Market Line (SML), between the return of the security and its beta – the beta representing the sensitivity to the market portfolio . With regards to bonds, however, an initial difference with regards to their evaluation is highlighted here – the definition of a bond market portfolio is not so well defined. Furthermore, the nonlinearities inherent in the price yield relationship can make inferences on the basis of such modelling biased. With interest rate risk the dominant factor to which bond are exposed it would perhaps be considered that a single-factor model would suffice. However, this proves not the be the case. Various multi-factor extensions will be discussed throughout this chapter.

A brief overview of the most notable multi-factor model extensions is worthwhile, as this is highlighted as an interesting area for future research with respect to bonds, and one which has been receiving more interest of late. For example, the characteristics-based premia as discussed in Chapter 2 that are common to style analysis of equity funds – value, growth, momentum, low-risk/defensive – are underpinning new product construction. Furthermore, it is not always the case that the performance evaluation methods that are grounded in equity theory can be readily applied to bonds. Certain characteristics and data limitations often determine otherwise. The following discussion will highlight these where appropriate.

3.2.1 Multi-factor models

In terms of application to bonds, the model selection is more open in some respects than is the case for equities. For example, there does not exist a definitive bond market portfolio, and the characteristics-based factors, such as value, carry, defensive/low risk, size, and momentum, appear to be in early stages of development in terms of their application to bond portfolio management and analysis. A factor pricing model for bonds, incorporates risk adjustment for both term and default risk . Fama and French, (1993) assess the extent to which there may be commonality between the risk factors on stocks and bonds, and Gebhardt et al. 2005 test various factors to identify the most efficient for explaining the cross-section of Corporate bond returns, be it beta (systematic exposure) or characteristics.

$$r_i = \alpha + \beta(TERM) + \beta(DEF) + \varepsilon \tag{3.1}$$

Here, the Term and Default factors represent "the market". Fama and French (1993) found that including a stock market proxy along with these, did not add further explanatory power. The Gibbons et al. 1989 test of mean-variance efficiency was used to judge the reliability of their various candidate models for explaining the cross-section of equities and bonds. It is often the case in the bond fund performance literature, that multiple factors are used to represent the Term and Default factors. For example, both are split into various maturity ranges, such as intermediate term and another for longer term. By doing so, it adds further explanatory power and allows for identification of the loadings on these that are contributing to the overall performance.

In their study focused on explaining the cross-section of expected corporate bond returns, Gebhardt et al., (2005) evaluate the use of corporate bond characteristics for factor construction. For example, credit ratings could proxy for default risk, or duration could proxy for term risk. Again, the GRS (1989) test was used to assess model suitability from the variety of factors they tested. They find there to be significant correlations between the beta and characteristics methods. However, the key question addressed is which has the greater explanatory power. Their results show that after controlling for duration and ratings, the default factors can explain the cross-section of bond returns. The converse also applies with the characteristics. Thus, this would indicate that the characteristics-based factors are not adding explanatory power over and above what is already captured by the systematic factors. These results suggest that ratings and duration characteristics are proxies for systematic factors; default and term respectively.

Although the linear factor model methodology as discussed thus far is underpinned by theory relating to equities, it may still be applied to the evaluation of bond portfolios and is the most common method of doing so. With a wealth of literature existing regarding performance evaluation of managed funds, interest rate risk identified as a dominant risk factor, and the relatively predictable nature of bond cash flows, the performance of such portfolios could perhaps be perceived as a relatively simple process; at least when compared to the heterogeneity that exists with equity funds. (Campbell, Lo, and MacKinley (1997)). This proves not to be the case in some respects, as the following discussion of the common methods of evaluating bond portfolios highlights.

An early examination of analysing bond portfolio returns by Dietz et al. (1980) acknowledges that there has been a growing interest in active bond management, however, techniques for performance evaluation were not well developed. Furthermore, the products have been increasing in complexity. A wide range of derivatives may be used for risk management and return enhancement, such as interest rate and sector swaps for example. They decompose the total return into three components; changes in the term structure, sector/quality selection, and the impact of the portion of return unrelated to either. The yield to maturity (YTM) is cited as being the known contribution to return. An issue here, however, is the assumption of no changes in the rate, which is unlikely to be the case. Problems highlighted when evaluating actively managed funds on a risk adjusted basis are non-linearity, call risk, and risk resulting from commonalities with stock portfolios.

Bond returns exhibit a nonlinear relationship with duration, which can impact inference of portfolio performance evaluation. A linear relationship would only hold if there were parallel shifts in the yield curve, which is not likely to be the case. (Dietz et al. 1981) conducted cross-sectional regressions using data on US Government securities from 31 December 1972 until 30 September 1977, divided into quarterly periods. The results indicate that nonlinearity is present in most of the sample. Furthermore, in the regression equation that includes the nonlinear (logarithmic) term, the coupon term is also significant for 10 of the 19 quarters. This shows that there are other factors, such as supply and demand, that influence bond prices, which are not captured within the present value of coupons already reflected in the measurement of duration.

Their identification of nonlinearity has interesting implications for portfolio performance evaluation and assessment of manager skill. The lack of explanatory power in relation to the individual bond returns, however, should not present a major cause for concern. When evaluating portfolios, it will be the case that much of the idiosyncratic risks will be diversified away anyway. On the other hand, the nonlinear relationship does have the potential to cause some issues when conducting portfolio performance evaluation. This contrasts to the linear relationship that is illustrated by the securities market line (SML), whereby expected returns as per the CAPM are plotted. Managers may have the opportunity to appear more skilled due this inherent nonlinearity. For example, if a lower risk posture than the market portfolio was assumed, a return above the line would be near enough guaranteed on this basis alone. As such, further emphasis should be applied to the decomposition of alpha generation. For example, identify if they have been able to do this through market timing ability, security selection, etc. They conclude by asserting that it is apparent the theories of bond risks require further analysis and development, to arrive at a more robust application for portfolio performance evaluation. Furthermore, the authors note that consideration within a multi-asset framework will bring additional insights.

3.2.2 Skill assessment

In terms of fund manager skill, bond funds are often considered to be market timers as opposed to security selectors. This is in relation to duration management whereby managers must anticipate interest rate movement and adjust their exposure accordingly. The CAPM-based market timing model of Treynor and Mazuy (1966), takes the form of a quadratic regression:

$$R_i = \alpha + \theta(R_m - R_f) + \gamma(R_m - R_f)^2 + \varepsilon \tag{3.2}$$

where $(R_m - R_f)$ refers to the excess return of the market portfolio, θ the corresponding exposure, and γ captures the manager's timing ability. If the coefficient θ is positive and significant, then it can be concluded that the manager exhibits either positive or potentially negative market timing ability. A notable point to consider with respect to timing ability is the extent of portfolio turnover. For example, Chen et al (2010) propose that due to the high turnover observed in bond funds, this must be indicative of active management. However, high turnover is a common feature of bond funds. This is due to the constant maturating of the underlying securities which require to be replaced to maintain the appropriate exposures, contingent upon their investment mandate. With regards to the potential for value to be added from timing ability, it is perhaps the case that the transaction costs involved will erode any of the potential gains from doing so. It has also been identified that depending upon the strategy, the options available for timing can be limited. An alternative timing model, although not so prevalent in the literature reviewed here, is that of Henriksson and Merton (1981);

$$R_i = \alpha + b(R_m - Rf) + c(R_m - R_f)D + \varepsilon \tag{3.3}$$

In this case, D represents a dummy variable; $D = 1$ when $R_m > R_f$ otherwise $D = 0$. Used as a robustness test by Huang (2015) when evaluating the timing ability of government bond fund managers. Their analysis uses portfolio-holdings data, thus providing a more granular perspective as to the allocation decisions of managers.

When evaluating funds using either timing model, the alpha is a biased measure of outperformance. This is due to the incorporation of the squared term as an explanatory variable. In this case, adjustments need to be made when interpreting alpha. Furthermore, the definition of a "market" portfolio remains open to greater subjectivity with regards to fixed income performance evaluation than is perhaps the case with equities. In this respect, the funds may be judged to time an alternative benchmark, such as a style-specific model as according Morningstar category definitions. For example, Corporate, High Yield, or perhaps Government. (Y. Chen et al. 2010) However, there remains ambiguity as to what exactly the managers are timing, and the style-specific models remain open to greater subjectivity which is likely to inhibit comparison across similar studies.

In addition to the definition of an appropriate benchmark model, the non-linearities inherent in the price-yield relationship can also present issues when attempting to infer skill. Sources of non-linearity that are not related to timing ability include stale pricing, interim trading, nonlinear relation to market factors, and public information. These characteristics are addressed by Y. Chen et al. (2010) in their study on the timing ability of corporate bond funds; the results of which will be discussed further in the following literature review.

3.2.3 Conditional performance review

An issue that has been highlighted with various fund performance applications, is that dynamic trading is not appropriately accounted for when deriving alpha. Ferson and Schadt (1996) examine the impact of conditioning information on a sample of US equity funds. They find that the overall performance of the funds they analyse looks better, when conditioning information has been included. Conditioning information is incorporated into the linear factor model as per;

$$\beta_p(Z_{t-1}) = \beta_{0p} + \beta_p^{'} z_{t-1} \tag{3.4}$$

where, $z_{t-1} = Z_{t-1} - E(Z)$ reflects the deviation of the conditioning variables from the average, as denoted $E(Z)$. The impact of the conditioning variables is captured by $\beta_p^{'}$ with $\beta_0 p$ representing the average beta. Incorporating these factors into the regression model used to derive Jensen's alpha, gives:

$$r_p, t = \alpha_p + \beta_{0p} r_{mt} + \beta_p(z_{t-1}, r_{m,t}) + \varepsilon_{p,t} \tag{3.5}$$

This model therefore now allows for the conditional alpha to be estimated. Although this adaptation allows for time-varying betas, the assumption of constant alpha remains. However, a further adjustment can be made to allow for such dynamics of alpha to also be accommodated. Again, the time-variation is assumed to be a linear function of the conditioning variables, as denoted Z_{t-1}:

$$\alpha_p(Z_{t-1}) = \alpha_{0p} + A_p^{'} z_{t-1} \tag{3.6}$$

As such, with incorporation of both the time-varying alphas and betas, the fully conditional model is represented by:

$$r_p, t = \alpha_{0p} + A_p^{'} z_{t-1} + \beta_{0p} r_{mt} + \beta_p^{'}(z_{t-1}, r_{m,t}) + \varepsilon_{p,t} \tag{3.7}$$

The null hypothesis proposes that performance is constant. Therefore, in this case the coefficient $A_p^{'}$ should be zero. Otherwise, the average alpha is captured by the intercept term, α_{0p}.

In terms of the conditioning variables that have been identified in the bond fund performance evaluation, commonly tested for their predictive power are the term spread, inverse relative wealth, and the January effect. (Ilmanen 1995) Although the initial findings by Ferson and Schadt (1996) found favourable results for equity funds when the dynamic environment was considered, this does not appear to apply to the same extent in fixed income studies. Here, the performance discrepancy between unconditional and conditional models is not uniform. For example, Kryzanowski and Ayadi (2005) find mixed performance in the Canadian market, with the conditional models having both positive and negative effects. Furthermore, various studies suggest that the predictive content of the conditioning variables is limited. Ilmanen (2011) notes the poor relation between spreads and expected return. Also suggested in Robeco factor investing paper the poor predictive content and Ferson (2006) suggests so too.

An alternative approach to incorporate conditioning information is to use the dummy variable method. By doing so can condition on various economic states by assigning dummy variables accordingly (Ferson and Qian 2010) A benefit of this approach, rather than using instrumental variables, is the assumption of a linearity may be avoided. However, for this method to yield results that can be interpreted in terms of the potential for optimal allocation, further information with regards to the definition of the economic states would be useful. For example, how persistent the economic state is and how the performance of fixed income relates to other asset classes during such times.

Another required consideration with regards to conditioning information relates to the significance of the instrumental variables over time. This needs to be tested for, as it is not necessarily consistent between markets and periods. In this thesis, consistency was maintained by using the same model in each period to allow for relative comparison. A conditional evaluation is left for future work.

3.2.4 Term structure models

As discussed in the previous chapter, interest rate risk is the main factor to which bonds are exposed. An alternative method for performance evaluation of bond funds are the use of term structure models. The development of affine term structure models has received significant attention within academic literature, however, their application in the fund performance evaluation literature is limited. To the best of my knowledge, the key papers that make use of such models for this purpose are those of Ferson et al (2006).

In order to apply term structure models for this purpose, estimation using GMM in the stochastic discount factor framework is required. Two models that are often classified as equilibrium models are those of Vasicek (1977) and Cox et al. (1985) (CIR(1985)). Equilibrium models specify the behaviour of the term structure of interest rates. This is contingent upon the state of the economy as defined by assumptions regarding the dynamics of the underlying state variables. Specified explicitly in these models are the market prices of risk.

With variation of the level factor dominating the changes observed across the yield curve, the instantaneous short rate, $r(t)$, is the factor for most one-factor models.

The Vasicek (1977) single-factor affine model takes the form:

$$dr(t) = \kappa[\hat{r} - r(t)]dt + \sigma dW(t) \tag{3.8}$$

The Vasicek (1977) model is characterised by mean reversion of the short-term rates, under the assumption that the economy tends towards equilibrium. Mean reversion implies that the short rates converge towards a long-term target. In this sense, the impact of factors known to change the level of the short rate dissipate over the long-term. A low mean-reversion parameter means persistent effects from any shock to the short rate. The converse also applies. This assumption of mean reversion is a strong one to make. It may perhaps be reasonable given stable macro-economic conditions, however, certain situations such as hyperinflation or other catastrophic ones will challenge mean reversion. One aspect of the Vasicek (1977) model that has previously been criticised is the allowance for negative interest rates. Given that major economies in recent years have cut interest rates to below zero, this aspect of the model may be considered more legitimate. Another single-factor affine term structure model is that of Cox Ingersoll and Ross (1985):

$$dr(t) = \kappa[\hat{r} - r(t)]dt + \sigma\sqrt{r}(t)dW(t) \tag{3.9}$$

Here, the volatility parameter is constant, however, annualized basis point volatility equals $\sigma\sqrt{r}(t)$, increasing with the level of the short rate.

A differentiating feature between the Vasicek (1977) single factor and CIR (1985) is that although both specify mean reversion of the short rate, the latter does not impose constant variance. Instead, the variance of the short-rate according to CIR (1985) changes in proportion to the level of the short rate.

Although the single-factor models are relatively simple, they are not without their limitations. Both the Vasicek (1977) and CIR (1985) models specify a relatively simple volatility structure. However, this is not the conclusion as presented by empirical evidence. With regards to their application for derivative pricing, this is a key point to consider that renders them ineffectual; the specification of the volatility structure is a critical input for such purposes. Thus, even with the dominance of the level factor, further extensions are still required and multi-factor models often favoured. Brennan and Schwartz (1979) – two-factor not affine:

$$dr(t) = \beta_1(r, l, t)dt + \eta_1(r, l, t)dW_1(t) \tag{3.10}$$

$$dr(t) = \beta_2(r, l, t)dt + \eta_2(r, l, t)dW_2(t) \tag{3.11}$$

where, $\beta_1(r, l, t)dt$ and $\beta_2(r, l, t)dt$ represent the drift terms of the short and long rate, and likewise $\eta_1(r, l, t)dW_1(t)$ and $\eta_2(r, l, t)dW_2(t)$ the volatility terms. In this two-factor model, both the effects of the level and the slope are accommodated; represented by the dynamics of the short rate r(t), and the long-term yield, l(t) : The slope measure is estimated by the difference between the long and short rates.

With regards to applicability in practice, it is more likely to be the case that arbitrage models are used, rather than equilibrium models for trading purposes. The calibration of arbitrage models results in the model price of the underlying security being consistent with the market price.

In terms of deriving estimates of factor risk premiums, the term structure model approach for fund performance evaluation is perhaps more ambiguous. However, there are some benefits in terms of their use for evaluating fixed income funds. Continuous time nature of the interest rate process allows for further explanatory factors to be derived in the model. The method is also accommodating of nonlinearities as it avoids the assumption of the linear functional form as is the case when using beta factor models.

As noted, the use of term structure models within the performance evaluation literature on bond funds is to the best of my knowledge, limited. The data used in the Ferson (2006) papers covers from 1985 until 2000. Therefore, the application of these models during times in which the interest rate environment has been significantly different, such as historically low rates, would make for interesting further research and a basis of comparison with the more commonly used methodologies. Additionally, their use in another market aside from the US would allow for insights to be gained from their use in fund performance evaluation using a new dataset.

3.3 Performance evaluation of bond funds

The central contribution of this thesis is to provide an evaluation of UK fixed income mutual funds. No study to date provides such an analysis of this market; most of the literature reviewed is therefore focused on the US and international examples. Much of the literature begins with a statement regarding the lack of studies relative to equity fund analysis. For example, it is noted by Ang and Goetzmann (2009) when they conduct their review of the active management of the Norwegian pension fund that the lack of academic research on fixed income funds is unfortunate, especially given that this fund at the time of their analysis, was 40% invested in fixed income. Furthermore, the growth in the size of the fixed income markets is also often cited.

An early study by Cornell and K. Green (1991) investigates the performance of US low-grade bond funds over the period 1960-1989. The authors seek to identify if low-grade bonds offer greater risk-adjusted returns than their more highly rated counterparts. Equally-weighted portfolios are constructed of both types of funds. Using a simple two-factor model they regress the returns of the bond funds on Treasury bond returns and stock market returns. The mean return on a portfolio of low-grade mutual bond funds is higher than the mean of a high-grade corporate bond index and the mean return on a Treasury bond index – consistent with expectations given the different the risk profiles. Low-grade bond funds are much more sensitive to changes in stock prices than high-grade bonds. Duration is likely to be shorter for the low-grade funds, which is due to characteristics such as greater coupons and/or weaker call protection. Given that duration is the measurement of sensitivity to changes in interest rates, it is for this reason that such funds have a greater sensitivity to changes in the stock market. Perhaps contrary to expectations, is the observation that the standard deviation of low-grade fund returns is lower than that of the high-grade funds. This is also due to the shorter durations. Over the sample period the standard deviation of changes in long-term interest rates was large relative to that of stock returns, as such the low-grade funds were less exposed. Findings from this study indicate that over the long-term, the returns on low-grade bond funds are approximately equal to the returns provided by an index of high-grade bonds.

Two papers which have been central to much of the bond fund performance evaluation literature that has followed are those of Blake et al. (1993) and Elton, Gruber, and Blake, (1995). In the earlier paper, they calculate the excess returns of the bond funds by simply subtracting either the risk-free rate, or the benchmark return that has been assigned by Morningstar. It is recognised, however, that this is a relatively naïve measure of alpha. A more appropriate method requires the abnormal return to be calculated on a risk-adjusted basis, which they do using OLS regressions. As such, single and multi-factor models are used to mimic the risks that are inherent in the funds' strategies. As noted, the weightings on these indices may be viewed as investable passive alternatives, allowing for identification of whether the manger is able to outperform through active management. Six unconditional models are used for the evaluation, five of which are presented in table 3.1.

A limitation of regression analysis for mutual fund evaluation, however, is that it often assigns extreme and negative weights to the factors. This is potentially problematic as a common restriction for mutual funds is that they may only hold long positions. A popular method in the literature to overcome this or add a robustness test, is to use the quadratic programming technique of Sharpe (1992), which does not allow for negative weights. EGB (1993) apply this as their sixth method of modelling and conclude that the styles as assigned by Morningstar are reasonable, as they are concurrent with the

Table 3.1: Model construction

This table presents the models as per Elton, Gruber, and Blake (1993). This study was among the first in academic literature to conduct a comprehensive evaluation of bond funds. The sample of funds was US-focused. A selection of indices from Lehman brother and Citigroup were used to proxy for the risk factors

Model Type	Risk Factors	Representative Index(s)
Single-index		
1: Market	Bond market aggregate	Lehman Brothers Aggregate
2: Own	Style-specific	Morningstar style-specific
Multi-index		
3: Risk 3	Credit	Aggregate High Yield Mortgage Backed Securities (MBS)
4: Maturity 3	Maturity and credit	Government 1-10 Government 10+ High Yield
5: Reg 6	Maturity and credit	Government 1-10 Government 10+ Credit 1-10 Credit 10+ High Yield Mortgage Backed Securities (MBS)

allocation as derived by Sharpe's method. Across the variety of models, the findings conclude that the performance of bond funds is not as sensitive to the specification of the benchmark index as is the case with equity fund performance, observing that the (adj)R^2 terms and the sign of alphas are similar, regardless if single or multi-index models are used. A point to note, however, is that the significance of the performance varies between the different samples, highlighting the sensitivity to the selected time-period. Only four of the six models are used for sample 2; Maturity 3 and the Sharpe (1992) method are excluded. Overall, the bond funds are found to underperform by a magnitude of costs. The authors note "striking" results from sample 2, as it is in this case whereby the significant negative results are observed.

Their latter paper Edwin J. Elton et al. (1995) take a two-step approach to fund evaluation. First, they use Likelihood Ratio tests to determine the validity of the models, before then using them for the performance evaluation of 123 US bond mutual funds. The factor combinations in the four models they test are as follows:

1. Index1: Aggregate bond index
2. Index 4: Aggregate bond index + stock market index + default effect + optionality effect
3. Fundamental 4: Aggregate bond index + aggregate stock index + unexpected changes in inflation + unexpected changes in real GNP
4. Fundamental 4: All the influences from models 2 and 3 as above

Their sample contains portfolios of bonds beyond just Treasuries, which they note has been the focus of prior studies in this area. Corporate and mortgage securities are included also. They conclude that indices are required for explaining the time-series of returns, however, the fundamental factors are required for explaining the cross-section of expected returns. They use Likelihood Ratio tests, evaluating restricted and unrestricted versions of the models. the validity of a model is determined by its ability to price the test assets, as defined by a variety of passive indices. A point to note is the lack of available data with which to test models applicable for High Yield funds. Therefore, they have excluded them from the sample.

A performance evaluation is conducted; however, this differs in a number of ways from that in the authors' 1993 study. As discussed, the earlier paper compares performance on a risk adjusted basis in terms of comparable passive portfolios. In this case, they make use instead of their relative pricing models they have parameterised and tested. As such, the prices of risk have already been estimated and thus used to obtain expected returns for the mutual funds. By doing so, in addition to evaluating fund performance, this approach provides an out-of-sample test of the models whereby it will be evident from the reasonableness of the return predictions whether they have been correctly parameterised. They examine the sensitivities of the categories to changes in the variables in each of the models. Inflation is a significant risk factor for bonds –

rising inflation corresponds to rising interest rates, which would have a negative impact on bond prices. Therefore, a negative relationship is found between the performance and factor representing unexpected changes in inflation. The relationship with the GNP factor, however, is slightly more ambiguous. An unexpected rise in GNP would indicate improving economic conditions, which is likely to benefit corporate bonds as the credit of the issuing company strengthens as a result. However, with this usually comes a rise in interest rates, which would hurt the returns of the other funds. To determine overall performance, each category of funds is evaluated with each of the four models. There is no evidence of superior performance. When the evaluation is conducted across all the funds, the result is negative with each of the models. In this case, the most negative result is obtained when using Fundamental 6 – to the magnitude of -0.1158% per month, significant at 1%. The Fundamental 4 model gives the least negative result of -0.0647%, although this is not significant. To further their analysis, the regressions are conducted using the gross excess fund returns, rather than net. As such, this would allow for an indication as to whether the managers were able to provide sufficient superior performance to at least cover costs. Even on this basis, however, the findings are disappointing. The alphas are range from -3.5bps to +1bps as a monthly average, which is hardly compelling evidence that these managers provide significant outperformance, even on a before-cost basis. To conclude, they recommended that investors seek low cost funds.

The classification of bond funds into categories, for example as defined by Morningstar is relatively broad. Therefore, within each there will remain a degree of heterogeneity in terms of style and overlap between them. Comer and Rodriguez (2013) aim to identify the extent of performance discrepancy between US High Quality Corporate, General Corporate, Government Treasury, and General Government funds, using models similar to those of Blake et al. (1993). They conclude that Corporate bond fund managers exhibit more skill and superior performance, relative to those of Government bond funds on a risk-adjusted basis. However, they still find that alphas for both Corporate and Government bond funds are negative; those of the Corporate bond funds just to a lesser extent. The worst performance is observed for the General Government category using the Maturity model, and the best is for the General Corporate using either the single-factor or Sector models. The annualized alphas are -1.17% and -0.58% respectively. T-tests show that a significant difference exists in the return premium between Corporate and Government bond funds; the former exhibiting superior performance to the degree between 8 and 53 bps annually, contingent upon the model used. Their results show that in each of the sub-periods considered, the two best performing categories are the General and High Quality Corporate funds. However, the period of 2007-2009 is an exception, whereby this is not the case. The results are not reported, therefore it is not clear to what extent, or in terms of category rankings, this differs. Contrary to what was found by other studies , they note that the differences are not only a function of the expenses. A comparison of the abnormal performance observed with the Total Expense Ratios (TERs) shows that some of the funds underperform by a lesser extent than the percentage of fees. Having identified the risk-adjusted performance, the authors then use these results to try to identify what is driving the flows of funds. It appears that investors are chasing returns; the highest performing decile of Corporate bond funds appear to attract the highest volume.

The papers discussed thus far adopt the beta representation for performance evaluation, in which the models are linear in form. It is a useful way by which to assess the value of active management in relation to an “otherwise equivalent” (Aragon and Ferson 2006) passive portfolio. However, there are feature of bonds and their returns that have encouraged alternative methodologies to be considered for this purpose. In addition to the issue of nonlinearities, the interim trading bias that results when fund managers trade within the return measurement window, may be exacerbated. This is due to bond funds being known to have a higher turnover than, for example, equity funds.

An alternative approach to fund performance evaluation, initially proposed by Z. Chen and Knez (1996), is the stochastic discount factor (SDF) framework. Ferson, Henry, and Kisgen (2006a) derive continuous time term structure models using this method. Their study focuses on US Government bond mutual funds from 1986 to 2000, which are aggregated into an equally-weighted portfolio. Given their sample is comprised of default-free securities, interest rate risk is the primary factor to consider; avoiding the need to incorporate further credit risk factors into the modelling. They present various reduced form single and multi-factor models as discussed in section 3.2.4, the most notable from the literature are being those of Vasicek (1977), Cox et al. (1985), and Brennan and Schwartz (1979).

Conditional performance evaluation considers the influence of time varying macroeconomic factors on performance and risk; as such they find it to be a particularly appropriate method to use as the representative benchmarks they select are found to vary in terms of returns and volatility during the different term-structure states in their sample period 1986-2000. The use of term structure models provides guidance in this respect, as they inherently indicate the selection of conditioning variables – various slopes of the term structure. As such, dummy variables are used to define a flat, upward, or inverted curve. Additionally, by using these models, the continuous time application overcomes the interim trading bias. This has been noted to be potentially be problematic when inferring managers' skill. For example, if not accounted for, it may appear that managers are acting on private information when they have just acted on publicly available information that was released within the measurement interval. The use of this methodology to overcome interim trading bias has an advantage in that it avoids the need for portfolio holdings, which is an alternative approach but suffers from data availability issues. Furthermore, the explanatory power of the reduced form models is increased as additional empirical factors are derived from the continuous time estimation. For example, the single factor model of CIR (1985) in this case has two factors due to the time aggregation of the underlying state variables.

Given the results of their model testing, which finds that the three-factor affine model is the most reliable, they focus on the results from this one when conducting the performance evaluation. They conclude that on an after-cost basis the return of the Government bond funds is overall negative; the unconditional alpha is -0.08 per month (t-stat = -4), and the conditional alphas range from -0.03% to -0.09% per month, with four of the six conditional alphas being statistically significant. In terms of economic significance, the alphas are small. However, the accuracy with which they are estimated is high, in part due to the relatively low volatility of the funds' returns. The extent of abnormal performance varies more in relation to economic states, as opposed to when the funds are grouped according to characteristics. The results are similar between models when the time averaged factors are excluded. They assess the significance of expenses, by regressing the conditional alphas on total costs; including both the total expense ratio (TER) and an estimate of trading costs based on turnover. Costs are estimated to be between 8bps to 12bps, and the regressions indicate a weak negative relationship between alphas and total costs.

Further to their research on Government bond funds, in a separate paper the authors also assess the extent of cross-sectional variation and any style dependent performance dispersion in a sample of US bond funds across economic states (Ferson et al., 2006b). The challenge presented here was to derive a stochastic discount factor that incorporates both the term structure, and account for the additional risk factors required for the non-default free bonds. Their sample contains Government, High Yield, Corporate, High-quality Corporate, and Mortgage bond funds. As such, default risk, mortgage prepayment risk, and potentially others, must be added to the models. Over their sample period 1985-1999, performance is found to be less than passive benchmarks on an after-cost basis for most of time periods considered. However, this varies with such funds appearing to offer positive excess return in times in times of high credit spreads, a factor synonymous with increased levels of volatility.

They find that the factors associated with poor bond fund performance are high short-term interest rates, steep term structure of interest rates, and high industrial capacity utilization. Fund style is identified to present a greater heterogeneity of performance than does the cross-sectional variation of fund characteristics. Of the categories included in their sample, the Mortgage funds perform the worst; underperforming their style-specific passive benchmark in all economic states. These excess returns as just described are based on the unconditional performance and also that which has been conditioned on various economic states, but prior to risk-adjustment using the SDF models. When the various risk models are applied, the performance observed is closer to neutral; the excess returns are very small, many of which are insignificant. Having adjusted for risk, the style groups show an absolute alpha of 13.2bps across the various states, which is notably less than the 22.4bps observed prior to risk-adjustment.

3.3.1 International evidence

It is well recognized in asset management that what goes on in the US markets has a knock-on effect on the rest of the world. As such, it is likely there is a degree of commonality between the results observed between countries. An international perspective is considered by Ilmanen (1995) in his analysis of time-varying expected returns across bond markets. The sample includes Government bond funds from US, UK, Canada, Japan, Germany, and France. These are the most liquid bond markets, and therefore issues with stale pricing should be minimized. Furthermore, as the sample consists only of Government bond funds, interest rate risk is the dominant factor. The inputs into the predictive regressions are used as conditioning variables in other subsequent studies; term spread, inverse relative wealth (IRW), January effect. (Silva et al., (2003)). The paper concludes that single-factor models are too simple and insufficient to explain bond return predictability, noting that theoretical justification is lacking to support interest rate risk being the only risk factor. Ilmanen (1995) Finds that there is a high correlation between the expected excess returns across international bond markets; more so than is the case between world stock and bond.

There has not as yet, to the best of my knowledge, been a study conducted that focuses specifically on the performance of UK bond funds. (D. Blake et al. 1998) assess 2,300 UK mutual funds from 1972-1995. The sample is comprised of funds that may have an investment focus of domestic equities, international equities, bonds, commodities, or property. Many different perspectives of performance are considered, such as persistence, survivorship bias, variation across asset classes, and distance between inception and termination dates. However, due to the limited availability of benchmarks, the risk-adjusted performance evaluation does not include the UK bond funds – it is limited to equity and balanced funds. They do have sufficient data to calculate the survivorship bias of the UK Gilt and Fixed Interest Funds, however, which is found to be 0.05% per month.

Although the volume of studies is limited by comparison, international evidence also indicates similar results to those in the US. Furthermore, due to relatively small sample sizes, the conviction of the inferences drawn at times may be questionable. Detzler (1999) conducted a performance evaluation of 19 global mutual funds over the period 1988-1995. The paper concludes that such funds do not exhibit outperformance, net of expenses. The conclusions also indicate that diversification benefits from including non-US bonds in a portfolio are minimal, and exchange rates are also an important explanatory factor to consider. This can be related to the findings of Ilmanen (1995) in terms of the correlations across global bond markets.

Silva, Cortez, & Armada, (2003) assess the performance of European bond funds, using conditional and unconditional single and multi-index models. The sample includes 638 bond funds - 58 Italian, 266 French, 90 German, 157 Spanish, 45 UK, and 22 Portuguese. The authors note that more research on the European market is warranted and highlight that it is the largest second to the USA. Although this may be a valid claim, it does not mean that European funds may be easily grouped for a performance evaluation. There will remain discrepancies between countries in terms of regulations and market structures, data availability issues; potentially making the application of common benchmarks for analysis problematic. For example, Ilmanen (1995) notes that there are different regulations in France relative to the other markets and as such the results vary. Additionally, the UK has always sought to retain a relatively separate identity and did not adopt the single currency.

Before the information variables are incorporated into the conditional models for performance evaluation, they are first assessed using regressions to see if they are indeed useful predictors in this sample of bonds. The unconditional models take the form of those as per Blake et al. (1993), in which beta is assumed to be constant. For evaluation of bond funds, this may be more of a problem than is the case for equities. As the risk of the bond is measured by duration, this changes over the time-series as a function of maturity, which inevitably leads to beta instability. Also, with interest rates as the dominant risk factor, fund managers must constantly adjust duration to manage this risk. Therefore, conditional models may be more appropriate, to allow for time-varying alpha and beta. The term spread is found to be highly significant but negative, which is contrary to the positive relation observed in the US market. Inverse Relative Wealth is most significant for the UK market, as identified across all models; not quite so for the other markets as only evident when using the 3-5 year maturity model. The UK stands out again in relation to the January effect, as this is the only market in which significance is not identified.

Overall, the paper concludes that bond funds exhibit negative performance - to the greatest extent in Italy, Spain, and Portugal, and UK Gilt funds. Neutral performance, however, cannot be rejected for most of German, UK Corporate, and "Other Bond" funds. Using the conditional model, there is evidence that betas and alphas may both be time-varying. From a comparison of their unconditional and conditional models, the study finds there to be a greater impact on performance assessment from the inclusion of additional risk factors, more so than for incorporation of predefined information variables.

A recent study by Grose et al. (2014) assesses the performance of European bond funds, primarily invested in Portugal, Ireland, Italy, Greece, and Spain. Such countries are known as the PIIGS. They evaluate the persistence of the funds' performance over the short, medium, and long-term. Various metrics are used to do so, which allows for a consistency check between them; the conditional CAPM, Sharpe ratio, and modified Sharpe ratio of Ferruz Agudo and Sarto Marzal (2004). The conditioning variables they use are the yield of a three-month note, difference between the yield of a ten-year and three-month note, the dividend yield spread, difference between the yields of BBB and AAA corporate bonds, and a January dummy to detect if any turn of the year effect. The persistence is measured over 6, 12, 24, and 36-month periods; measured relative to the median performance of the whole sample for each market. Contingency tables are used to categorize winners and losers in six-monthly periods. A Cross Product Ratio is then used to estimate the odds of recurring performance. They find evidence of short and medium-term persistence, but weaker over the longer horizon. The unconditional alphas are all negative, and significance is observed for all except for Greece. The same applies for the conditional model, although the magnitude of the negative results is to a slightly lesser degree. In terms of the contribution of this paper, they note that fund-of-funds and pension funds seek the best performing mutual funds for return enhancement. Given the long history of academic literature highlighting that active management fails to add much value to investors, this is debatable. Additionally, they note the diversification benefits from investing in government and corporate bonds. However, they haven't addressed the extent to which this may be the case, if any.

Aside from Europe, the Canadian market has been identified as of interest for further study, due to various distinctive characteristics. For example, the authors cite features such greater use of trailer fees by institutions, varying attrition rates, and lower levels of leverage employed by bank fund sponsors. Another study adopting the stochastic discount factor approach is that of (Kryzanowski & Ayadi, 2005) their performance evaluation of Canadian fixed-income funds. Both conditional and unconditional models are applied. The performance of Canadian fixed-income funds is found overall to be negative; consistent with most results from other countries.

3.3.2 Timing ability

In terms of fund manager skill, bond funds are known to be market timers as opposed to security selectors. This is in relation to duration management whereby managers must anticipate interest rate movement and adjust their exposure accordingly. As outlined in Section 3.2.2, there are two commonly applied measures of timing ability adopted within the academic literature; the Treynor Mazuy (1966) and Henriksson and Merton (1981) models. Studies have applied these to both returns and holdings-based data, with mixed results.

Boney et al. (2009) focus on the timing ability of High-Quality bond funds over the period 1994-2003 to determine the ability of managers in terms of their asset rotation skills. They find disappointing results from the perspective of a positive alpha; the managers tend to allocate to cash when bonds of all maturities outperform risk-free assets. Nonetheless, the authors note that a source of value may still be perceived by investors in terms of the potential for diversification benefits.

As noted, non-linearity is evident in the relationship between bond returns and risk. There are, however, various characteristics that may exacerbate this. Consequently, Chen et al.(2010) account for nonlinearities both related and unrelated to timing ability. Four potential sources of nonlinearity unrelated to timing are identified and adjusted for: 1) nonlinear relation between economic factors and benchmark portfolio, 2) interim trading, 3) stale pricing, and 4) publicly observed conditioning variables. They include such biases in addition to convexity, which is identified as the source of nonlinearity related to timing ability. They do not find the timing ability of bond fund managers to be a particularly significant influence on performance, finding neutral effects having grouped funds by style, and slightly positive performance on a cross-sectional basis. Consistent with much of the research on bond funds they find that on an after-cost basis the performance is significantly negative. Reference is also made to market timing ability on a more investor-specific basis; the ability of a manager to time market conditions and adjust to increased volatility could provide higher utility for more risk-averse investors, despite still observing a negative after-cost alpha. The authors acknowledge that forming such conclusions regarding skill on an aggregate basis may mask the ability of some individual managers. Thus, they evaluate the performance on a before-costs basis. From this, it is evident that approximately 75% of the funds do have the ability to generate positive alpha, although not all significant.

Although the returns-based approach dominates, with data availability and quality improving, the portfolio holdings method of analysis has gained popularity. Assessment of timing ability can focus on different aspects of performance; be it market level, volatility, or sector rotation. Comer (2005) looks at bond fund sector timing skills using a sample of US Government bond funds; a popular choice among such studies due to the higher levels of liquidity in their markets. The more traditional returns-based model of Treynor Mazuy (1966) is criticized for assessment of timing ability, finding that it requires managers to have significant levels of skill for the model to be able to detect positive timing ability. When the holdings-based approach is used a more positive result is obtained; managers are found to have neutral skill at worst, and in some cases positive timing ability.

Another early paper to adopt the holdings-based approach was that of Cici et al. (2012), focusing on US Corporate bond funds over the period 1995-2006 using quarterly snapshots of holdings. By combining six sources, all of which are survivorship-bias free, they construct a very comprehensive dataset. Having applied specific screens to determine inclusion, a final sample of 746 Corporate bond funds results; 537 Investment-grade and 209 are High Yield. The overall results indicate that the costs of active management outweigh the benefits. There is minimal evidence of stock selection ability; as the performance is neutral or weakly positive.

Huang and Wang (2014) study the timing ability of Government bond fund managers using monthly or quarterly holdings of Treasury securities during 1997-2006. Given that interest rate risk dominates for this type of fund, the focus on how to add value is likely to be with regard to market timing as opposed to security selection. Due to the greater liquidity of Treasury securities, they note that frequent trading is more likely and conducive to a greater extent of market timing. As a robustness check they extent their analysis past Treasury securities to consider the corresponding ability of more general government bond funds. The latter have the ability to invest in a wider range of assets, including mortgage and asset-backed securities. The results confirm that Treasury managers have a greater degree of positive timing ability; however, a conditional performance evaluation confirms that this is limited to their response to public information. As such, they conclude that they are better able to interpret the information in the economic news announcements. Although using holdings-based data gives a more detailed analysis, it has its limitations still in terms of availability. The authors apply relevant screens to their data, and after doing so results in a relatively small sample of 146 unique funds; small relative to many US studies that use a returns-based approach.

From the literature reviewed thus far, it is evident that active bond managers struggle to perform by a magnitude sufficient to cover costs. Moneta (2015) however, finds some indications that there may in fact be some evidence of superior performance. Using portfolio holdings, he finds evidence of the benefits of active management; on a gross basis, managers earn a positive alpha of 1% p.a. Sector timing ability proves to be neutral, however there appears to be some market timing ability exhibited by a sub-group. Bond funds are likely to exhibit significant time-variation in returns and characteristics, due to time-variation in duration and credit quality of the holdings. Using the holding-based measures is a way by which to overcome this, however, the availability of the data may also have some limitations. The conclusions of this paper propose that investors should look to the "Return Gap" (RG) measure, as opposed to just Jensen's alpha when selecting bond funds, finding this to be a more "precise" measure to indicate likelihood of performance persistence. Those investing based on a favorable past RG would expect to realize an excess of 2.3% p.a. The greatest extent of abnormal performance is found in the high yield debt market, perhaps due to more inefficiency in prices. Furthermore, diseconomies of scale are identified and evidence of a negative impact of past flows on performance. The results overall are not as negative as was observed Blake et al. (1993), however, the alphas identified by Moneta (2015) are not significant.

3.3.3 Persistence

The ability of active managers to consistently provide superior performance is coming under increased scrutiny; particularly given the expanding set of low-cost passive alternative investment vehicles that are now available for retail investors, such as ETFs and index-linked mutual funds. In the case of efficient markets, as discussed, such performance should not be possible.

The results of a study by Huij and Derwall (2008) indicate that US bond funds do exhibit performance persistence; concluding that managerial skill is a determinant of the returns observed from High-Quality and High-Yield bond funds. They find that investors can earn returns in excess of benchmarks when selecting on the basis of past fund returns, potentially by a magnitude of 3.5% p.a, which is the risk-adjusted return difference between the top and bottom deciles of funds ranked on prior performance. Furthermore, Gutierrez Jr et al. (2009) analyze performance persistence in a sample of US corporate bond funds. They find that their sample exhibits positive alpha on a before-cost basis for the subsequent four years, however, when expenses are included they underperform the benchmarks. In relation to the findings of Berk and Green (2004) whereby the larger funds suffered from diseconomies of scale, the results of Gutierez, Maxwell, and Xu (2009), do not find support of this. They conclude that there is no significant relation between performance and lagged fund size. Instead, they conclude that performance persistence observed is in fact due to managers' skill and is not just down to luck. In persistence studies there remains to be a degree of ambiguity as to what exactly defines persistence; over what length of time-period the manager needs to consistently perform is debatable between studies.

3.4 Conclusions

Although performance evaluation has received much attention in the academic literature, it remains to be a rich field for future research with respect to bonds. It is interesting to identify the extent of differences, if any, in terms of the alphas derived when various methods are used. The approach of Jensen's alpha (1968) remains to be the most common in the academic literature when determining the value added by active management. The alternative use of term structure models in the stochastic discount framework is far less common. (Ferson et al., (2006a,b))

Overall, the performance evaluation literature finds little support for active management in terms of outperformance relative to a benchmark model. This is most commonly evident when net returns are used with much of the literature concluding that funds underperform by a magnitude of costs. The continued rise in popularity of exchange traded funds (ETFs) and other low-cost vehicles, has added to the pressure for active management in this respect. Industry regulators have also been adding further focus. Hedge funds were traditionally the source of such scrutiny; however, mutual funds now also fall very much within scope. It is apparent that fund performance research regarding UK samples of funds is minimal.

There are studies, however, in which the results do indeed appear to be more favourable. For example, the use of holdings data allows for more granularity with regards to the actions of managers (Moneta (2015)), Huang (2014)), identifying that managers do in fact have potential to add value for investors. With data quality improving and disclosure requirements becoming more demanding with a view to improve transparency, it may be the case that further work will be possible using holdings data. If so, greater insights as to the selection ability of the managers will be possible.

Chapter 4

Model testing

Having reviewed the literature regarding the performance evaluation of bond funds, this chapter proposes and tests five candidate models that may be used for evaluating the sample of UK bond mutual funds in this thesis. The testing methodology is discussed, the data described, models specified, and the empirical results are presented.

4.1 Introduction

The specification of an appropriate benchmark model is central to evaluating the merits of active management with minimal bias. An interesting perspective with regards to the importance of model testing for bond fund performance evaluation is the observation that active management of this asset class has been found to outperform passive benchmarks over the past 20 years. (Frieda and Richardson 2016) The debate, however, as to whether this is in fact due to managerial skill is still very much open. One reason why it is difficult to ascertain the source of such outperformance is the diversity and scale of the fixed income markets. It is often the case that benchmarks will leave much of the investment opportunity set uncovered, therefore providing notable scope for managers to capitalise on return derived from off-benchmark positions and create "alpha". By conducting model testing prior to an evaluation of these actively managed funds, it will hopefully help to assess the prevalence of model biases and arrive at a more robust conclusion regarding outperformance in the empirical chapters to follow.

Although commonalities exist between the risk factors to which equities and bonds are exposed, specific models for each asset class are required to optimally capture the cross-sectional variation in returns. In a seminal paper, Fama and French (1993) seek to identify the extent to which a single model can indeed explain the returns of both stocks and bonds. Overall, they conclude that three equity and two bond factors are required; stock market excess return, size, value, term, and default respectively. Consistent with this, Ferson et al. (2006) notes that it is empirically difficult to find a model to price both stocks and bonds.

In the bond fund performance evaluation literature, evidence of model testing is limited. Many studies refer to (adj)R^2 when deducing which model is most appropriate and making comparisons. Alternatively, the Schwartz Information Criterion (SIC) statistic has been used (Ayadi and Kryzanowski (2011)). Perhaps is the case that these statistics give an overall test of the model fit, while penalising for the number of parameters; however, they don't provide further insights regarding what type of portfolios may present pricing problems for the models. This is an attraction of using the GRS (1989) test methodology. The corresponding alpha statistics allow for identification of the sources of potential biases in the evaluation of managed funds.

Focusing on interest rate risk as the dominant risk factor to which bonds are exposed, Ferson et al (2006a) derive term structure models using the stochastic discount factor (SDF) framework. Prior to evaluating the performance of US government bond funds, the authors test the ability of candidate models in pricing the returns of passive benchmarks and dynamic bond portfolio strategies. The models perform well, with the results indicating that a large proportion of the conditional expected returns on the set of test assets can be explained by the SDF models. The average absolute mean return before risk adjustment is 80bps per month, with the average absolute alpha after risk-adjustment being 8bps or less. As such, the pricing error on the test asset portfolio indicates that around 90 percent of the expected returns are captured by the risk factors of the SDF models.

Further to their analysis of Government bond funds, they extend the sample in a second paper to include High Yield, High Quality Corporate, and Mortgage funds (Ferson et al (2006b)). Therefore, additional risk factors must be added to the term structure models previously used when the focus was only Government bond funds. Joint estimation is used to identify the parameters of the SDF models and generate alphas relative to the benchmarks. Primitive assets are required to first specify the models; in this case the 90-day Treasury bill, and the twenty-year Treasury bond return. The benchmarks used are the returns on the one-year government bond, and the Lehman Brothers Government-Corporate index . Hansen's J-statistic rejects the models, finding them to be over-identified. However, further statistics are used to provide a basis for relative evaluation to decide upon the most reliable for the fund performance evaluation. The key metrics are the alphas and standard errors. When these are small, it indicates minimum bias and high precision, respectively. This is an observation common among model testing literature - models may appear to perform poorly in terms of satisfying conditions of the GRS or Hansen's J-statistic, however, they can in fact still be reliably applied for the purpose of evaluating active management. Considering additional perspectives, such as alphas and goodness of fit, is key to identifying the more appropriate model.

This chapter specifies and tests the candidate models that will be used throughout the subsequent empirical chapters in this thesis. The primary basis of comparison are those of Blake et al (1993), which have commonly underpinned the model construction in the bond fund performance literature. The overall aim of the chapter is to identify which of these can be considered as most reliable for performance evaluation of UK bond mutual funds. As such, the identification and significance of alpha across the test asset portfolios is the primary focus. By assessing the models' ability to explain the returns on passive portfolios (as represented by the test assets), it allows for insights as to the extent of any biases that may be induced when assessing active management.

Five models are proposed; two single-factor and three multi-factor benchmarks. There are two main results from this chapter. First, from the model performance evaluation tests a five-factor model is identified as most reliable. Second, the power of the GRS test is highlighted as a limitation when making judgements on this basis alone. For example, none of the candidate models pass this test of mean-variance efficiency during the recent sub-period from 2009 until 2016. However, the model fit is superior and alpha bias is minimal - which may appear to be counter-intuitive to such results. Overall, the five-factor model is found to be consistently reliable most appropriate to use for the fund performance evaluation in the following chapters.

The chapter proceeds as follows: First, the testing methodology is outlined. In section 2 the candidate models are specified. Descriptive statistics for the model testing data are then discussed in section 3. The empirical results are presented in section 4. The analysis is concluded in section 5.

4.2 Time-series tests of asset pricing models

The focus of this chapter is an evaluation of candidate models, seeking to identify which may be considered most reliable for the purpose of evaluating actively managed UK bond funds. The GRS test is used extensively throughout model testing literature in the equity space. Many candidate models have been proposed; so much so that the range of factors now available has been referred to as a "zoo". (C. R. Harvey and Liu 2019) A review of the extensive equity literature in this area is besides the scope of this thesis, however, specific examples from application of the methodology are worth highlighting.

A key observation in this respect is that studies most often find candidate models fail the GRS test. Various reasons may be at the root of this. However, it is not the case that such failure invalidates a model and concludes that it is "bad" for the purpose intended. In some instance, the opposite may in fact apply. For example, E. Fama and K. French (2012) conduct international tests of the their three-factor model (across North America, Europe, Japan, and Asia Pacific), augmented with momentum Carhart (1997). They do not find support for internationally integrated models – local models perform better than global, although still provide incomplete descriptions of average returns. Power issues are identified. The results highlight the impact of tight fit on the GRS test results; a characteristic to which it is sensitive. Fama and French (2012) find that when left hand side and right hand side variables are formed with respect to the same region, the models are rejected. This may appear to be counter-intuitive as they are otherwise appealing, for example, from an economic perspective.

Their five factor model is also tested on an international basis. Here, spanning tests find that the size factor is not important for describing average returns in Europe and Asia Pacific, but it is for North America. However, including it in every case results in improved precision, as it allows for more left hand-side variance to be absorbed. When it is omitted, there is a large decline in R^2. This is acknowledged to be an unwanted outcome, especially for the purpose of fund performance evaluation. Therefore, it is concluded that it is worthwhile to include the size factor, despite minimal explanatory contribution in Europe and Asia Pacific.

With existing literature such as the aforementioned studies predominantly concluding that candidate models fail the GRS test and on that basis are not mean-variance efficient, it would call into question their applicability and therefore accuracy of inferences drawn from their use. The GRS test is renowned for its power and parameter sensitivity. Levy and Roll (2010) show the extent to which this is the case. They provide evidence that with very slight parameter modifications, the market proxy portfolios are then deemed efficient. A reverse-engineering approach is adopted.[1] The results are promising and help to add some much welcome conviction to an area that so commonly faces criticism – the widespread use of apparently inefficient models.

Levy and Roll (2010) solve an optimisation that seeks to identify the minimal variation in sample parameters that would still allow for mean-variance efficiency to be concluded. Detailed comparison with previous results finds that the GRS test is particularly sensitive to the risk-free rate and length of sample, T. Levey and Roll (2010) find that although rejected by the GRS with monthly risk-free rate of 0.2%, using their range it cannot be rejected. The CAPM requires risk free to be below 0.3% for a rejection at 5% or above 1.3%. in between these values the model cannot be rejected. The authors conclude that mean-variance efficiency of a market portfolio proxy can be achieved with only very small adaptations to the sample parameters. By doing so a "perfectly efficient" market proxy may result. Therefore, in this case, the true expected returns can be estimated with relative precision from the sample betas.

As noted, similar examples are generally absent whereby studies evaluating models for bond fund performance are concerned. However, Gebhardt et al. (2005) conduct a study to determine if betas or security characteristics provide better power in terms of explaining the cross-section of Corporate bond returns. The reliability is judged using

[1]This is the case when using either sample parameters or applying shrinkage adjustments. MV efficiency tests are renowned for their sensitivity to inputs, hence the application of shrinkage

the GRS (1989) test. They conclude that systematic risk is important in explaining the cross-section of average corporate bond returns.

4.3 Methodology

4.3.1 GRS Test

Testing applies to time-series regressions, in which the factor(s) is a portfolio of test asset excess returns. There is the assumption that the risk-free asset exists, and investors will hold combinations of this and the market portfolio; and as such construct a mean-variance efficient portfolio. The expected return on an asset, or portfolio, implied by single or multi-factor linear models[2] is expressed as:

$$E(r_i) = \sum_{k=1} \beta_{i,k}^{K} \, \lambda_k \tag{4.1}$$

where r_i is the excess return of asset i, β_{ik} is the beta of asset i relative to factor k, λ_k is the risk premium for factor k and K is the number of factors in the model. When considering assets individually, time-series regressions can be run, and the significance of the resulting alpha tested;

$$r_{i,t} = \alpha_i + \sum_{k=1} \beta_{i,k}^{K} r_{k,t} + \varepsilon_{i,t} \tag{4.2}$$

where $\varepsilon_{i,t}$ is a random error term with $E(\varepsilon_{i,t}) = 0$ and $E(\varepsilon_{i,t} r_{i,t}) = 0$ for $k = 1, ...K$.

The hypothesis to be tested seeks to identify whether the asset pricing model provides sufficient pricing accuracy in terms of explaining the cross-section of expected

[2] For example, the CAPM, ICAPM, APT

returns. In this case, it is implied that the pricing error is equal to zero, as denoted by the alpha in the regression equation; $H_0 : \alpha_i = 0$ for $i = 1, ...N$. This is the same as was discussed in the context of Jensen's (1968) alpha for performance evaluation; however, here the pricing performance of the model is being tested, rather than ability of the fund managers. The alpha (pricing error) is calculated by subtracting from the asset return, the prediction of the pricing model being tested;

$$\alpha_i = r_1 - \sum_{k=1} \beta_{i,k} r_k \tag{4.3}$$

where $\beta_{i,k} r_k$ represents the weighting on each of the factors and corresponding risk premia that have been specified in the model. To test the significance of the alpha, t-tests are used.

The process as above is useful when evaluating the pricing errors in relation to assets or portfolios individually. However, when conducting an analysis at an aggregate level, it is more useful to determine if the intercepts are jointly equal to zero across the N portfolios of test assets. Such test assets should be selected to represent the investment opportunity set of the fund managers to be evaluated. To do so, a multivariate test is required. The Gibbons, Ross,& Shanken, (1989) (GRS) test is a powerful and absolute test of mean-variance efficiency. This allows for identification of the most reliable model that can be used in a wider application. The GRS test is a test of whether f, i.e. a combination of K factor portfolios is ex ante MV-efficient; whether it is on the MV frontier of the $N + K$ assets. Often it is the case that f (the combination of efficient factors) will lie inside the ex post MV frontier. The frontier has been calculated using sample moments, leading to the scope for misspecification; however, it should not be too far off. Cochrane (2005). For example, even if the selection of factors has been specified to be mean-variance efficient and lies on the frontier, in application it is likely to be the case that there will be returns which will outperform this combination. Thus, the combination of K factor portfolios will no longer be the most efficient selection and therefore will lie inside the ex post MV frontier.

The GRS test statistic is derived as follows:

$$\theta = \frac{(T-N-K)}{N} \quad * \quad \left[1 + [\,\mu' \,\Omega^{-1}\mu\,]^{-1}\right]\alpha' \sum^{-1} \alpha \tag{4.4}$$

Where T the number of time-series observations, N = the number of test asset portfolios, K the number of factor portfolios, $\alpha = (N * 1)$ column vector of regression intercepts, $\mu = (K * 1)$ column vector of average factor portfolio excess returns, $\sum =$ maximum likelihood estimator of the $N * N$ variance-covariance matrix of residuals from the N time-series factor regressions, $\Omega =$ maximum likelihood estimator of the $(K * K)$ variance-covariance matrix of the factor portfolio excess returns.

The hypothesis we wish to test is: $H_0 : \alpha_i = 0$ for $i = 1, ...N$ for a given level of significance, i.e. are the alphas that results from regressing the test asset portfolios against the candidate models jointly equal to zero? Upon rejection, it indicates that the combination of K factors does not provide a mean-variance efficient benchmark model. Therefore, doubt could then be cast on the reliability of this model for performance evaluation.

In most cases, performance evaluation of active funds is concerned with the alpha derived by managers in excess of an appropriately specified benchmark model. To evaluate the ability of active management, the benchmark model should represent a passive alternative on a risk-adjusted basis. Thus, if the alphas are not jointly equal to zero and mean-variance efficiency is rejected, alpha bias will be introduced when evaluating the active funds. As such, it will be more unclear to what extent the alpha observed is due to model misspecification or is indeed due to the ability of the fund. In the context of the Efficient Markets Hypothesis by evaluating the models to identify which can be considered the most reliable, this helps to appease the joint hypothesis problem. (Fama 1970) As noted, the aim of the fund performance evaluation is to obtain a clear picture of the merits of active management, which can be assisted having first reviewed a selection of models that may be most appropriate for this purpose.

4.3.2 Alpha tests

The GRS test as discussed is considered a powerful and absolute test. Comparison between candidate models cannot be made on this basis alone. The variance-covariance matrix of residuals is not constant across the competing models, which is inevitable as different combinations of factors are used in each test. Therefore, to conduct a relative evaluation, other statistics may be calculated using the alphas from the time-series regressions. With the purpose of the models in this study being used to evaluate overall fund performance relative to passive benchmarks, the interpretation of alpha is of key concern. By looking at the alpha statistics in addition to the GRS test, it helps to identify to what extent the alpha observed from the funds may be biased by potentially problematic securities that the models struggle to explain. The statistics that have been used for this purpose are defined as follows (Fama and French 2015):

$A|\alpha|$ - the average absolute alpha that results from regressing the portfolios of test assets against each of the competing models. Lowest values are desirable here. However, the model fit is important too. The alphas can be interpreted with most conviction if the corresponding (adj)R^2 is high, which is an indication that the proposed model is well suited to explaining the test asset portfolio(s), and that the standard error of alpha is low.

Sharpe $|A|\alpha|$ - defined as a Sharpe Ratio for the intercepts (unexplained average returns) of a model, which is considered as the core of the GRS statistic. A lower Sharpe Ratio of intercepts indicates better model performance. This statistic is useful whereby it combines estimates of the alpha dispersion, with the covariance-matrix of regression residuals. However, with respect to this, the Sharpe ratio must be considered alongside the scope for estimation error. For example, high values of (adj)R^2 and low values of the standard errors of alpha for the corresponding model will help to add conviction to this measure. In terms of model comparability, this statistic cannot be used alone to conclude which model is most reliable.

$A|\alpha_i|/A|r_i|$ - This ratio relates the average absolute value of the intercepts, to the average absolute value of the deviation of the test assets from the cross-sectional average excess return across the N test assets. As such, it captures the dispersion in alphas relative to the dispersion in average excess returns.[3] An alternative to this ratio would be to use the market return. It has been noted in literature that the market portfolio may be preferred, due to its central role in asset pricing. The test assets, however, vary between samples. In relation to CAPM-type models for fixed income evaluation, the definition of the market portfolio is more subjective. Thus, the cross-sectional average excess return across the N test assets has been used in this study. A lower value of this ratio is desirable, as it represents the proportion of the cross-section of expected returns that the candidate models are unable to explain.

ASe$(\alpha^2)/A(\alpha^2)$ – Contrary to the previous ratio, high values are preferred in this case. This would indicate that sampling error is driving the dispersion of intercept estimates, rather than the dispersion of the true intercepts themselves.

In summary when identifying what could be considered a mean-variance efficient benchmark model we are hoping that the hypothesis of the GRS test, that the alphas across the test asset portfolios are jointly equal to zero, cannot be rejected. High p-values would be indicative of this. However, the sensitivity and power of the GRS test can at times be a limitation. For example, in the case of a tight fit (high (adj)R^2 and low standard errors), the model is more likely to be rejected. For the purpose of performance evaluation, the primary concern is identifying a model with good fit, from which the smallest absolute alphas result. Furthermore, using the alpha statistics allows for a

[3]Double adjustment ratios may also be used, such as $A|\alpha_i^2|/A|u_i^2|$, which captures the proportion of variance in the N expected returns, left unexplained by the model. However, these can often result in extreme values. This is likely to be more problematic in samples for which there is a limited time-series available, limited range of test assets, and for international samples. As such, reliable inference from these ratios is limited here and the statistics have not been presented.

relative comparison across candidates that the GRS test does not accommodate. Next, in the context of the methodology and tests as just described, the models and data are presented. The models are based upon those discussed in the literature review and the factors known to be important in explaining bond fund returns.

4.4 Models and data

As discussed in the literature review, there are various methodologies and performance metrics that may be used to evaluate actively managed funds. A benefit of the beta representation and comparison to passive benchmarks is that it is relatively easy to interpret. For example, although the indices themselves are not directly investable investors can relate them to ETFs/Index trackers to evaluate potential investment alternatives; which in many cases come with a lower cost in terms of management fees charged. The models are required to be specified as linear beta representations, in either single or multi-factor form, as the methodology that will be used to evaluate the funds in the following chapter is that of Jensen (1968);

$$R_p = \alpha_p^J + \beta_p R_m + u_p \tag{4.5}$$

$$R_p = \alpha_p^M + \sum_j = 1...K\ \beta_p j R_j + u_p \tag{4.6}$$

where α_p^J denotes Jensen's alpha, β_p is the sensitivity to the factor representing the market, and R_m is the corresponding excess return on this factor, and u_p the residual term. In the multi-factor Arbitrage Pricing Theory framework (Ross 1976), the alpha α_p^M results as the abnormal return derived relative to a combination of various factors; $\sum_{j=1...K}$. The sensitivities to which and the corresponding excess returns are captured by β_{pj} and R_j respectively.

The indices used as factors have been obtained from the Morningstar Direct database, which has been the primary data source throughout this thesis. The return properties of the models are further discussed in section 4.6. Notable variation exists between studies on bond fund performance in terms of the data sources and the models constructed for evaluation purposes. Furthermore, in general there is a lack of testing of linear factor models in this area. The equities space, on the contrary, has a wealth of literature directed at this topic. Thus, it highlights there are significant gaps with regards to model testing for bond fund performance evaluation. Table 4.1 presents the selection of models that will be tested in the following analysis. These take the form of a linear beta representation. Chapter 2 outlined the primary risk factors to which bonds are exposed - Term and Credit. Thus, the models have been defined in accordance.

Table 4.1: Model Selection

This table presents the five candidate models that have been specified for testing throughout this chapter. The representative indices are available in the Morningstar Direct database.The two single-factor models proxy for aggregate representations of the UK bond market. The multi-index models also proxy for the UK bond market, however, use a variety of both term and credit factors.

Model Type	**Risk Factors**	**Representative Index(s)**
Single-index		
Model 1: Aggregate	Bond market aggregate	Barclays Sterling Aggregate
Model 2: Credit	Bond market aggregate	Barclays Sterling Aggregate Credit
Multi-index		
Model 3: Risk 2	Aggregate and credit	Barclays Sterling Aggregate BofAML Sterling High Yield
Model 4: Maturity 3	Maturity and credit	BofAML UK Gilts 10+ Yr BofAML UK Gilts 1-10Yr BofAML Sterling High Yield
Model 5: Maturity 5	Maturity and credit	BofAML UK Gilts 1-10Yr BofAML UK Gilts 10+ Yr Barclays Sterling Non-Gilts 10+ Yr BofAML Sterling Non-Gilts 1-10Yr BofAML Sterling High Yield

4.4.1 Single-index models

Contrary to equities whereby a representative market index is readily available, this is not the case for bonds. The magnitude of bond markets made identifying a reasonable proxy challenging. For example, a company's equity is usually represented by a share. Debt on the other hand, is likely to span across many issues, potentially all with various characteristics. Furthermore, the transparency of bond markets is rather more opaque, with trading at times limited. This presents challenges with specifying bond market representations that do not apply to equity markets in such an extent. For example, commonly noted problems in the fixed income space relate to diversification, liquidity, and transparency. When selecting an index as a benchmark another issue to consider is whether it is float/non-float. The former is preferable, as this means that the index is more representative of the investment opportunity set. In the academic literature, indices used as bond market representations are commonly those provided by Lehman Brothers, which have been rebranded under Barclays Capital as of 2008. Two single-index models are proposed in this thesis as bond market aggregate representations. Both of which have been used in the literature, although with a US investment objective.

Model 1 uses the Barclays Sterling Aggregate Index as the market factor. This is known as a broad-based flagship benchmark, incorporating exposures to treasuries, government-related, investment-grade credit, and securitized issues. Although it does capture various bond classifications, it has a bias towards lower risk investments due to the high proportion of Treasuries. The index includes callable bonds; therefore, it may therefore help to match nonlinearities prevalent in the funds' returns. Model 2 is an alternative bond market representation. This index is known to represent the non-securitised component of the Barclays Sterling Aggregate and contains a greater proportion of corporate securities. By testing two single-index models, it can be identified to what extent the selection of the bond market index has on the pricing performance.

4.4.2 Multi-index models

Throughout the existing literature on bond fund performance evaluation, multi as opposed to single-factor models are used. Gultekin and Rogalski (1985) provide some early insights into the challenges of model specification for assessing bond portfolio performance. Their sample focuses on US Treasury bond portfolios over the period 1960-1979, to which they apply both CAPM and Arbitrage Pricing Theory (APT)-type models. Given that their sample focuses on a specific sector and the dominant risk exposure is likely to be interest rates, it would perhaps be expected that a single-factor model would provide sufficient explanatory power.

Nonetheless, both the single and multi-factor models failed to provide a complete explanation of the risk and return relation in the US Treasury market. The results marginally favoured the multi-index models. Such findings are also consistent with those of Elton et al (1995). When they test their four models, it is evident that the addition of the fundamental factors to the single index model results in an improved fit. Given the variety of bond fund categories sampled in this thesis, it seems logical that a multi-index framework would be better suited to capture the risks to which they are exposed. Thus, multi-index models have been also been specified and will be tested alongside the single-factor models to determine the most reliable for performance evaluation.

The Risk 2 model contains two factors combining the Aggregate bond market representation with the High Yield index. This is a similar specification to the Risk 3 model of Blake et al. (1993), which included a Market representation, High Yield index, and a Mortgage-Backed Securities index. The latter has not been included in this study due to the lack of such funds in this sample. Optionality is captured by the securitised component of the Aggregate index. This is evident from the index description that notes callable bonds are included. The High Yield index is added to account for credit risk, particularly in relation to low-grade issues.

Unfortunately, holdings data is not available for the full sample of actively managed funds for the full time-series. However, the portfolio allocations as provided by Morningstar as at July 2016 in Table 4.2 below, show that many of the funds aside from those in the High Yield category do invest in securities of a low credit rating. Therefore, the inclusion of a High Yield index is necessary to account for such risks.

Table 4.2: Holdings Data

This table presents details of the fund holdings as provided by Morningstar. N refers to the number of funds in each category for which this data is available. Each category value presents the proportion in percent (%) of each rating. The data is taken as at the end date of the sample (July 2016), it has not been evaluated over the lives of the funds.

Category	N	AAA	AA	A	BBB	BB	B	B-	N/R
Corporate	120	12.39	10.89	24.91	39.19	5.41	1.70	0.48	5.03
Diversified	62	32.87	25.78	12.74	16.24	3.31	0.99	1.47	6.61
Flexible	14	10.97	17.55	9.41	26.60	15.87	10.06	1.09	8.45
Government	48	-8.62	82.37	1.00	1.35	0.14	0.09	2.00	21.68
High Yield	14	2.28	1.57	2.55	11.99	34.94	31.99	7.92	6.76

As discussed, interest rate risk is the primary factor to which bonds are exposed. The Maturity 3 model incorporates two Government bond indices; the first of which is to represent intermediate term exposures, and the second is for longer-term. In addition to specifying maturity ranges for the Government indices, the Maturity 5 model also does so for Corporate bond indices, again for intermediate and long-term exposures. The High Yield index has been included in each of the multi-factor models. This is consistent with prior studies such as Blake et al (1993). However, Ayadi and Kryzanowski, (2011), do not include this index as they note that there are few High Yield bond funds in the Canadian market. As illustrated above, this is not the only reason for which a High Yield index would be an appropriate addition to the models. Given that the other indices used as factors in the models do not contain low-grade or illiquid securities, but it appears the funds in all other categories do, the High Yield index has been included in each of the models to increase the explanatory power in relation to these holdings. Furthermore, a common proxy to assess exposure to the Credit risk premium is to use the excess returns on a High Yield index. (Palhares and Richardson 2018)

4.4.3 Macro environment

The time periods selected for analysis have been determined by the data availability and the interaction of key macroeconomic variables prevalent at such times.

- Whole Period = April 2002 until July 2016: 172 months - Earliest date at which the Markitt iBoxx benchmarks used to represent the passive portfolios is April 2002. Therefore, the model testing has been conducted since then.
- Period 1 = April 2002 until August 2009: 89 months - Incorporates data up until the most recent period of historically low interest rates.
- Period 2 = September 2009 until July 2016: 83 months – Period of experimental monetary policy whereby interest rates reached historic lows. Continued uncertainty regarding economic outlook. September 2009, however, was not the date at which the Bank of England reduced them to such levels. This was in March 2009. September 2009 allows for a lag period to better all for incorporation of the effects on performance. Additionally, inflation expectations reached a low here; inflation being another risk factor to which bonds in terms of their future valuations are exposed.

As noted, the data in this thesis spans periods of interesting macro-economic events. For example, the period of Quantitative Easing (QE) that was enacted by the BoE in 2009. Existing literature attempts to examine the effects of this upon both the wider economy and asset prices. However, attributing to the specific impact of QE is a challenging task – one which remains ambiguous. Event studies have been used to address the initial impact of QE, however, these are less suitable for assessing over an extended period. However, Breedon et al. (2013) conclude that the impact of QE is significant and economically meaningful on the UK bond market, with regards to the effect on longer-term yields. This is implemented through the portfolio-balance effect. In terms of the wider repercussions from QE on the economy and other assets, the outcome remains ambiguous. Given the that QE coincided with a credit crunch, whereby even the effects of more standard central bank tools are ambiguous, isolating the impact that may be specifically attributable to QE is likely to continue as unresolved for a duration yet.

Although such experimental monetary policy is acknowledged as being a noteworthy consideration in this study, its specific impact upon the results of model testing is less so. It is not apparent that either the test asset or factor portfolios experienced greater sensitivity to time variation than did the other. Furthermore, as indices have been used in both cases, their formation is strictly rule-based and transparency is limited regarding the specific constituents. To account for time-variation, this thesis selects sub-periods to identify if the performance of active funds is significantly different between them. This will be further discussed in Chapter Five. An alternative option in terms of model specification is to condition the risk factors on predetermined variables. For example, the level of interest rates, a measure of credit spreads, or perhaps the January effect (Ilmanen (1995), da Silva et al (2003)) By doing so, time-varying alphas and betas may be incorporated as per the method of Ferson and Schadt (1996). However, the findings from existing bond fund performance literature indicate that the addition of risk factors, as opposed to conditioning variables, has a more notable effect on the performance observed. (da Silva et al (2003)) [4]

The performance evaluation in this thesis is conducted by means of various methodologies, including standard OLS regression as applied in Chapter 5, the entire-cases method of bootstrapping and adjustment for false discovery rates in Chapter 6. The incorporation of conditional variables can be problematic when simulations are involved. Thus, an unconditional approach is applied throughout. An alternative approach to accommodate time-variation is to use the method of dummy variables, as per Ferson and Qian 2010 This defines term structure states as low, normal, or high, and assigns either a one or zero accordingly. By doing so, it avoids the assumption of linearity between the predictor variables and the risk factors in the model. This method has not been adopted

[4] This observation is noted with specific reference to bond funds. The impact of conditioning variables on equity fund performance may have a more of an impact.

in this thesis, however, it is acknowledged as a potential direction for further research.

4.5 Data description

4.5.1 Test assets

The selection of test assets plays an important role when assessing the reliability of models. It is generally considered beneficial to limit the number of test assets. However, that in itself can pose problems if the selection fails to incorporate assets that may not be well explained by the competing models. Here, this limitation is presented due to the lack of an available High Yield index to use as a test asset portfolio. This issue is not isolated with respect to UK coverage, but has also been encountered in previous studies with a US focus. (Dannhauser 2017; Ferson et al. 2006; C. R. Blake et al. 1993). The High Yield market is relatively thin, hence the lack of available data.

Table 4.3 shows the selection test asset portfolios. The index provider of the test asset portfolios is Markit iBoxx.

Table 4.3: Test Asset Portfolios

This table shows the indices that have been selcted as test assets to represent passive portfolios. The index provider is Markit iBoxx in each case. There are 4 to span various credit ratings: A, AA, AAA, and BBB. Another 5 have been selected to cover maturitity ranges: 1-3,3-5,5-7,7-10, and 10-15. The indices are have a rules-based selection scheme, rebalanced monthly based on market capitalisation, and total return methodology.

Markit iBoxx Indices - Test Asset Portfolios		
GBP Corp A TR	GBP Corp BBB TR	GBP Gilts 5-7 TR
GBP Corp AA TR	GBP Gilts 1-3 TR	GBP Gilts 7-10 TR
GBP Corp AAA TR	GBP Gilts 3-5 TR	GBP Gilts 10-15 TR

By using these indices, it attempts to minimise overlap between the test assets and the factors included in the models. For example, many indices are subsets of aggregates. In this case, the tests would be asking the models to price many of the securities that are already included within the factors and therefore inducing a bias towards zero pricing errors. By keeping the index providers separate between the test assets and factors it aims to minimise this effect. A difference is worth noting here between the data used for the test assets here relative to other literature that uses the same methodology for model testing. For example, throughout much of the tests of equity factor models the test asset portfolios are based upon portfolios constructed according various characteristics known to be explain the cross-section of returns. In the case of indices, however, specific details regarding their composition are not always available, hence constituent transparency is limited.

Table 4.4 presents descriptive statistics of the test asset portfolios. The observations from the full sample from 2002 until July 2016 are first presented, with those from the two sub-samples also highlighted.

Table 4.4: Descriptive Statistics - Test Asset Portfolios

Descriptive statistics for the test asset portfolios are presented in the table below. The data is in excess return format. Four credit rating categories and five maturity buckets are represented. The mean and standard deviation data is presented monthly and in percent (%). The t-statistic of the mean and Sharpe ratio (monthly basis) are also included. Data covering the whole period (2002-2016) is presnted along with two sub-periods which represent before (2002 to 2009) and after (2009 to 2016) the golbal financial crisis. Each test asset portfolio is represented by an index as outlined in Table 4.1

Apr 02 - Jul 16	***AAA***	***AA***	***A***	***BBB***	***1-3***	***3-5***	***5-7***	***7-10***	***10-15***
Mean (%)	0.352	0.318	0.286	0.432	0.102	0.208	0.272	0.350	0.410
t mean	2.257	2.545	1.834	3.049	3.324	3.578	3.199	2.963	2.825
Std dev (%)	2.043	1.640	2.047	1.857	0.403	0.761	1.116	1.549	1.905
Sharpe Ratio	0.17	0.19	0.14	0.23	0.25	0.27	0.24	0.23	0.22
Apr 02 - Aug 09									
Mean (%)	0.035	0.055	-0.047	0.128	0.110	0.161	0.202	0.211	0.219
t mean	0.184	0.337	-0.208	0.627	2.032	1.809	1.655	1.327	1.210
Std dev (%)	1.814	1.544	2.117	1.928	0.511	0.838	1.153	1.496	1.706
Sharpe Ratio	0.02	0.04	-0.02	0.07	0.22	0.19	0.18	0.14	0.13
Sep 09 - Jul 16									
Mean (%)	0.691	0.600	0.643	0.757	0.094	0.258	0.347	0.500	0.616
t mean	2.829	3.215	3.055	3.986	3.522	3.507	2.937	2.846	2.686
Std dev (%)	2.224	1.701	1.918	1.731	0.243	0.671	1.077	1.599	2.088
Sharpe Ratio	0.31	0.35	0.34	0.44	0.39	0.38	0.32	0.31	0.29

Considering the whole period, all the mean excess returns are significantly different from zero, except for the portfolio of A rated bonds (at 5%). The 1-3 year portfolio has the lowest return, 0.10% and has also the lowest risk, 0.40%. In terms of the highest reward for risk, this is exhibited by the 3-5yr bonds, as evidenced by the highest Sharpe ratio of 0.27 (monthly basis). Sharpe ratios are generally higher across the Gilt portfolios. Long-term bonds have been found to exhibit much greater volatility of returns, relative to shorter-term bonds. The Sharpe ratios of longer-term bonds are usually found to be less than those of shorter-term bonds, which is the case here. The highest in this respect is 0.27 for the 3-5 year maturity bucket, and the lowest is 0.22 for 10-15 years.

Although the rating portfolios have been selected to represent credit risk, they are also subject to interest rate risk. This is inevitable for debt securities unless it is hedged away. Longer term bonds are subject to greater interest rate risk (duration), driven by uncertainty regarding future term structure movements. The riskiest credit rating of the selection available, BBB, exhibits the highest return 0.43% and a standard deviation of 1.86%. The return of the 10-15yr Gilt portfolio is almost the same, 0.41%, with a standard deviation of 1.91%. It is noted by Ilmanen (2010) that the evidence of the return advantage of credit over Treasury securities is relatively limited. For example, US data shows that the longer-term returns are slightly higher with lower volatilities. Other factors to consider in terms of the attraction of Treasuries over credits is the hedging and diversifying abilities with regards to both recessions and equities, with greater liquidity also provided by Treasuries. Corporates on the other hand are more subject to stale pricing, which can understate correlations with equities and the volatility of the corporate bond indices.

Comparing the descriptive statistics pre and post crisis, a number of characteristics are worth highlighting. In the earlier sub-period all the mean excess returns are indistinguishable from zero, except that of the 1-3yr portfolio which is significant at 5%. The portfolio for A bonds is the only one for which a negative excess return is observed. Relative to the whole period, there is greater variation between the credit and Treasury portfolios in terms of reward for risk. For example, the Sharpe ratios of the credit portfolios range from -0.02 (A) to 0.07 (BBB), whereas those of the Treasuries range from

0.13 (10-15yr) to 0.22 (1-3yr). Overall, the characteristics of the test asset portfolios indicate that the benefits of holding credits relative to Treasuries are minimal, if any, from a reward for risk perspective.

In contrast, the recent period shows that all mean excess returns are significantly different from zero. The magnitude of mean excess returns is far greater than the earlier period. Additionally, there is more consistency in Sharpe ratios across both the credit and Treasury portfolios. However, now the lowest is 0.29 (10-15yr) and the highest is 0.44 (BBB). The standard deviation is highest for AAA and again lowest for 1-3yr. The longer maturity buckets are riskier, particularly 10-15yr Gilts.

Table 4.5 presents correlations between the Test Asset portfolios for each of the sub-periods. Variation in the correlations may give insights as to whether there has been any change in the dispersion of returns the factor models are being required to explain.

Table 4.5: Test Asset Correlations

This table presents the correlations between indices that represent the test asset portfolios to be priced by the candidate models. Correlations for 2002-2009 are presnted in the top section, with those from 2009-2016 below

	AAA	**AA**	**A**	**BBB**	**1-3 yr**	**3-5 yr**	**5-7 yr**	**7-10 yr**	**10-15 yr**
AAA	1.00	0.86	0.75	0.62	0.36	0.53	0.50	0.53	0.57
AA	0.86	1.00	0.94	0.84	0.21	0.35	0.35	0.41	0.48
A	0.75	0.94	1.00	0.91	0.01	0.16	0.21	0.29	0.36
BBB	0.62	0.84	0.91	1.00	0.04	0.16	0.19	0.24	0.29
1-3 yr	0.36	0.21	0.01	0.04	1.00	0.95	0.86	0.73	0.67
3-5 yr	0.53	0.35	0.16	0.16	0.95	1.00	0.93	0.82	0.78
5-7 yr	0.50	0.35	0.21	0.19	0.86	0.93	1.00	0.95	0.91
7-10 yr	0.53	0.41	0.29	0.24	0.73	0.82	0.95	1.00	0.99
10-15 yr	0.57	0.48	0.36	0.29	0.67	0.78	0.91	0.99	1.00

	AAA	**AA**	**A**	**BBB**	**1-3 yr**	**3-5 yr**	**5-7 yr**	**7-10 yr**	**10-15 yr**
AAA	1.00	0.89	0.80	0.64	0.64	0.76	0.82	0.85	0.88
AA	0.89	1.00	0.95	0.85	0.55	0.67	0.72	0.75	0.77
A	0.80	0.95	1.00	0.95	0.40	0.51	0.56	0.59	0.61
BBB	0.64	0.85	0.95	1.00	0.28	0.39	0.43	0.44	0.44
1-3 yr	0.64	0.55	0.40	0.28	1.00	0.94	0.89	0.83	0.79
3-5 yr	0.76	0.67	0.51	0.39	0.94	1.00	0.98	0.93	0.89
5-7 yr	0.82	0.72	0.56	0.43	0.89	0.98	1.00	0.98	0.95
7-10 yr	0.85	0.75	0.59	0.44	0.83	0.93	0.98	1.00	0.99
10-15 yr	0.88	0.77	0.61	0.44	0.79	0.89	0.95	0.99	1.00

For example, tighter correlations would be suggestive that the test asset portfolios present an easier hurdle. The average correlation across the test asset portfolio is 0.60 pre-crisis and 0.76 post-crisis, with the average correlations of the AA and A portfolios increasing by the largest margin of 0.19 between periods. Pre-crisis, the 10-15 year bucket exhibited the highest average correlation of 0.67, and post-crisis the highest observed was 0.82 for the 7-10year bucket. Conversely, the BBB portfolio increased the least, by 0.12. It also exhibited the lowest average correlation in both pre and post-crisis periods; 0.48 and 0.60 respectively. In general, the correlations of all the credit portfolios have increased with the rates portfolios.

4.5.2 Factors

Table 4.6 presents descriptive statistics of the indices used as factor portfolios. Data from the full sample from 2002 until July 2016 is first presented, with those from the two sub-samples also highlighted. These are represented by the Barclays and BofAML indices as per Table 4.1.

Table 4.6: Descriptive Statistics - Factors

This table presents descriptive statistics for the seven factor portfolios. The data is in excess return format. The mean and standard deviation data is presented monthly and in percent (%). The t-statistic of the mean and Sharpe ratio (monthly basis) are also included. Data covering the whole period (2002-2016) is presented along with two sub-periods which represent before (2002 to 2009) and after (2009 to 2016) the global financial crisis. Each factor portfolio is represented by an index as outlined in Table 4.1

Apr 02 - Jul 16	***Agg***	***Credit***	***HY***	***G 1-10***	***G 10+***	***C 1-10***	***C 10+***
Mean (%)	0.338	0.340	0.893	0.230	0.508	0.260	0.444
t mean	2.839	2.795	3.676	3.267	2.499	3.303	2.533
Std dev (%)	1.562	1.598	3.187	0.922	2.665	1.031	2.298
Sharpe Ratio	0.22	0.21	0.28	0.25	0.19	0.25	0.19
Apr 02 - Aug 09							
Mean (%)	0.119	0.081	0.630	0.160	0.217	0.062	0.118
t mean	0.792	0.492	1.710	1.584	0.889	0.555	0.510
Std dev (%)	1.420	1.546	3.474	0.954	2.308	1.060	2.177
Sharpe Ratio	0.08	0.05	0.18	0.17	0.09	0.06	0.05
Sep 09 - Jul 16							
Mean (%)	0.573	0.619	1.176	0.304	0.819	0.471	0.794
t mean	3.109	3.494	3.771	3.127	2.501	4.466	3.031
Std dev (%)	1.679	1.614	2.840	0.887	2.985	0.962	2.386
Sharpe Ratio	0.34	0.38	0.41	0.34	0.27	0.49	0.33

The High Yield index exhibits the highest mean excess return of 0.893%. Gilts 1-10yrs has the lowest, 0.230%. The properties of the Aggregate and Credit indices are very similar, despite the Aggregate comprising of a far greater proportion of Treasuries. As already discussed with regards to the test asset portfolios, a distinct return advantage to credit over Treasuries is not often apparent. Here, the Aggregate has a very slightly higher Sharpe ratio (0.22 vs. 0.21), which would again indicate a favourable risk/reward trade-off.

In both sub-periods, similar trends are again apparent as was the case with the test asset portfolios. In the data up until 2009, none of the mean excess returns are indistinguishable from zero. Here, the Treasury-specific factors offer an increased reward for risk relative to the credits. For example, the Sharpe ratio of the Gilts 1-10 is 0.17 vs 0.06 for Corporates 1-10. Also, an interesting comparison on this basis may be made between the High Yield index and Gilts 1-10. The Sharpe ratios are 0.18 and 0.17 respectively. This is perhaps surprising given the notably different risk profiles that would be assumed between High Yield and Gilts. The Aggregate index Sharpe ratio is greater than that of the Credit index here, although both are very small in magnitude; 0.08 and 0.05 respectively.

The performance of the credit indices is more favourable in the recent period. Across all the factors the mean excess returns are significantly different from zero. The Corporate 1-10yr portfolio has the highest Sharpe ratio of 0.49, followed next by 0.41 for High Yield. A comparison of the Aggregate and Credit indices also shows the positive contribution from a higher weighting to credit in the latter, as evidenced by a higher mean excess return and higher Sharpe ratio of 0.38 for the Credit index, compared to 0.34 for the Aggregate.

Table 4.7 presents correlations for the factor portfolios in the two sub-periods. Pre-crisis, the average correlation across the factors is 0.64, which rises marginally to 0.66 post-crisis. The lowest correlation pre-crisis is 0.25 as exhibited by the High Yield index, and the highest is 0.77 as per the Corporate 10+, Aggregate and Credit indices. Similar applies in the post-crisis period with the High Yield again the lowest (0.30), with Corporate 10+ and the Credit index again the highest (0.79). The average correlation of the Gilts 1-10 factor increased the most, although by only a small margin (0.09). A slight decrease in average correlation is seen for three of the factors; by -0.01 for Gilts 10+ and by -0.02 for Corporate 1-10 and the Aggregate index.

Table 4.7: Factor Correlations

This table presents correlations between factors that are used in combination to comprise the five candidate models to be tested. Correlations for 2002-2009 are presented in the top section, with those from 2009-2016 below.

	G1-10	**G10+**	**C1-10**	**C10+**	**HY**	**Agg**	**Credit**
Gilts 1-10	1.00	0.84	0.49	0.62	-0.23	0.84	0.57
Gilts 10+	0.84	1.00	0.49	0.81	-0.15	0.96	0.70
Corp 1-10	0.49	0.49	1.00	0.82	0.52	0.70	0.93
Corp 10+	0.62	0.81	0.82	1.00	0.31	0.92	0.97
HY	-0.23	-0.15	0.52	0.31	1.00	0.05	0.42
Agg	0.84	0.96	0.70	0.92	0.05	1.00	0.86
Credit	0.57	0.70	0.93	0.97	0.42	0.86	1.00

	G1-10	**G10+**	**C1-10**	**C10+**	**HY**	**Agg**	**Credit**
Gilts 1-10	1.00	0.82	0.48	0.52	-0.28	0.79	0.50
Gilts 10+	0.82	1.00	0.56	0.80	-0.17	0.95	0.71
Corp 1-10	0.48	0.56	1.00	0.84	0.42	0.78	0.94
Corp 10+	0.52	0.80	0.84	1.00	0.32	0.92	0.97
HY	-0.28	-0.17	0.42	0.32	1.00	0.05	0.39
Agg	0.79	0.95	0.78	0.92	0.05	1.00	0.89
Credit	0.50	0.71	0.94	0.97	0.39	0.89	1.00

The High Yield correlations with the Gilt factors increased, although still remained negative. Insights can be gained from the correlations regarding the composition of the bond market proxy indices. Aggregate has a high correlation with both Gilts and Corporate 10+. The Credit index is highly correlated with the two Corporate indices and Aggregate, with its lowest correlation being with the High Yield index.

In summary, the risk and return characteristics of the test asset portfolios appear to be well matched by the seven factors. In terms of outliers, the test asset portfolio representing A rated bonds stands out for its poor performance in the earlier period. It has a standard deviation higher than that of the low-rated BBB bonds and is the only one to have a negative mean excess return. It is unfortunate that there is not an additional High Yield index that could be used as a test asset portfolio. However, given that the number of High Yield bond funds and exposure to speculative-grade securities from the other funds is relatively low, it is not considered to present a significant hurdle with respect to judging the most appropriate model for performance evaluation purposes.

In terms of time variation, it is not the case that either the dependent or independent variables have been impacted to a relatively greater extent than the other.

The correlations diverge marginally between periods. Those of the Test Asset portfolios increased to a greater degree than did the factor indices. This indicates that there is greater concentration within the test asset portfolios for the candidate models to explain. There is a greater dispersion across the factor indices – lower correlations, higher standard deviations – which are helpful characteristics in terms of providing explanatory power.

4.6 Empirical results - model performance

4.6.1 GRS test and alpha statistics

The goal when specifying an asset pricing model is to completely capture the expected returns of the test assets in question. If this were to hold, the regression of the assets' excess returns on the candidate factor model would result in zero intercept (alpha), as proposed by the hypotheses underpinning the GRS test of mean-variance efficiency. However, this is renowned as a high hurdle to pass. Asset pricing models are imperfect descriptions of reality and the GRS test is powerful and sensitive to even marginal adjustments in test design. (Levy and Roll 2010)

As such, literature has broadened the range of evaluation criteria that helps to conclude the suitability of candidate models. This is consistent with the observation of a common pitfall in empirical work being that conclusions are based on consideration of statistics in isolation. There is a need to look beyond the decision based on a p-value alone. The analysis in fact only begins here. (Harvey (2020)) This is a particularly relevant point with regards to model testing and conclusions of mean-variance efficiency. A large literature exists debating this issue, however, primarily in the context of equity factor models. With this in mind, the empirical results are presented across a number of dimensions in order to decide the most appropriate for fund performance evaluation. Table 4.8 presents the results.

Table 4.8: Model Performance Statistics

This table presents the model testing results using the GRS (1989) tests and associated alpha statistics. The whole period (2002-2016) and the two sub-periods (2002-2009 and 2009-2016) have been considered. The GRS statistic and its pvalue are in the first two columns.

2002 - 2016 (172)	GRS	p value	$A\|\alpha\|$	Se$A\|\alpha\|$	(adj)R^2	SR$A\|\alpha\|$	$\frac{A\|\alpha_i\|}{A\|r_i\|}$	$\frac{ASe(\alpha^2)}{A(\alpha^2)}$
Risk 2	3.110	0.002	0.148	0.053	0.738	0.441	1.924	0.117
Maturity 3	1.997	0.043	0.091	0.050	0.826	0.365	1.190	0.195
Maturity 5	1.936	0.050	0.018	0.033	0.916	0.370	0.228	1.272
Aggregate	2.652	0.007	0.064	0.066	0.616	0.393	0.828	0.823
Credit	2.702	0.006	0.097	0.065	0.535	0.396	1.266	0.416
2002 - 2009 (89)								
Risk 2	1.400	0.203	0.182	0.082	0.673	0.410	2.530	0.190
Maturity 3	0.925	0.509	0.112	0.075	0.775	0.344	1.553	0.297
Maturity 5	0.493	0.875	0.027	0.045	0.910	0.262	0.379	2.479
Aggregate	0.844	0.578	0.095	0.101	0.535	0.311	1.324	1.087
Credit	0.884	0.543	0.112	0.095	0.482	0.318	1.560	0.623
2009 - 2016 (83)								
Risk 2	3.163	0.003	0.105	0.063	0.831	0.713	0.558	0.300
Maturity 3	2.898	0.006	0.086	0.059	0.883	0.705	0.479	0.380
Maturity 5	3.035	0.004	0.032	0.034	0.952	0.757	0.179	0.860
Aggregate	4.077	0.000	0.093	0.080	0.727	0.750	0.523	0.381
Credit	3.693	0.001	0.072	0.089	0.608	0.723	0.404	1.641

Model testing provides interesting insights into the scope for bias that may arise. The initial model tests have been conducted to answer the following hypotheses: H_0: Alphas across the test asset portfolios are jointly equal to zero vs. H_a: Alphas across the test asset portfolios are not jointly equal to zero. Table 4.8 presents details from the model evaluation which includes the GRS test results and associated alpha statistics. These were discussed in section 4.3.2 and include: the average absolute alpha, the standard error of absolute alpha, the Sharpe ratio of alpha, the ratio to capture dispersion in alphas relative to the dispersion in average excess returns, and finally the ratio of standard error α^2 relative to α^2; a high value of which would show that sampling error is driving the dispersion of intercept estimates, rather than that of the true intercepts themselves.

The results indicate that only the Maturity 5 model passes the GRS test when the whole period is considered. However, this is only marginal at 5% level of significance and would fail at 10%. The GRS p-values for all the other models are closer to zero, hence the hypothesis that alphas across the test asset portfolios are jointly equal to zero can be rejected. In other words, they do not satisfy the mean-variance efficiency condition of the GRS test. Although a range of GRS statistics have been generated by the candidate models, this is an insufficient basis to judge their relative performance. For example, comparing Maturity 3 (GRS=1.997) with the Aggregate index (GRS=2.652), the magnitude of these statistics would not provide a robust conclusion as to which is superior. This is due to the method of calculation – they have each been formed under a different hypothesis and are therefore not comparable.

Aspects of model fit and absolute alphas are an important consideration. Given that there are limitations imposed by the power and sensitivity of the GRS test, they could in fact be considered of greater relevance than the GRS test alone. The results here show that absolute alphas are the smallest, standard error of alphas the smallest, and the (adj)R^2 the highest for the Maturity 5 model. Despite the marginal results with respect to failing the GRS test, the Maturity 5 model exhibits otherwise preferential characteristics. The absolute alpha is only 0.018%. The next smallest is 0.064% for the Aggregate index. Furthermore, Maturity 5 has the lowest standard error and highest adjusted R^2. This gives conviction to the conclusion that this model is the most reliable. For the purpose of fund performance evaluation higher values of R^2 are desirable. It may even be the case that a factor appears not to be significant in, but if it's inclusion in the model increases the R^2 then it is preferable for it to be left in. (Fama and French 2012) As a basis of comparison, the single-index models reflect opposing results in this respect. Higher absolute alphas, higher standard errors of alpha, and lower R^2 indicate that insufficient explanatory power is offered here.

Until 2009 all the models could be considered mean-variance efficient. Again, the Maturity 5 has the lowest GRS statistics (GRS = 0.493, pvalue = 0.875) and the most attractive alpha statistics. The ratio $A|\alpha_i|/A|r_i|$ shows that the proportion of alphas dispersion unexplained relative to ri is 37.86%. For all the other models, this ratio is greater than 1.00. A high value of ASe$(\alpha^2)/A(\alpha^2$ means that the alpha dispersion is due to sampling error – high values therefore being preferable. Maturity 5 model performs the best with regards to both these ratios.

The recent period, however, highlights the somewhat counterintuitive nature of the GRS test. Here, all the models do in fact possess more appealing characteristics in terms of inducing minimal alpha bias and being most appropriate for fund performance evaluation. The absolute alphas are all lower (except that of Maturity 5), lower standard errors, and higher adjusted R^2s. The ratios are also favourable, indicating less relative dispersion of alphas. However, results from the GRS test show that none of the models can be considered mean-variance efficient. The failure of the GRS test does not, however, present a cause for concern here. This is because it can be attributed to increased tightness of fit - which is in fact a desirable property. Furthermore, the absolute alphas are tiny. This indicates that the when Maturity 5 is used to evaluate active management in the following chapters, the performance results will not be biased by potentially large alphas from exposure to passive strategies (as represented by the test asset portfolios).

4.6.2 Regression Coefficients

To gain further insights into what is driving the results of the model testing statistics, analysis of the regression output from which they are derived provides a means to do so. Indications of the models providing good explanatory power across the test assets are small and insignificant alphas and a high (adj)R^2s

To support the conclusion that Maturity 5 is the optimal candidate model, various aspects of the regression output are worth highlighting. There is only one significant alpha and a high R^2 for each of the test asset portfolios. The lowest is 0.78 which is for the AAA portfolio. The beta loadings are consistent with expectations as to the sources of explanatory power. For example, the BBB test asset portfolio has a negative exposure to the Gilt indices, but a positive and significant exposure to the Corporate and High Yield indices. The exposures across the maturity buckets also fit with expectations. The Gilts 1-10 index contributes explanatory power to all the maturity buckets, whereas Gilts 10+ does so for only the 7-10 and 10-15 yr. portfolios.

There is minimal significant exposure to the High Yield factor in the Maturity 5 model. The most prominent is in relation to the BBB test assets. Significance is observed to a greater extent in the other multi-factor models, as these lack any corporate-specific factors. Unfortunately, there is limited data available for the UK with regards to specifying a High Yield test asset portfolio. However, this is not considered to be problematic in terms of assessing model suitability as there is a low number of High Yield funds in the performance evaluation to follow. Despite this, there is value to be gained from the addition of a High Yield factor in the models. It allows for a more complete range of credit exposures to be considered. Its inclusion presents a tougher hurdle when evaluating the funds. An alternative option to proxy for higher yielding securities would be to use a composite of various B and C rated bonds. Ferson et al (2006) adopted this approach. However, by doing so this has potential to induce an upward bias in the performance observed. REF also acknowledged this a shortcoming of the Ferson analysis. Furthermore, in unreported tests the High Yield factor proved to be significant in spanning tests and increased R^2, highlighting its worthy inclusion even in the absence of many High Yield funds or speculative grade holdings of other funds.

The impact of using factors that contain maturity ranges can be clearly seen from a comparison of Maturity 3 or 5 with either of these Aggregate indices. The (adj)R^2s are all notably increased and in excess of 0.80 when using the maturity models on these buckets. It appears that the 1-3yr Gilt is the most problematic of these. As discussed in

Table 4.9: Regression Coefficients - 2002 to 2016

This table presents the regression coefficients from all candidate models from 2002 to July 2016. The nine Test Asset portfolios are evaluated using each of the five candidate models. Alphas and sensitivities to each factor are provided, with the correponding t-stats included in the row below.

April 02 - July 16		**AAA**	**AA**	**A**	**BBB**	**1-3 yr**	**3-5 yr**	**5-7 yr**	**7-10 yr**	**10-15 yr**
Risk 2	Alphas (%)	-0.059	-0.129	-0.318	-0.152	0.086	0.139	0.147	0.153	0.145
		-0.734	-2.348	-4.247	-2.443	3.277	3.517	3.146	3.265	3.143
	Agg	1.140	0.874	0.935	0.702	0.139	0.360	0.594	0.898	1.143
		23.524	26.163	20.556	18.580	8.731	15.015	20.904	31.569	40.838
	HY	0.028	0.170	0.323	0.388	-0.034	-0.059	-0.085	-0.119	-0.135
		1.161	10.394	14.480	20.935	-4.411	-5.013	-6.082	-8.554	-9.866
	adj R^2	0.76	0.83	0.79	0.83	0.35	0.59	0.73	0.86	0.91
Maturity 3	Alphas (%)	-0.109	-0.150	-0.307	-0.184	0.019	0.021	-0.009	-0.010	-0.014
		-1.118	-2.084	-3.294	-2.405	1.287	1.164	-0.496	-0.364	-0.408
	Gilts 1-10	0.171	0.020	-0.225	0.239	0.587	0.995	1.230	1.118	0.934
		0.937	0.147	-1.283	1.660	21.458	29.504	36.587	21.812	14.949
	Gilts 10+	0.583	0.470	0.562	0.299	-0.093	-0.087	-0.015	0.203	0.405
		9.331	10.153	9.397	6.081	-9.936	-7.588	-1.317	11.590	19.014
	HY	0.140	0.252	0.402	0.458	-0.005	0.003	0.007	0.000	0.004
		4.841	11.761	14.517	20.149	-1.132	0.563	1.339	0.031	0.396
	adj R^2	0.67	0.72	0.70	0.75	0.81	0.92	0.96	0.95	0.96
Maturity 5	Alphas (%)	0.012	0.002	-0.096	-0.005	0.010	0.018	-0.012	-0.002	0.000
		0.152	0.042	-3.276	-0.134	0.717	1.019	-0.683	-0.069	-0.006
	Gilts 1-10	-0.373	-0.325	-0.818	-0.001	0.543	0.913	1.239	1.245	1.126
		-1.973	-3.586	-11.906	-0.006	15.921	21.745	28.888	19.572	14.772
	Gilts 10+	0.415	0.104	0.106	-0.210	-0.044	-0.041	-0.007	0.116	0.271
		4.049	2.123	2.845	-4.753	-2.367	-1.779	-0.298	3.371	6.559
	Corp 1-10	1.040	0.755	1.236	0.619	0.061	0.130	-0.019	-0.197	-0.295
		5.539	8.392	18.098	7.652	1.802	3.104	-0.454	-3.121	-3.896
	Corp 10+	0.116	0.388	0.458	0.578	-0.066	-0.070	-0.009	0.125	0.193
		0.940	6.566	10.216	10.881	-2.986	-2.549	-0.302	3.013	3.888
	HY	-0.120	-0.034	-0.009	0.141	0.003	-0.003	0.014	0.003	0.006
		-3.309	-1.943	-0.697	8.981	0.488	-0.313	1.693	0.221	0.379
	adj R^2	0.78	0.92	0.97	0.95	0.82	0.92	0.96	0.96	0.96
Aggregate	Alphas (%)	-0.035	0.017	-0.040	0.182	0.056	0.088	0.074	0.050	0.028
		-0.450	0.257	-0.371	1.603	2.104	2.156	1.488	0.928	0.506
	Aggregate	1.143	0.890	0.965	0.738	0.136	0.354	0.586	0.887	1.130
		23.510	20.877	14.238	10.364	8.090	13.820	18.721	26.118	32.272
	R^2	0.76	0.72	0.54	0.39	0.28	0.53	0.67	0.80	0.86
Credit	Alphas (%)	-0.012	-0.019	-0.133	0.074	0.077	0.129	0.141	0.146	0.144
		-0.131	-0.560	-3.063	1.194	2.561	2.481	1.941	1.538	1.287
	Corp/Gov	1.066	0.990	1.232	1.052	0.074	0.233	0.386	0.599	0.782
		19.769	47.893	46.107	27.813	4.020	7.319	8.680	10.278	11.349
	R^2	0.70	0.93	0.93	0.82	0.09	0.24	0.31	0.38	0.43

Table 4.10: Regression Coefficients - 2002 to 2009

This table presents the regression coefficients from all candidate models from 2002 to August 2009. The nine Test Asset portfolios are evaluated using each of the five candidate models. Alphas and sensitivities to each factor are provided, with the correponding t-stats included in the row below.

					April 2002 - August 2009					
		AAA	**AA**	**A**	**BBB**	**1-3y**	**3-5y**	**5-7y**	**7-10y**	**10-15y**
Risk 2	Alphas (%)	-0.104	-0.143	-0.359	-0.201	0.117	0.162	0.191	0.186	0.176
		-0.831	-1.672	-3.117	-2.338	2.635	2.559	2.588	2.608	2.542
	Agg	0.972	0.832	1.019	0.800	0.188	0.390	0.620	0.904	1.080
		11.133	13.978	12.736	13.409	6.089	8.863	12.076	18.228	22.472
	HY	0.038	0.157	0.303	0.370	-0.046	-0.075	-0.099	-0.132	-0.136
		1.057	6.466	9.250	15.180	-3.673	-4.189	-4.729	-6.523	-6.908
	adj R^2	0.59	0.73	0.74	0.83	0.35	0.51	0.65	0.80	0.86
Maturity 3	Alphas (%)	-0.188	-0.171	-0.341	-0.248	0.023	0.015	0.006	0.009	0.005
		-1.253	-1.540	-2.446	-2.270	1.151	0.538	0.206	0.203	0.098
	Gilts 1-10	0.326	-0.112	-0.592	0.077	0.723	1.087	1.241	0.951	0.771
		1.194	-0.554	-2.337	0.387	20.069	21.805	22.522	12.235	9.196
	Gilts 10+	0.411	0.481	0.742	0.403	-0.129	-0.135	-0.031	0.258	0.430
		3.743	5.935	7.269	5.036	-8.908	-6.699	-1.386	8.244	12.753
	HY	0.130	0.221	0.363	0.439	-0.001	0.002	0.006	-0.010	-0.004
		2.987	6.876	8.963	13.843	-0.190	0.208	0.715	-0.836	-0.321
	adj R^2	0.43	0.57	0.64	0.73	0.87	0.91	0.94	0.93	0.94
Maturity 5	Alphas (%)	0.045	0.057	-0.042	-0.022	0.020	0.034	-0.003	-0.012	-0.010
		0.418	1.227	-1.215	-0.449	1.010	1.271	-0.080	-0.274	-0.216
	Gilts 1-10	-0.012	-0.164	-0.855	0.375	0.668	1.080	1.253	1.063	0.919
		-0.044	-1.437	-10.087	3.135	13.656	16.558	16.658	10.361	8.243
	Gilts 10+	-0.059	-0.197	0.006	-0.543	-0.078	-0.190	-0.013	0.234	0.360
		-0.312	-2.387	0.090	-6.270	-2.192	-4.019	-0.237	3.143	4.456
	Corp 1-10	1.276	0.878	1.410	0.405	0.065	0.078	-0.048	-0.221	-0.248
		5.279	8.329	17.959	3.652	1.443	1.289	-0.681	-2.330	-2.405
	Corp 10+	0.237	0.517	0.478	0.870	-0.063	0.041	-0.009	0.065	0.116
		1.162	5.805	7.215	9.304	-1.658	0.807	-0.156	0.804	1.325
	HY	-0.163	-0.077	-0.020	0.128	0.005	-0.024	0.017	0.011	0.008
		-3.468	-3.790	-1.317	5.954	0.531	-2.010	1.278	0.619	0.413
	adj R^2	0.72	0.93	0.98	0.95	0.88	0.92	0.94	0.94	0.94
Aggregate	Alphas (%)	-0.081	-0.046	-0.172	0.027	0.088	0.115	0.130	0.105	0.092
		-0.652	-0.452	-1.084	0.171	1.884	1.693	1.596	1.222	1.090
	Aggregate	0.977	0.851	1.055	0.844	0.182	0.381	0.608	0.888	1.064
		11.127	11.782	9.400	7.446	5.510	7.916	10.590	14.722	17.846
	R^2	0.58	0.61	0.50	0.39	0.26	0.42	0.56	0.71	0.78
Credit	Alphas (%)	-0.042	-0.022	-0.151	0.040	0.103	0.143	0.175	0.168	0.166
		-0.380	-0.455	-2.029	0.406	1.969	1.758	1.606	1.259	1.142
	Corp/Gov	0.960	0.955	1.292	1.097	0.083	0.219	0.341	0.522	0.662
		13.352	30.590	26.765	17.342	2.430	4.146	4.825	6.014	7.036
	R^2	0.67	0.91	0.89	0.77	0.06	0.16	0.21	0.29	0.36

Table 4.11: Regression Coefficients - 2009 to 2016

This table presents the regression coefficients from all candidate models from September 2009 to July 2016. The nine Test Asset portfolios are evaluated using each of the five candidate models. Alphas and sensitivities to each factor are provided, with the correponding t-stats included in the row below.

		September 2009 - July 2016								
		AAA	**AA**	**A**	**BBB**	**1-3y**	**3-5y**	**5-7y**	**7-10y**	**10-15y**
Risk 2	Alphas (%)	-0.043	-0.139	-0.263	-0.080	0.049	0.101	0.084	0.099	0.086
		-0.500	-2.107	-2.891	-0.898	2.506	2.364	1.565	1.670	1.460
	Agg	1.258	0.901	0.863	0.619	0.107	0.340	0.581	0.898	1.195
		27.582	25.793	17.943	13.160	10.270	15.040	20.393	28.632	38.390
	HY	0.011	0.190	0.350	0.411	-0.015	-0.032	-0.059	-0.097	-0.132
		0.406	9.190	12.317	14.753	-2.347	-2.396	-3.524	-5.219	-7.161
	adj R^2	0.90	0.90	0.85	0.83	0.57	0.73	0.84	0.91	0.95
Maturity 3	Alphas (%)	-0.069	-0.164	-0.272	-0.099	0.019	0.020	-0.033	-0.041	-0.053
		-0.624	-1.906	-2.413	-0.953	1.450	1.015	-2.174	-1.544	-1.250
	Gilts 1-10	0.164	0.250	0.188	0.356	0.326	0.862	1.235	1.393	1.282
		0.722	1.399	0.805	1.646	12.044	21.554	38.934	25.484	14.533
	Gilts 10+	0.653	0.428	0.413	0.230	-0.028	-0.042	-0.008	0.131	0.332
		9.662	8.045	5.943	3.591	-3.463	-3.521	-0.832	8.052	12.667
	HY	0.148	0.288	0.442	0.476	-0.001	0.009	0.010	0.008	0.006
		4.358	10.789	12.680	14.785	-0.331	1.495	1.999	0.981	0.441
	adj R^2	0.85	0.84	0.78	0.77	0.82	0.95	0.99	0.98	0.97
Maturity 5	Alphas (%)	0.032	-0.034	-0.108	0.033	0.008	0.001	-0.039	-0.021	-0.012
		0.362	-0.844	-3.145	0.965	0.618	0.078	-2.599	-0.824	-0.289
	Gilts 1-10	-0.298	-0.084	-0.454	-0.386	0.290	0.785	1.143	1.425	1.460
		-1.234	-0.758	-4.800	-4.119	8.319	15.515	27.942	20.057	13.187
	Gilts 10+	0.428	0.094	0.028	-0.047	0.010	0.023	0.022	0.069	0.183
		4.393	2.117	0.733	-1.233	0.735	1.130	1.323	2.408	4.107
	Corp 1-10	0.661	0.433	0.899	1.081	0.067	0.137	0.151	-0.066	-0.315
		2.419	3.467	8.410	10.211	1.706	2.403	3.270	-0.822	-2.518
	Corp 10+	0.325	0.487	0.557	0.396	-0.057	-0.098	-0.046	0.092	0.224
		2.619	8.588	11.486	8.237	-3.205	-3.758	-2.176	2.525	3.934
	HY	-0.145	-0.007	0.000	0.048	0.003	0.010	-0.013	-0.010	0.004
		-2.745	-0.291	0.010	2.344	0.365	0.902	-1.402	-0.611	0.143
	adj R^2	0.91	0.97	0.98	0.98	0.84	0.95	0.99	0.98	0.98
Aggregate	Alphas (%)	-0.031	0.080	0.142	0.395	0.033	0.064	0.016	-0.013	-0.067
		-0.379	0.916	0.993	2.489	1.720	1.553	0.290	-0.202	-0.950
	Aggregate	1.259	0.908	0.874	0.633	0.107	0.339	0.579	0.895	1.191
		27.567	18.228	10.768	7.031	9.901	14.498	18.946	24.726	30.012
	R^2	0.90	0.80	0.59	0.38	0.54	0.72	0.81	0.88	0.92
Credit	Alphas (%)	-0.029	-0.033	-0.082	0.138	0.049	0.104	0.076	0.080	0.058
		-0.207	-0.704	-2.145	1.884	1.970	1.646	0.798	0.581	0.327
	Corp/Gov	1.163	1.024	1.171	1.001	0.072	0.249	0.438	0.679	0.902
		14.235	37.078	52.849	23.573	4.910	6.787	7.871	8.509	8.801
	R^2	0.71	0.94	0.97	0.87	0.23	0.36	0.43	0.47	0.49

section 4.8, the absolute alphas were highest for Risk 2. This is evident from the high proportion of significant alphas that result from the regression; 8 out of the 9 test assets. The results show that the most problematic test asset portfolio is that of A rated bonds. For all models, except the Aggregate index, the alpha is significant. However, the R^2 is only 0.54 for this model, but at least 0.70 for all others which indicates greater scope for bias with the Aggregate index.

Contrary to the high explanatory power provided by the Maturity 5 model, the single-index models provide the least explanatory power across all the test asset portfolios, as evidenced by the lower R^2s across all the Test Asset portfolios. The Aggregate index is better fitted to the Maturity bucket portfolios than is the Credit index. However, the factor loadings and R^2s with regards to the 1-3 and 3-5yr portfolios indicate insufficient explanatory power. The alphas here are significant for the single-factor models.

A comparison over time indicates that there are more significant alphas in the later period. Maturity 5 has two significant alphas at 5%. However, the magnitude of these is very small; -0.108% for the A portfolio and -0.039% for the 5-7yr portfolio. The R^2s have increased for all test asset portfolios using this model, except for the 1-3yr bonds. The alphas observed here would not present a cause for concern regarding the bias they would impose when evaluating active management.

Overall, the improved precision in the latter period is the key contributing factor as to why the models fail the GRS test. For the purpose of fund performance evaluation, high R^2s are an appealing characteristic. However, with this comes an increased likelihood of failure with regards to concluding a model is mean-variance efficient (E. Fama and K. French 2012). With the diversity and magnitude of bond markets potentially problematic when specifying appropriate benchmark models, high explanatory power for the test asset portfolio is desirable. Although there is an increased number of significant alphas in the recent period, these do not present a cause for concern in terms of impinging upon model suitability. They are all tiny in magnitude. It is the high R^2s/low standard errors that are causing the models to fail the GRS test. With Maturity 5 having performed consistently the best with respect to these appealing characteristics, it has been concluded to be the most suitable to be used for the fund performance evaluation in the chapters to follow.

4.7 Conclusions

With a distinct lack of model testing being identified in the bond fund performance evaluation literature, it presents a rich field of research questions to be explored. From all the tests conducted, it is evident that the Maturity 5 model is consistently the most reliable. It performs best with respect to minimising the alphas observed in the Test Assets which have been selected to represent passive portfolios. These results can be interpreted with conviction due to the high (adj)R^2 and low standard error of alpha statistics which support a good model fit.

The bad model problem is prevalent throughout fund performance evaluation. Although some of the models may fail in this particular data set, this is not to say that they are of no value in another set up. Various factors must be considered as to why this may be the case. For example, the selection of Test Assets is likely to be paramount. Model selection is also likely to be biased by the purpose for which the model is required. If an attribution analysis is the primary intended purpose, then then the inclusion of additional factors may be necessary; even although they are concluded to be redundant in terms of providing additional explanatory power to the variance of average returns.

Furthermore, in terms analysis of active management, such funds are likely to tilt towards certain strategies, or invest in securities not covered by the test assets, meaning that the selection of test assets may not be fully representative of certain pricing problems that need to be explained by the models. This is recognised as a shortcoming of this analysis. Unfortunately, there does not appear to be another High Yield index with a UK investment orientation, aside from that which is already used as an explanatory variable. Additionally, from the description of the Flexible and High Yield categories, it is evident that some of these funds have ability to invest in concentrated and potentially illiquid positions. These may not be fully represented by the test asset passive portfolios.

Regarding model selection in related studies, a degree of variation exists. This is of course highly dependent upon the data available, and the purpose of the study. A key difference, however, is that model testing is limited. Many use the (adj)R^2 values as the determinant of model fit. A point to notes as a difference between this study and others, is the absence of the mortgage-backed security funds. To provide appropriate risk-adjustment when these funds are included within the sample, a specific risk factor is included within the models. For example, the Risk 3 model of Blake et al., (1993) is frequently used in the literature. This highlights a difference in the investment opportunity set available, as such funds are not included in the UK sample, but are indeed very prevalent in the US market.

The discussion regarding performance evaluation methods has been primarily based upon Jensen's alpha (1968). An alternative approach from the literature is the implementation of term structure models using the stochastic discount factor framework (Ferson et al., (2006a,b)).Model testing here takes and alternative form whereby Hansen's J-statistic is used. It would be interesting to observe the effects of this alternative method when both the crisis period and historically low interest environment are incorporated. This approach may warrant future work. Another potential consideration going forward would be to take inspiration from an approach such as that of Roll and Levy (2010) and reverse engineer the parameters to ascertain the required adjustments to arrive at a mean-variance efficient model for bond fund performance evaluation.

Chapter 5

Bond Fund Performance Evaluation

This chapter uses the five-factor Maturity 5 model, as defined and tested in Chapter Four, to evaluate the performance of a survivorship-bias free sample of UK bond mutual funds. Three sub-sample periods are considered to identify the extent of time-varying performance. Overall, the results show that there is evidence of improved performance in the post-crisis period. The timing method of TM has been applied to determine if the funds appear to exhibit skill with regards to timing a bond market proxy or a style-specific benchmark

5.1 Introduction

Active management has been facing increased scrutiny from regulators on a global basis, with more pressure being placed on managers with regards to transparency and the value they are returning to investors. A further development of note in the UK market is the introduction of the Pensions Freedoms reform, which was enacted in the 2015 budget. Now, it is considered that much risk has been passed to the individuals saving for retirement; they now have much greater discretion with which to invest, as opposed to the previous annuity auto-enrolment. The Financial Conduct Authority (FCA) is undertaking a review of the implications of this thus far. Fixed income funds are renowned for being a popular choice for long-term savings; regarded often as a relatively safe-haven investment vehicle and allow for asset-liability matching. With a wealth of literature directed at analysing equity portfolios, that which relates specifically to bond funds is scarce by comparison.

One of the key factors known to impact the value returned to investors, is the fees charged by the funds. Traditionally, Hedge Funds have been under the spotlight in this respect; however, mutual funds are now also very much within scope. New product development has contributed significantly to this issue. Exchange Traded Funds (ETFs), for example, charge very low fees, with total expense ratios (TERs) even as low as 0.1%. When active management has offered disappointing returns on a net basis, these low-cost alternatives provide an attractive alternative in many cases.

The Efficient Markets Hypothesis in its strongest form asserts that it should not be possible to consistently achieve superior performance. (Fama 1970) It further implies that active management is a value-decreasing activity. Studies of flows of funds, question the rationality of investors in this sense. (Berk and R. Green 2004) If according to theory it is inevitable they will suffer poor performance, why would they continue to allocate money to them? Thus, the performance of actively managed funds has been an extensively studies topic in asset pricing literature, and now is facing further scrutiny from industry regulators as they demand more transparency with regards to fees and value creation.

It is often noted in performance evaluation studies that investors would be best advised to seek out low-cost passive alternatives due to the underperformance that is often identified in the academic literature. However, there are many factors that will impact the alpha observed, in addition to the varying utility derived from such investors from these funds. Net alpha should not be considered as a one size fits all statistic. Nonetheless, given that there is a notable lack of bond fund performance evaluation literature, further research as to the prevalence of alpha in this area is clearly warranted. By doing so, it can help to encourage further analysis regarding the drivers of returns, and consideration of various methodologies with which to evaluate performance and the merits of active management.

As at the date of writing, to the best of my knowledge, no literature has specifically addressed the performance of UK bond mutual funds. The majority is focused on the US, and UK funds are contained within a wider sample incorporating various European countries (Silva et al. 2003). Perhaps given the recent departure of the UK from the EU, it is now even more important that it be considered separately from the rest of Europe within such studies.

The work of Elton, Gruber and Blake (1993, 1995) has been instrumental for much of the subsequent performance evaluation literature on bond funds, with their models underpinning the specification of those that followed. Their study involves two samples of US bond funds; one of which is subject to survivorship bias, the other is not. The funds are evaluated using the methodology of Jensen (1968). The funds' returns are regressed against both single and multi-factor models. As a robustness test, the Quadratic Programming technique of Sharpe (1992) is used to test performance, alongside the linear factor models. This method is a way of testing the accuracy of the fund's strategy definitions. The results of this show that categories, as defined by Morningstar, are consistent with the self-declared styles of the funds. Overall, they do not find evidence of superior performance, relative to any of the models. Furthermore, the alphas are relatively consistent, regardless of model used. The significance of expenses is tested and it is concluded that the funds are found to underperform by a magnitude of costs.

Taking a similar approach, Comer and Rodriguez (2013) evaluate whether significant differences exist between the performance of US Corporate and Government bonds, using a five-factor sector model and a six-index maturity model. They also look at fund flows, aiming to determine the sensitivity of these to the alpha observed. The flows of funds indicate that investors are chasing the highest alpha. Again, the significance of expenses is tested. However, on the contrary to the findings of Blake et al (1993) it is not always the case that funds underperform by a magnitude of costs.

An attraction of using the Jensen's alpha approach, is that is allows for specification of the models on an "otherwise equivalent" basis. This means that the factor(s) have been defined to best mimic the risk of the funds' and represent the performance that would be attainable through investment in a passive alternative. In this sense, alpha is easily interpreted as the extra return generated by active management. It is a commonly used method throughout the equity fund literature, and that of hedge funds too. However, the characteristics of fixed income securities may encourage a different approach to be adopted; for example, the use of term structure models due to the dominance of interest rate risk. Ferson (2006a,b) used the stochastic discount factor (SDF) approach to evaluate the performance of US bond funds; one study considering only Government, and another with an extended cross-section to include a variety of styles. Overall, they find that the samples underperform by approximately a magnitude of costs. Variation in performance is noted, depending upon the economic state prevalent at such time.

When evaluating performance of active funds and managers' skill, the returns-based approach remains to be the most common. An alternative would be to use holdings data, if available . This allows for greater insights as to the actual trading conducted by the fund. On this basis, Moneta (2015) finds that there is evidence of skill. In terms of conclusions regarding the ability of managers, it provides greater clarity in this respect. Also using holdings data, Huang and Wang (2014) show that managers of government bond funds exhibit ability to generate superior performance. Given findings such as these, it highlights the differences in conclusions that may be drawn regarding performance and skill.

In addition to looking at the absolute performance relative to a benchmark, some studies attempt to identify the skills with which the managers are attempting to do so. Given the dominance of interest rate risk, active management is often considered to be a result of market timing as opposed to security selection, with managers attempting to create value through correct anticipation of interest rate movements. (Chen et al., (2010)) Otherwise, timing could be with respect to sector allocation. (Comer (2005)) An issue with timing strategies, however, is the potential for costs to be increased significantly due to trading turnover. Therefore, even if the security selection appears to be optimal, the scope for value additivity is likely to be limited.

Overall, the literature indicates that bond mutual funds struggle to outperform passive benchmarks and return value to investors after costs have been accounted for. Given the increased scrutiny regarding fees, and the lack of literature on the performance of UK bond funds, it highlights a gap in the literature for further research. This chapter aims to identify how the performance of a sample of UK bond mutual funds compares to such findings. They are evaluated relative to the five-factor model, Maturity 5, as found in Chapter Four to be the most reliable benchmark for such purposes. Given the variation in economic conditions over the during the sample period, the data has been sub-sampled to allow for identification of time-varying performance. With fees known to be a key determinant of the value returned to investors, the significance of the Total Expense Ratio (TER) has been determined.

There are three main results from this performance evaluation. First, over the whole period (1999 - 2016) equally-weighted portfolios of UK bond mutual funds underperform the five-factor benchmark model when net excess returns are used. The category of All funds shows a significant monthly alpha of -0.06%. In terms of style-specific performance, the High Yield category exhibits the least negative results, showing neutral performance with an insignificant alpha of -0.01%. Conversely, the worst performing are the Flexible bond funds for which a significant alpha of -0.09% is identified. On this basis, the findings are consistent with much of the academic literature; when evaluated at an aggregate level, the active funds fail to outperform the benchmark model after risk adjustment and costs. The second notable finding identifies that performance is found to vary between sub-sample periods. The results post-crisis period (2009 - 2016) are neutral with Wald tests indicating that Corporate, Government, and Diversified bond funds all experienced significantly improved performance during this time.

To gain some further insights as to the driving factors behind these results the classic timing model of Treynor Mazuy (1966) has been applied, which leads to the third key result. This analysis evaluates the timing ability using two models; a bond market aggregate as proxied by the Barclays Sterling Aggregate index, and the style specific benchmark as assigned by Morningstar. Minimal evidence of timing ability is found using the aggregate market proxy. However, when both the global financial crisis and low interest rate environment are considered, 25% of Corporate bond funds are found to exhibit positive timing ability.

The chapter proceeds as follows; Section 1 outlines the methodology used for the evaluation, Section 2 describes the sample of funds and corresponding descriptive statistics, Section 3 presents the empirical results, Section 4 concludes and provides suggestions for future research.

5.2 Methodology

Investment performance evaluation is most often conducted in terms of comparing the returns of an actively managed portfolio, with those of what should be an "Otherwise Equivalent" benchmark. (Aragon and Ferson 2006) As such, this should allow for identification of the extent of skill the portfolio managers possess to return value to investors. Jensen (1968) shows that, for example in the context of the CAPM, investors will allocate between cash (risk free security), and the market portfolio, depending upon their level of risk aversion. The benchmark portfolio should be a combination of allocation to safe assets and to risky assets, such that the weighting results in the same betas as the fund.

$$R_{p,t+1} = \alpha_p^J + \beta_p R_{m,t+1} + u_{p,t+1} \tag{5.1}$$

$$\alpha_p^J = E\{R_p\} - \beta_p E\{R_m\} = E\{R_p\} - \Big[\beta_p E\{R_m\} + (1-\beta_p)R_f\Big] \tag{5.2}$$

The betas represent weights that specify the "Otherwise Equivalent" strategy. The Jensen's alpha that results is indicative of the superior or inferior performance of the actively managed portfolio, relative to the passive alternative. The Arbitrage Pricing Theory (APT) was developed by (Ross 1976), proposing a multi-factor structure. (Connor and R. a. Korajczyk 1986) adopt this approach for the evaluation of mutual funds. Aragon and Ferson (2006) define multi-beta models in the context of allowing for optimisation of mean-variant efficient portfolios, with incorporation of hedging other relevant risk factors. Jensen's alpha is easily generalized for the multi-beta framework:

$$R_p = \alpha_p^M + \sum_{j=1\ldots K} \beta_{pj} R_j + u_p \tag{5.3}$$

Where Rj, j=1...K are excess returns on the K hedge portfolios. The alpha is derived as the managed portfolio return in excess of the weighted combination of hedge portfolios and the risk-free asset:

$$\alpha_p^M = E\{R_p\} - \sum_j \beta_{pj} E\{R_j\} \tag{5.4}$$

$$= E\{R_p\} - \Big[\sum_j \beta_{pj} E\{R_j\} + (1 - \sum_j \beta_{pj}) R_f\Big] \tag{5.5}$$

The models are estimated using OLS regressions. Newey and West (1987) corrections have been applied to the standard errors in the regression results. By doing so, this adjusted for heteroskedasticity and autocorrelations. The sample of funds was tested for these features and it was identified that this was evident in the majority.

An issue with using OLS regression to evaluate mutual funds, is the tendency for negative weights to be assigned to the factors. In this context, this would indicate that the manager has taken a short exposure. However, mutual funds are subject to investment restrictions that prohibit them from doing so. As discussed, the Sharpe QPS (1992) method can be used to test the fund's strategy classification. Furthermore, it is a way to overcome the issue of negative weights. The results from the model testing in Chapter 4 show minimal evidence of negative exposure. Additionally, findings from existing literature that has used QPS, show that the results are consistent with those of the models estimated by OLS regression in the same studies. Thus, negative weights have not been considered an issue here.

As discussed in the literature review, there are various methods that could be used for performance evaluation of actively managed funds. In this chapter, the Jensen's alpha (1968) approach has been applied. Given that this remains to be the most commonly used method in fund performance evaluation, it allows for a degree of comparison with existing literature, although there is a lack of UK focused studies. The Maturity 5 model will be applied, as this was found to be consistently the most reliable. Results from using a selection of alternative models are presented in the Appendix.

5.3 Data and sample

5.3.1 Data sources

The funds included in the sample have been selected from the Morningstar Direct database. Morningstar has emerged as a powerful participant in the asset management industry, from the perspective of both investors and academics. The data provider conducts its own research and analysis of the extensive universe of funds on its platform. As such, investors with access to this, will have the insights of many analysts and performance metrics at their disposal for making their investment decisions. In terms of data availability for research purposes, it offers coverage of a global database and various types of investment vehicles, e.g. mutual funds, closed-end funds, exchange traded funds (ETFs), and to an extent hedge funds. [1] However, with much of the academic literature focused on the US market, the most commonly used source is the CRSP database. This provides a very comprehensive set of survivorship bias-free returns and fund characteristics. An unfortunate limitation of focusing on the UK market is the lack of data in this respect. Although Morningstar provides an easily accessible history of fund performance, often certain characteristics are limited, e.g., fund size, expense ratios, turnover.

Table 5.1 provides a comparison of the data sources used in a selection of existing literature across various markets. A point to note with regards to fixed income fund performance evaluation is the lack of readily available data with which to construct the required factor models. For example, a style analysis of equity funds is relatively easy to implement, given the accessibility of the commonly used factors of Fama and French (1993). This is not currently the case for fixed income. Thus, there exists further heterogeneity with regards to the specification of factors used through similar studies of fund performance.

Morningstar allocates each fund to a category based upon the self-declared style

[1] Data availability for hedge funds is often limited due to a lack of performance reporting, as a result of the light regulations previously imposed.

Table 5.1: Comparison of Data Sources

This table presents a summary of data sources used in existing academic literature that focus on bond fund performance. The list is by no means exhaustive, but has been selected to provide a reasonable cross-section of the sources and highlight the variation

Study	**Market**	**Dates**	**Data Source**
Blake et al (1993)	US	1987-1991	Morningstar
Elton et al (1995)	US	1980-1992	Morningstar
Ferson et al (2006a)	US	1986-2000	CRSP, Morningstar
Ferson (2006b)	US	1985-1999	CRSP, Morningstar
Moneta (2015)	US	1997-2003	Morningstar Principia CD-Roms
Ayadi & Kryzanowski (2005)	Canada	1985-2000	Financial Post Mutual Fund Database
Da silva (2003)	Europe	1994-2000	Micropal and Datastream
Grose et al (2014)	Europe	2000-2010	Bloomberg and Datastream
Leite et al (2016)	Europe	2001-2012	Portuguese Securities Market Commission and Datastream

of the fund, as noted in their prospectus. As such, these classifications rely upon the accuracy of the funds' disclosures. Even with each category, however, Comer and Rodrigues (2013) note that there is likely to remain a degree of heterogeneity between them. Another way by which the style of the funds could be determined is to use the method of Sharpe (1992). [2] It was found by Chen et al., (2010) and (C. R. Blake et al. 1993) that both this approach and Morningstar result in very similar classifications.[3] The category descriptions show that there has been some progression in terms of the strategies available. This is illustrative how products are developed in accordance with macro-economic conditions, past performance, and the changing needs of investors, as discussed in Chapter 2. Previous studies have predominantly been based upon the more traditional classifications, such as Government and Corporate bond funds.[4] The Flexible and the Diversified categories are not readily identifiable as either one of these. Three new categories were added to the Morningstar database in 2010; Diversified Short-Term, Inflation-linked, and Flexible bond funds. However, the Short-term and Inflation-linked categories have not been included throughout the rest of this thesis due to limited data availability and modelling complexities with regards to inflation.

Over the sample period from 1999 until 2016, the Corporate bond category had

[2]Quadratic programming techniques as discussed in literature review

[3]As also identified by Blake et al. (1993), the results of the regression analysis when comparing those of the QPS with those based on the indices were almost identical

[4](Edwin J. Elton et al. 1995), Ferson (2006), Comer and Rodrigues (2013)

the greatest number of funds. The Diversified category has seen the largest decline in numbers, the majority of which occurring in 2009. At any one point, from January 1999 until July 2016, the maximum number of funds across the categories was 325.[5] The analysis of the attrition has been based upon closures at the fund level. Total of 481 closures on a share class basis for the relevant categories in the sample – mergers or liquidations. As can be seen from the Figure 5.1 below there is a spike in liquidations in 2009, corresponding primarily to the Diversified funds.

Figure 5.1: Bond fund attrition

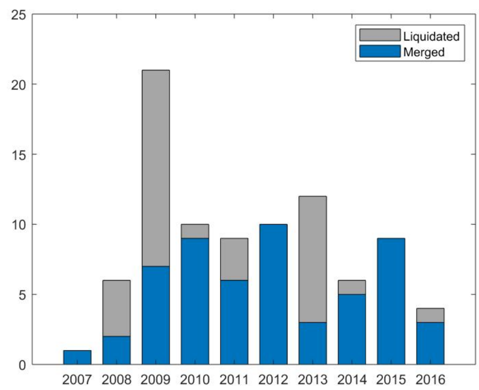


Figure 5.2: Number of Funds

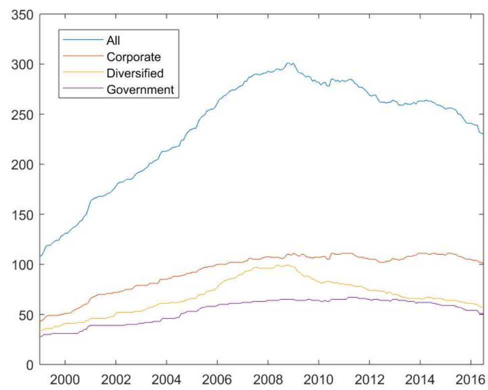


Figure 5.2 shows the trend in fund numbers from 1999-2016. Various reasons may exist for the pattern of closures, aside from the commonly perceived poor performance. For example, the industry may experience further consolidation through more mergers and acquisitions.

Greater concentration of the investor base has become apparent, exacerbated by mergers among UK institutional investors. To ensure successful issuance and gain the attraction of these large issuers, the terms must be attractive, i.e., higher yields. Furthermore, concentration affects liquidity – a small number of large investors holding majority shares. As such, foreign bonds are perhaps perceived as more attractive, unless induced by higher yields. These factors have, to an extent, influenced the higher yielding issued being offered from certain Corporate bond market participants, e.g., those with a lower credit rating or limited access to foreign debt markets. High Yield issuance has

[5]This is having edited the number of funds according to sample selection criteria.

seen an increasing share of the global markets since 2007, albeit from a low level. Other structural changes have been highlighted as a reason for the fall in sterling corporate bond issuance. For example, the UK annuities reform and increased competition from foreign markets. (Elliott and Middeldorp 2016)

5.3.2 Selection of funds

The funds are selected from the Morningstar Direct Global database. They are then required to have a United Kingdom investment orientation, have Sterling base currency, and a minimum return history of 12 months. As discussed, there does not exist to date as far as I'm aware, of a bond fund performance study that is focused on the UK market. By isolating this country orientation, it allows for greater precision in terms of benchmarking. Morningstar provides Total Return data on both a gross and net basis; the difference being the deduction of the funds' Total Expense Ratio (TER) to arrive at the net return. The sample has been limited to the funds for which both net and gross returns are available, allowing for an estimation of costs to be derived for the full time-series of returns. The Morningstar Direct database does provide details regarding the expenses deducted from the funds. However, a full history is not available for the whole sample period of this study, which is from January 1999 until July 2016. As such, the difference between the gross and net returns was considered the most appropriate estimate of costs. A further cost that may be incorporated is that of trading expenses (Ferson et al (2006a)). An estimate may be made using the turnover statistics, however, again such data was incomplete to allow for this.

Multiple share classes are available for most of the funds, however, the one with the longest series of data is selected for inclusion in the sample. Aggregation across the share classes on a total net asset (TNA) weighted basis would be preferable. However, as complete TNA data is not available this has not been possible. Using the asset data that is available, a sub-sample was tested for the sake of comparison. The returns calculated on either basis were almost identical. Excess returns are calculated by subtracting the monthly UK three-month T-bill rate obtained from Datastream.

Equally-weighted (EW) portfolios of funds are formed for each style-specific category as defined by Morningstar. Various approaches may be adopted in terms of creating portfolios, however, data limitations apply. As insufficient fund characteristics data is available for the full sample, sorts on this basis have not been possible. Although it would be useful to potentially allow further insights as to the return drivers, it has been noted in previous literature that bond funds have a stricter adherence to style than perhaps other mutual fund styles do (Ferson et al. (2006b)). As such equally-weighted portfolios based on style are most likely to be an appropriate aggregation method.

5.3.3 Descriptive statistics

Table 5.2 presents the descriptive statistics of equally-weighted portfolios of net excess returns of the sample of funds. The period of analysis is from January 1999 until July 2016, incorporating a total of 211 months. N is the number of funds in the category as at the end of the sample period. The returns are presented monthly and in percent. Age is the average number of years for which the fund has been live. The Total Expense Ratio (TER) has been estimated from the difference between the net and gross returns as provided by Morningstar. The autocorrelations at lags 1 and 2 are denotes by ρ. Inactive refers to both liquidated and merged at the fund level as opposed to share class level. The top table is the whole sample which is survivorship-bias free, extending from January 1999 until July 2016, and the table below presents the statistics for inactive funds for the same period.

Table 5.2: Descriptive Statistics of Sample

This table presents the descriptive statistics of all funds included in the sample. The top section includes all funds - live and closed. Age is in years. Index refers to the return of the category-specific benchmark as assigned by Morningstar. The mean, min, max, and std figures are monthly and in percent. The Sharpe ratio is monthly and in percent. TER refers to the total expense ratio estimated as the differece between the gross and net returns of the funds. ρ_1 and ρ_2 represents the autocorrelations at lags 1 and 2. The bottom section refers only to the inactive funds in the sample. The statistics presented are the same, except the number of merged and closed funds in each category are also included.

	N	**Inactive**	**Age**	**Index**	**Mean**	**Min**	**Max**	**Std.**	**Sharpe**	**TER**	ρ_1	ρ_2
All	372	138	12	0.26	0.21	-3.88	4.47	1.36	0.15	0.84	0.06	0.02
Corporate	146	45	12	0.28	0.22	-5.73	4.87	1.53	0.14	0.81	0.15	0.07
Diversified	115	55	12	0.26	0.17	-3.59	4.16	1.27	0.13	0.93	0.04	0.03
Flexible	16	5	10	0.26	0.23	-8.00	4.46	1.51	0.15	1.15	0.27	0.09
Government	79	27	14	0.27	0.21	-5.16	6.07	1.71	0.12	0.62	0.00	-0.02
High Yield	16	6	13	0.67	0.35	-16.61	8.24	2.27	0.15	1.12	0.32	0.07

	Merged	**Closed**	**Age**	**Index**	**Mean**	**Min**	**Max**	**Std.**	**Sharpe**	**TER**	ρ_1	ρ_2
All	80	58	11	0.26	0.17	-3.92	4.53	1.38	0.12	0.94	0.02	0.04
Corporate	23	22	11	0.28	0.19	-5.65	5.14	1.58	0.12	0.97	0.1	0.04
Diversified	37	18	10	0.26	0.12	-3.28	4.81	1.26	0.09	0.98	-0.02	0.09
Flexible	2	3	8	0.26	0.34	-7.23	7.58	1.71	0.2	1.22	0.24	0.06
Government	17	10	12	0.27	0.15	-5.55	5.14	1.71	0.09	0.67	-0.06	-0.02
High Yield	1	5	11	0.67	0.28	-18.3	8.81	2.37	0.12	1.16	0.34	0.13

Government bond funds have the highest mean average age of 14 years. Second to this is the High Yield category which has an average age of 13. The monthly Sharpe ratios are reasonably consistent across the categories. The High Yield, Flexible, and All categories of funds all have the highest ratio of 0.15, indicating these strategies offer the most incremental excess return per unit of risk. Government bond funds are the least attractive in this respect with the lowest ratio of 0.12. A comparison with the Flexible category for example shows that the Government funds offer a lower average return but have a higher standard deviation.

The High Yield category has the greatest variation in returns, as would be expected. It has the highest mean monthly return of 0.35%, but also the lowest minimum across the categories of -16.61%, which occurred in October 2008, just subsequent to the Lehman Brothers collapse. Its mean monthly return, however, is significantly lower than that of BofAML High Yield index to which Morningstar has benchmarked it. This is the case for all the categories relative to their style-specific benchmarks, although not to the same extent. Ferson et al (2006) found the same underperformance relative to benchmarks. However, not in the case of the High Yield funds as they found the converse applied. In the case of the Ferson et al (2006) example, the selection of the style-specific benchmark is likely to have had an impact on the positive excess returns observed. They measured the returns of the High Yield funds relative to a selection of bonds that were of a higher grade, than those usually classed as High Yield. As such, the return of the benchmark they specified is likely to be downward biased, and thus present a lower hurdle for the High Yield portfolios to beat. The Diversified category offers the lowest mean monthly return. The inactive funds within this category are likely to be a contributing factor, as their mean return is even lower at 0.12%.

Overall, the standard deviation across all the categories is slightly higher than that which has been documented in other international studies. It was found by da Silva et al (2003) that the standard deviation of the UK fund returns was higher than that of the other European markets. Additionally, when compared to the US studies it is also greater. Ferson (2006b) finds that the standard deviations across Government, High Yield, Corporate, High-Quality Corporate, and Mortgage funds are all approximately 1% to 1.5% per month, with mean returns of around 0.6% and 0.7% per month. The standard deviations for the UK funds range from 1.27% for the Diversified category, to 2.27% for the High Yield funds, with mean excess returns of 0.17% and 0.35% respectively. As such, the UK bond funds in this sample appear to be providing a less attractive risk and return opportunity on that basis, with lower mean monthly returns but higher standard deviations.

Autocorrelations can be indicative of stale pricing, which is a feature of bond returns that may bias the performance inferences observed. This is due to the lower transparency of the bond markets, as much trading is conducted over-the-counter (OTC), and less frequently than is the case for equities. Funds that invest in more illiquid securities are likely to have higher first order autocorrelations. It is evident that this is the case from the sample of funds; the High Yield and Flexible categories having significantly higher values of this characteristic than any of the others. These descriptive statistics have been presented for the entire period available; January 1999 until July 2016, however, further examination of separate time periods reveals that the variation in autocorrelations is likely to be indicative of changes in bond market liquidity. For example, during the period of falling interest rates, from July 2007 until July 2009 during the height of the crisis, liquidity was most severely impacted at this point. This is reflected in the autocorrelations, those of the High Yield, Flexible, and Corporate categories being 0.47, 0.45, and 0.34 respectively. In comparison, for the same categories the autocorrelations in the most recent period from August 2009 until July 2016 are 0.07, 0.08, and -0.04. Varying autocorrelation were also observed by Ang & Goetzmann (2009) in relation to the fixed income component of the Norwegian Pension Fund, finding them to be distinctly higher at the time of crisis.

Survivorship bias is another concern that must be addressed in performance evaluation studies. Depending upon the database used, samples may omit the data for all the funds that have ceased to exist over the period of analysis. It is likely the case that many of such funds have suffered poor performance. Consequently, if they are to be omitted, it will cause an upward bias to the average returns observed across the whole sample. By using the Morningstar Direct database, it ensures survivorship bias is mitigated to as great an extent as is possible, as the data on liquidated and merged funds is included. The descriptive statistics show that more than one third of the whole sample is inactive, highlighting that this is a significant proportion of data to include. Blake et al. (1993) found that survivorship bias is not as problematic for bond funds as is the case for equity funds. This is concluded to be primarily due to the lower risk of the bond funds, and as such it is implied they are perhaps less likely to liquidate. However, it is evident from the attrition rate of the sample used in this thesis that both liquidations and mergers have been notable. Given how significantly the asset management industry has been affected by the global financial crisis, this is understandably different to what may have been observed in earlier studies.

Bond fund survivorship bias has more recently been investigated by Rohleder et al. 2012; a motivation for which being the greater diversity of strategies now available. The authors propose that increased variability in returns may contribute to higher attrition rate than has previously been documented. Their study uses a probit analysis to assess the extent of disappearance across styles of bond funds and identify if there are different factors influencing them. They find that fund return has little impact. Fund size, flows and expense ratios are more significant factors; the larger the fund, higher the flows, and lower the expense ratio the better. Their findings regarding the impact of survivorship bias are contrary to those observed by Blake et al. (1993). This paper finds that in their sample the survivorship bias may overstate the returns of the average bond fund by up to 40%. As such, they cast doubt on the results of previous studies whose data has included this bias, for example Cornell & Green, (1991). A rationale given to why size is the dominant factor is that the larger funds can generate sufficient management fees for their sponsor. However, they also note that those funds with the higher expense ratios, of which management fee is a component, are more likely to disappear. Furthermore, it is evident from fee trends across the industry that there are significant economies of scale with regard to fees and fund size. As the funds grown, their fees tend to reduce. Such factors appear to be contradictory to their conclusions about size. Additionally, definitions in terms of what expenses are being considered may vary between samples. Nonetheless, regardless of the specific characteristics contributing to disappearance, the extent of the bias identified in their sample highlights that it is an important data concern to acknowledge.

Of the fund sample selected in this thesis, it appears that the impact of survivorship bias could be quite notable. Across the full sample of funds, 36% are inactive; 21% have closed and 15% merged. The Diversified category has the highest proportion of inactivity, with almost half (48%) either closed or merged. The mean across All funds is 0.18% inactive vs 0.21% whole sample. The highest mean for the whole sample is 0.35% for High Yield, and for the inactive its 0.34% for the Flexible. The lowest mean for whole vs inactive is for the Diversified; 0.17% vs 0.12%. The Flexible category has the highest TER of 1.22 then the High Yield of 1.16; both of these categories had TER of 1.14 in the whole sample. High Yield funds have the lowest minimum return in both samples; -18.33% and -16.61% for inactive vs whole. This category also has the highest standard deviation in both, 2.37% and 2.27%. Inactive funds are younger on average; 10 vs 12 years for the All category.

A characteristic associated with inactive funds is the fees they charge. Fees on an industry wide basis have been trending down over the years, in part due to increasing pressure from both regulators and investors for active managers to justify their costs. Another reason as mentioned by Morningstar is that fees tend to decrease as AUM increases. An analysis of the sample of funds from January 1999 until July 2016 shows that there has been a slight decrease. In general, however, the fees charged by fixed income funds are lower than those charged by equity, or alternative funds. Therefore, the trends observed are likely to be less pronounced. In January 1999 the highest fee charged was 1.29% by the High Yield funds and the average was 0.88%. The highest charge of the period was by the Flexible funds, most recently in April 2003. By July 2016 the average had fallen to 0.85% and the maximum to 1.15% charged by the Flexible funds. Although for a different market so not a like for like comparison, but it can be seen that the fees in the Blake et al (1993) sample were marginally higher.

5.4 Empirical results

The results of the performance evaluation will be primarily presented for the Maturity 5 model as it was concluded in Chapter 4 to be consistently the most reliable. It had the lowest absolute alpha, lowest standard error of alpha, and the highest adjusted R^2. Thus, indicating that the results of the performance evaluation can be interpreted with most conviction for this model, as minimal bias will be induced. As robustness checks and to provide further basis of comparison with existing literature, the results for a performance evaluation using three alternative models are presented in table 5.11 in the Appendix; the Aggregate index and Risk 2 models from Chapter Four, and the style-specific benchmark as assigned by Morningstar. The discussion to follow will be based on the results from Maturity 5, with significant differences relative to the other models highlighted where appropriate.

As has been identified in previous studies although interest rate risk is the dominant factor that causes variation in expected bond returns, it is not theoretically sound that a single-factor model can capture such variation. (Ilmanen (1995), Gultekein (1991)) To better capture the risks across all categories, various indices may be combined to do so. Furthermore, multi-factor models provide use over and above that of single-factor models, as they allow performance attribution analysis, to an extent. The loadings on various indices should give an indication as to what exposures are driving the results. As noted, there have not been any studies, as far as I'm aware, that have focused specifically on the performance of UK bond funds. Thus, there are no directly comparable models that have been previously used.

Table 5.3 presents the results using net excess returns and the Maturity 5 model. A multi-factor model is likely to present a higher performance hurdle and will therefore result in negative alphas of a greater magnitude than would be the case with single-factor models. All categories, except for Government, have negative loadings on the index representing intermediate-term Gilts. This could be interpreted as hedging interest rate risk. All the beta loadings are significant. The Government category has negative exposure to High Yield index, indicating inverse relationship. This is understandable given equity-like characteristics of High Yield bonds and the dominance of interest rate risk for Government bonds.

The Maturity 5 models results in five significant negative alphas. The category of "All" funds shows a significant monthly alpha of -0.06%. Of the categories, Flexible funds perform the worst with a significant alpha of -0.092%, followed by Diversified, -0.08%. The best performance is observed for the High Yield funds with an insignificant alpha of -0.01%, indicating neutral performance. Adjusted R^2s are all high; the lowest is 0.82 for High Yield, 0.90 for Flexible, and 0.96 for all the other categories. There is no evidence of significant positive performance identified for any category. In comparison, the results for single factor style-specific models as presented in the Appendix, show only the government bond funds to have significant underperformance, with an alpha of -0.06%. Otherwise, the alphas derived using the style-specific models are neutral.

Table 5.3: Net Results - Maturity 5 Model

This table presents the results from evaluating equally-weighted portfolios of funds using the Maturity 5 model. Funds are grouped according to Morningstar strategy definitions: High Yield, Government, Flexible, Diversifed, and Corporate. The returns are on a net basis in excess of the risk-free rate. Alphas α are monthly and in percent.T-stats are presented below each coefficient. These are based on standard errors corrected for serial correlation and conditional heteroskedasticity as per Newey and West (1987). Four time periods are considered.

	January 1999 until July 2016							**January 1999 until August 2009**						
	α **(%)**	**Gilts 1-10**	**Gilts 10+**	**1-10yrs**	**10+ yrs**	**HY**	**adj**R^2	α **(%)**	**Gilts 1-10**	**Gilts 10+**	**1-10yrs**	**10+ yrs**	**HY**	**adj**R^2
All Funds	-0.061	-0.107	0.215	0.484	0.203	0.054	0.96	-0.080	-0.094	0.184	0.519	0.226	0.054	0.96
	-3.808	-2.484	5.720	9.057	4.717	7.073		-6.047	-1.324	3.316	5.415	3.272	6.539	
High Yield	-0.013	-0.556	-0.152	0.630	0.254	0.432	0.82	-0.020	-0.323	-0.361	0.542	0.458	0.434	0.80
	-0.200	-3.029	-1.294	2.524	1.506	8.806		-0.267	-1.177	-1.591	1.606	1.584	8.593	
Government	-0.046	0.262	0.568	0.079	-0.023	-0.013	0.96	-0.065	0.349	0.520	0.042	0.021	-0.016	0.95
	-2.433	3.965	14.583	1.284	-0.491	-1.672		-3.365	5.429	13.045	0.513	0.363	-1.634	
Flexible	-0.092	-0.513	0.208	0.749	0.109	0.192	0.90	-0.111	-0.574	0.204	0.800	0.173	0.201	0.91
	-2.634	-4.837	4.997	10.470	2.028	12.668		-2.631	-4.959	2.970	7.487	2.012	10.348	
Diversified	-0.080	-0.024	0.236	0.397	0.163	0.033	0.96	-0.099	-0.009	0.189	0.443	0.180	0.031	0.96
	-4.905	-0.653	8.036	7.692	4.868	3.799		-7.062	-0.194	4.981	6.459	3.813	3.410	
Corporate	-0.065	-0.309	0.053	0.779	0.352	0.041	0.96	-0.083	-0.358	0.061	0.845	0.354	0.041	0.96
	-3.895	-6.015	1.205	12.722	7.043	6.202		-5.921	-4.352	0.979	8.209	4.907	4.857	

	July 2007 until August 2009							**September 2009 until July 2016**						
	α**(%)**	**Gilts 1-10**	**Gilts 10+**	**1-10yrs**	**10+ yrs**	**HY**	**adj**R^2	α **(%)**	**Gilts 1-10**	**Gilts 10+**	**1-10yrs**	**10+ yrs**	**HY**	**adj**R^2
All Funds	-0.095	0.076	0.038	0.323	0.409	0.038	0.97	0.00	-0.04	0.20	0.33	0.23	0.06	0.97
	-1.594	0.955	0.806	3.034	6.314	1.948		0.07	-0.48	2.88	3.46	2.60	6.60	
High Yield	-0.396	0.537	-0.838	-0.316	1.052	0.514	0.79	0.05	-0.94	0.03	0.96	0.02	0.37	0.89
	-0.904	1.450	-4.599	-0.468	2.518	3.835		0.45	-3.05	0.29	3.03	0.14	5.86	
Government	0.019	0.483	0.354	-0.105	0.226	-0.063	0.97	0.00	0.10	0.62	0.10	-0.06	-0.01	0.97
	0.270	6.492	6.755	-1.384	3.520	-2.931		-0.10	1.73	27.56	1.60	-2.22	-0.93	
Flexible	-0.239	-0.440	0.049	0.517	0.418	0.177	0.92	0.04	-0.39	0.18	0.70	0.04	0.15	0.96
	-1.568	-2.543	0.546	2.498	2.930	3.762		1.17	-4.77	4.91	6.85	0.85	6.95	
Diversified	-0.106	0.105	0.084	0.302	0.319	0.013	0.98	-0.03	0.09	0.21	0.22	0.21	0.04	0.97
	-2.190	1.714	1.813	4.490	5.149	0.809		-1.00	1.15	4.43	2.34	3.43	2.54	
Corporate	-0.098	-0.191	-0.075	0.653	0.513	0.035	0.98	0.01	-0.12	0.00	0.52	0.41	0.06	0.96
	-1.638	-1.776	-0.902	6.328	5.848	1.508		0.44	-1.18	-0.05	4.34	4.10	5.41	

Characteristics of High Yield funds can be problematic when applying standard performance measures. (Lipton and Kish 2010) compare the performance of High Yield mutual funds, ranked according to various measures. It is identified that this will differ if the persistence of returns (serial correlation) is adjusted for in the calculations. They evaluate the performance relative to market risk using single and multi-index models in the Jensen's alpha framework, and relative to total risk using Sharpe ratios. Two versions of the Sharpe ratio are compared; the first is the standard calculation, and the second is adjusted for serial correlation using (Lo et al. 2002) methodology. This is because they tested the monthly returns of the funds for serial correlation using the LB Q-test, finding that all exhibit serial correlation to some degree. Given that investors will likely have access to Sharpe ratios when making they allocations, it is important to assess the extent to which these could be biased if certain return characteristics are not accounted for. Sharpe ratios are affected by autocorrelations, skewness, and kurtosis. Overall, they don't find much support in favour of value added by fund managers. The risk-adjusted performance using single and multi-index models shows that the funds are not able to outperform the benchmark index.

A comparison with the results using gross returns, as presented in Table 5.10, shows the impact of expenses. In this case, when the same portfolios are evaluated using Maturity 5 the performance for every category is neutral. Furthermore, in all cases the alphas are positive, except for that of the Diversified funds. Thus, it would indicate that the fund managers do have the skills to outperform the passive benchmark on a risk-adjusted basis, however, not by a great enough margin to cover costs.

5.5 Significance of expenses

The expenses charged by active managers have come increasingly under scrutiny. Greater transparency in this respect has been demanded by both investors and regulators. Previously, it was the fee structure of hedge funds that was most notably the focus of investigation; however, mutual funds also face pressure to justify the costs they charge. Regressions have been conducted to identify if the impact of the fees charged by the funds have a significant impact on the performance. This is consistent with the approach of various similar studies, such as Blake et al (1993), Elton et al., (1995), and Comer and Rodrigues (2013). Table 5.4 presents the results that illustrate the impact of expenses for each category of funds.

Table 5.4: Significance of Expenses

This table presents the results from the regression analysis to identify whether the total expense ratio (TER) is a significant determinant of the alpha observed. Alphas have been estimated for individual funds in each category using the Maturity 5 model from 1999-2016

All Funds	**Avg Exp**	α	**t-stat**	**Exp**	**t-stat**
All	0.84	0.050	0.309	-1.014	-5.911***
Corporate	0.81	0.028	0.142	-0.771	-3.652***
Diversified	0.93	0.072	0.178	-1.441	-3.838***
Flexible	1.15	0.052	0.057	-0.413	-0.593
High Yield	1.12	2.314	2.030**	-2.454	-2.559**
Government	0.62	-0.110	-0.380	-0.635	-1.615

The significance was determined by regressing the alphas for each category of funds against the average total expense ratio (TER) for each category. Here, the alphas are the dependent variables and were derived by regressing the funds in each category against the Maturity 5 model. For the expenses to be considered significant, this would be indicated by a small and insignificant intercept term, but significant slope coefficient. The period of analysis is from January 1999 until July 2016. Time-variation of expense significance has not been analysed, nor has a comparison across models been conducted.

The results are generally consistent with those of existing literature, that the funds underperform on an aggregate basis by a magnitude of costs. As can be seen, the slope coefficient which denotes the total expense ratio (TER), is significant for the categories of All, Corporate, Diversified, and High Yield, but not so for the Flexible and Government funds. Furthermore, although the TER is significant for the High Yield funds, the intercept is also. Thus, it would indicate that the excess performance is not entirely driven by costs.

Some studies find that the performance observed is not directly attributed to the expenses charged. Comer and Rodriguez (2013) show that this is the case, as indicated by the significant intercept for some of the categories. For example, when their sample of high-quality funds are evaluated using any of their single, maturity, or sector models, a significant intercept is observed in each case. The only significant slope coefficient (expense ratio) for these funds results from the single-factor mode. For the other categories (General Corporate, Government Treasury, Government General) the findings are very mixed across the models. Another important point to note in this respect is the key role model specification plays when determining the relationship observed with the expenses. If attempting to draw such conclusions, this highlights the value of conducting model testing, as this helps to select the model likely to impose the minimal alpha bias.

The issue with fees charged by active managers continues to be very much an area of keen interest from both practitioners and academics. As can be seen from Table 5.5 the discrepancies between the fees charged by active managers in general and the low-cost alternatives is quite striking. Therefore, it is not surprising the level of scrutiny that active management faces. However, although disappointing with regards to the value returned to investors, it does continue to present an interesting basis for further academic research. For example, continued development and refinement of methodologies to evaluate skill, greater variety of investment vehicles, and much more competition; all of which adds scope in terms of identifying sources of outperformance. The pressure regarding fees could in fact help to generate improved performance to an extent. For example, Cremers et al. (2016) find that the increased presence of index funds provides incentives for their actively managed counterparts

Table 5.5: Comparison of Fees Across Vehicles

This table presents a summary of representative fees charged across a variety of vehicles

Style	**Fund Name**	**TER**
Bond ETF	iShares Global Aggregate Bond UCITS ETF	0.10%
Bond index-tracker	State Street Global Aggregate Bond Index Fund	0.25%
Bond mutual fund	Average UK bond mutual fund	0.85%
Equity ETF	iShares FTSE 100 ETF	0.07%
Equity index-tracker	Fidelity Index UK	0.06%
Equity mutual fund	Average equity mutual fund (ICI)	1.25%

5.6 Time-variation

The results discussion thus far has evaluated the performance over the entire lives of the funds. It spans an extensive sample period from January 1999 until July 2016, incorporating a variety of macroeconomic conditions. It is identified in the literature that bond fund managers are better known for their market timing ability than their security selection skills. This is to an extent expected due to the sensitivity of the asset class to macroeconomic activity – the management of duration risk being a key driver of performance. Therefore, given the distinct variation in this respect over the whole period, three sub-samples have been specified to see how the performance differs. For consistency, the periods are the same as those per the model testing as far as possible.

- Period 1 = January 1999 until August 2009: 128 months – Incorporates data up until the most recent period of historically low interest rates. Benchmark data is available from January 1999 in this case, therefore the analysis has been conducted since then. This differs slightly from the range used in Chapter 4 to conduct the model testing, as in this case the benchmark data was only available from April 2002.

- Period 2 = July 2007 until August 2009: 24 months – Relates to the global financial crisis period.

- Period 3 = September 2009 until July 2016: 83 months – Period of experimental monetary policy whereby interest rates reached historic lows. Continued uncertainty regarding economic outlook. September 2009, however, was not the date at which the Bank of England reduced them to such levels. This was in March 2009. September 2009 allows for a lag period to better all for incorporation of the effects on performance. Additionally, inflation expectations reached a low here; inflation being another significant factor for performance evaluation.

Previous findings, such as those of Blake et al (1993) and Ferson (2006b), show that time variation impacts the significance and magnitude of the alphas. The former used two samples; one of which included the funds since inception while another was over a five-year period. Only when making the adjustment to specify a certain five-year period were the results significant; otherwise they could not be concluded as being different from zero. This would indicate favourable performance over a shorter-term period. In the evaluation of bond funds across economic states, Ferson (2006b) finds some evidence of significant positive excess returns; all fund styles have a positive performance in states of high credit spreads, additionally High Yield funds earn superior performance during times of low short-term rates, and low volatility. Due to the inverse relationship between interest rates and bond prices, performance is likely to suffer during times when rates are rising. This was the case for some of this earlier period, otherwise interest rates have been on a downward trend.

5.6.1 January 1999 until August 2009

Although there are no directly relatable studies available using UK data, it allows for comparison with those that do exist for different markets, as much of these use pre-crisis data. The results for this earlier sub-period are similar to those when the whole sample from January 1999 until July 2016 is considered. This early sample is the one during which the most negative results arise. Furthermore, there is the greatest degree of significance and consistency across models, as presented in the Appendix. This is comparable with the findings of Blake et al. (1993). There is no evidence of superior performance identified across any of the categories. Alphas range from the worst, -0.11% which is observed for the Flexible funds, to the High Yield which is again insignificant, -0.019%. The category of All funds shows a significant alpha of -0.08% per month.

5.6.2 July 2007 until August 2009

The alphas for all the categories when using the Maturity 5 model are insignificant, except for the Diversified funds, which significantly underperform by 0.1058%. It must be noted, however, that the period under analysis here is only 24 months long and as such the robustness of conclusions that may be drawn here is relatively limited. This period relates to the lowest levels of liquidity in the UK corporate bond market, as discussed by the Financial Conduct Authority (FCA).

Measurement of bond market liquidity can be problematic, particularly due to the OTC trading, and there is no commonly accepted definition of liquidity. As such, the authors use five different measures that have been used throughout literature and practice. They found that liquidity peaked at its worst in Q3 of 2009, gradually improving since then and levelling off from 2011 onwards. Although the concerns regarding bond market liquidity are not quite as pronounced as previously thought in terms of the potential threat to systemic risk, it is still a factor that has potential to impact performance evaluation. For example, as examined by Cici et al. (2011), mutual fund managers may engage in return smoothing behaviour, as they have a degree of discretion with which to mark the holdings. In times of reduced liquidity, this may be more likely and pricing less transparent, as managers rely more so on their own methods of marking. Liquidity is not a factor that has been examined empirically in this thesis, however, it is worth noting as this may have an impact on the results observed.

Liquidity risk presents an additional challenge for bond fund managers relative to equities. With much of the trading conducted OTC and infrequently, this is not surprising. There is a debate as to whether this risk is priced in security markets. Theory proposes that investors should require a premium as compensation for bearing liquidity risk. However, the extent to which this translates to an attractive investment opportunity is debatable. Palhares and Richardson (2018) construct factor-mimicking portfolios to assess the risk and return characteristics and they find that this is not the case.

The results from a selection of alternative models have been presented in Table 5.11 This analysis highlights if there are any notable differences in the active management observed, contingent upon the model selection. It has been previously indicated in existing literature that bond funds are not as sensitive to the choice of model, given the dominance of interest rate risk (Blake et al (1993)). Evidence to contend this can be primarily seen from the more recent data. For example, during the crisis period, the Risk 2 model would attribute significant negative performance across All funds, and the Flexible, Diversified, and Corporate strategies. On the other hand, the Government funds would appear to have significantly positive performance here with this model. Furthermore, the alphas are of a much higher economic magnitude; for example, the Corporate category realising a significant alpha of -0.5160%, rather than the -0.0976% as per the Maturity 5 model.

In terms of comparable results, the scope for this is limited, as many of the bond fund performance evaluation studies use data prior to this time. Comer and Rodriguez (2013) used sub-samples, one of which was from 2007 until 2009, relating directly to the financial crisis. They note that the risk-adjusted performance here differs from the other period they consider in their samples; however, they do not report the results so unfortunately comparison cannot be made.

5.6.3 September 2009 until July 2016

With the experimental policy having been in place since 2008, setting interest rates at historic lows, it makes for a particularly interesting environment in which to observe and evaluate bond fund performance. Monetary policy was loosened in October 2008 when the BoE made the initial rate cut; it was further reduced to its lowest at 0.5% in March 2009. An asset purchase programme was also enacted which contributed to loosening of monetary policy further. As there is an inverse relationship between bond prices and rates, it would perhaps be expected that a low rate environment is conducive to positive performance. This appears to be the case from September 2009 until July 2016. With the Maturity 5 model, alphas are positive for all categories, except for the Government and Diversified funds. However, none of such alphas are found to be significant, therefore indicating neutral performance.

The alpha for the portfolio of All funds is insignificant and almost zero; 0.0017%. High Yield funds show the largest alpha of 0.0514%, ranging to the worst which is observed for the Diversified category. This period further highlights the differences that may be observed between using a single or multi-factor model. For example, Table 5.11 shows that when using the Barclays Aggregate index as the benchmark model, four positive and significant alphas result. The categories of All, High Yield, Government, and Corporate show alphas of 0.1349%, 0.6969%, 0.3645%, and 0.2413% respectively. Additionally, the Corporate bond funds exhibit significant positive performance of 0.06% when their style-specific benchmark, as assigned by Morningstar is used. In terms of the results from the model testing, the Risk 2 model was identified to have potential for the most alpha bias. Table 5.11 shows that this model would result in negative performance across All funds, although not significant. Conflicting signs also appear on a style-specific basis too, with the Risk 2 model showing negative alphas for High Yield, Flexible, and Corporate categories, although again insignificant.

Therefore, considering both the crisis period and subsequent low interest rate environment, this provides evidence of how the conclusions regarding performance may be altered contingent upon the model chosen. It is interesting to note that it is here that this is the case. Much of the bond fund performance literature includes data prior to such times, and therefore the conclusions that bond funds are not as sensitive to the choice of model may be more relevant then. Regardless, it highlights the key role that model specification and testing plays in the determination of superior performance. Furthermore, it is a dynamic are of research warranting continued attention as the universe of funds develops.

5.6.4 Significance of time-variation

Assessment of performance across various market cycles is important in many respects. An interesting point here is with regards to the utility provided by the active managers. For example, a key point to consider is whether they able to provide their investors with adequate performance when it matters to them most, such as during bad economic conditions. (Koswoski (2006)) It has even been noted in bond fund performance literature that value may be derived from the actively managed funds, even in the absence of significant positive alpha. For example, Chen et al (2010) in their study of managers' timing ability suggest that diversification gains can still be made.

An early study that used the dummy variable approach to assess the significance of policy decisions on the performance observed was that of (R. A. Korajczyk and Viallet 1989). The periods of interest related to changes in capital controls across global financial markets. For example, in their study a dummy variable $D_7$4was incorporated into the regression of market indices against various factor models to isolate the performance attributable to the period up until 1974. This date is of significance as it relates to various notable events. These include the change from fixed to floating exchange rate mechanisms, elimination of interest equalisation tax in the US, and a more flexible approach to capital inflows was adopted across various markets.

Recent data has of course given rise to further market disruptions. Leite et al. (2016) examine the performance of euro-denominated bonds available in Portugal during periods of recession and expansion. Two crisis periods are defined in accordance with the Economic Cycle Research Institute; February 2008 until December 2012 which relates to the global financial crisis, and April 2001 until December 2012 which relates to the euro sovereign debt crisis. Aside from these dates, all other periods are considered expansionary. Their sample of funds includes 39 actively managed bond funds, from January 2001 until December 2012. Risk-adjusted alphas are derived from regressing an equally weighted portfolio of the funds against a five-factor model comprised of indices representing Government, Credit, High Yield, Stock arket excess returns, and an additional one to capture the return differential between the PIIGS and other European countries. They find different impacts from each of the recessionary periods; the performance attributable to the global financial crisis is insignificant, while that of the sovereign debt crisis is significant.

This thesis adopts a similar approach to investigates if the performance differs significantly over the range of data available. Dummy variables are used to isolate the performance attributable to sub-periods;

$$R_p = \alpha_p + \alpha_1 D_1 + \alpha_2 D_2 + \beta_p K + \epsilon_p \tag{5.6}$$

where, α_p represents the fund performance from January 1999 until July 2016, α_1 the performance attributable to the period defined by the dummy variable D_1, and α_2 the performance attributable to the period defined by the dummy variable D_2, K represents the factors in the Maturity 5 model, and β_p represents the sensitivity to this model. D_1 represents the performance during the global financial crisis, with a dummy variable of 1 being applied to dates from July 2007 to August 2009, and zero otherwise. D_2 represents the performance during the recent period of historically low interest rates, from September 2009 until July 2016, with a 1 denoting this period and zero assigned otherwise.

$$d_{t,1} = \begin{cases} 1, & \text{if } t = \text{crisis period.} \\ 0, & \text{if } t \neq \text{crisis period.} \end{cases} \tag{5.7}$$

$$d_{t,2} = \begin{cases} 1, & \text{if } t = \text{low rates.} \\ 0, & \text{if } t \neq \text{low rates.} \end{cases} \tag{5.8}$$

Wald tests have been used to determine the significance of the periods as denoted by the dummy variables. Two hypotheses are considered - Hypothesis 1: The fund performance during the global financial crisis is equal to the whole period considered, i.e. $\alpha_c risis = \alpha_{Jan99-Jul16}$ Hypothesis 2: The fund performance during the low rate environment is equal to the whole period considered, i.e. $\alpha_{lowrates} = \alpha_{Jan99-Jul16}$.

Table 5.6: Significance of Sub-periods

This table presents the results from evaluating the significance between periods using the Wald test. α_1 denotes the performance differential during the global financial crisis relative to the whole period. $Wald_1$ is the corresponding statistic to determine if this is significant. α_2 denotes the performance differential during the post-crisis period relative to the whole period. $Wald_2$ is the corresponding statistic to determine if this is significant.

	All	**High Yield**	**Government**	**Flexible**	**Diversified**	**Corporate**
α_1	-0.027	-0.196	0.087	-0.240	-0.040	-0.029
$Wald_1$	0.367	8.604	4.620	5.351	0.432	0.404
α_2	0.041	-0.052	0.061	-0.004	0.040	0.045
$Wald_2$	3.581	1.705	5.212	0.984	4.639	4.152

The high values for the Wald tests indicate that the performance of the High Yield, Government, and Flexible categories during the first sub-sample from July 2007 until August 2009 is significantly different from that over the whole period. For the High Yield and Flexible funds, this period has a detrimental effect on their performance as is evident from the negative alphas. This finding is consistent with these strategies having greater equity-like characteristics, and even allowances for some equity allocation. Furthermore, High Yield funds have been found to behave in part like small cap stocks. (Domian and Reichenstein 2008) The converse applies for the Government category whereby improved performance was observed, which is consistent with the inverse relationship between interest rates and equity markets here.

From September 2009 until July 2016, this period is conducive to better performance for the Government, Diversified, and Corporate categories. This is an interesting point to note, as it has been identified that the low interest rate environment and generally unstable market conditions present a challenge for fund managers. With much of the bond fund performance literature incorporating earlier data, the scope for identifying superior performance on a risk-adjusted basis is perhaps yet to receive significant investigation. Thus, the findings from this performance evaluation highlight that UK bond funds may have the potential to generate more positive performance than is more commonly observed in the academic literature.

5.6.5 Timing ability

The analysis thus far has indicated that there is scope for improved performance in the post-crisis period. Determining the source of this, however, is challenging. In the absence of holdings data it is difficult to ascertain specifics regarding the trading activities of managers. This is due to the complexities involved in differentiating between the effects attributable to term and/or credit factors. The ability to do so is limited when using a regression-based approach. [6]

Furthermore, models such as those available in the equities space that allow for attribution relative to characteristics-based factors are not well developed for fixed income. However, indications as to the positioning in a more general sense can be achieved. For example, the classical timing model of Treynor and Mazuy (1966) allows for inference as to whether funds have gained or lost in relation to market movements. The model takes the form;

$$(R_i - R_f) = \alpha + \beta(Agg) + \gamma(Agg)^2 + \epsilon_i \tag{5.9}$$

The alpha is not of concern here in terms of defining the abnormal risk-adjusted return in excess of the benchmark model. This is due to the squared term not being in

[6] Complex return-splitting algorithms are required for this. Hence the need for fixed-income attribution systems, such as Barclays Point/Bloomberg PORT

excess return format.

Findings from existing literature on bond fund timing abilities finds that there is minimal positive contribution to performance from this aspect. Chen et al (2010) study the timing ability of US bond mutual funds. They conduct an extensive analysis and examine the bias that may be presented by other sources of non-linearities that are inherent in the data.

With interest rate risk the primary driver to which bonds are exposed, timing can often be thought of as indicating the duration management ability of fund managers. A limitation with regards to bond funds relative to equity funds (as has been recurring throughout this thesis) is the composition of the "market" portfolio. A review of the data provided by Morningstar shows that there is a relatively diverse cross-section of Primary Prospectus Benchmarks designated to the funds in each category.

Table 5.7 presents results of the timing ability of the UK bond mutual funds. The Barclays Aggregate index has been first used as the market model. A purpose of which is to allow some degree of comparison between similar studies, albeit a different market such as the US. Furthermore, this flagship model may be used as a default benchmark in databases such as Morningstar. Next, the timing results are presented using the category-specific benchmark as assigned by Morningstar. It has been assumed that this has been selected by the database to provide a reasonable representation of the category characteristics.

Across all categories there is minimal evidence of positive contribution. The Diversified category shows that 3.48% exhibit this and 6.25% of the Flexible category do. Timing coefficients are overall neutral, as although the majority are negative, the proportion which are significant is relatively low. Corporate funds fare the worst in this respect, with 14.28% significantly negative. On an average basis, the High Yield funds exhibit the most negative coefficient (-5.02). Only the Flexible funds have a positive one (1.28). The timing ability using the category benchmarks as assigned by MS is

Table 5.7: Timing Ability 1999 - 2016

This table presents the results a timing study using the TM methodology. Two representative "market" indices have been used. The first is the Bloomberg Barclays Aggregate index; $(R_i - R_f) = \alpha + \beta(Agg) + \gamma(Agg)^2 + \epsilon_i$ and the second is the category benchmark as assigned by Morningstar; $(R_i - R_f) = \alpha + \beta(MS) + \gamma(MS)^2 + \epsilon_i$. Both the fund and index data is in excess return format. For the Diversifed and Flexible categories these are the same. β denotes the sensitivity to this index, γ represents the timing coefficient and $t(\gamma)$ refers to the t-statistic of this coefficient. These values are average terms. The rows below present the proportions within each category of funds which have either positive or negative exposure to these coefficients.

Aggregate Index					
1999 - 2016	**Corporate**	**Diversified**	**Flexible**	**Government**	**High Yield**
$\beta(Agg)$	0.83	0.80	0.67	1.05	0.25
$\gamma(Agg)^2$	-4.43	-5.01	1.28	-1.44	-5.02
$t(\gamma)$	-1.74	-1.36	-1.61	1.20	-1.33
% +ve	6.16	30.43	18.75	83.54	12.50
% +ve (sig)	0.00	3.48	6.25	0.00	0.00
% -ve	93.84	69.57	81.25	16.46	87.50
% -ve (sig)	14.38	13.91	12.50	0.00	6.25
R^2	0.62	0.90	0.40	0.89	0.01
Morningstar Sector					
1999 - 2016	**Corporate**	**Diversified**	**Flexible**	**Government**	**High Yield**
$\beta(MS)$	0.84	0.80	0.67	1.01	0.63
$\gamma(MS)^2$	0.52	-5.01	1.28	-0.37	-0.95
$t(\gamma)$	1.17	-1.36	-1.61	-0.64	-0.94
% +ve	63.01	30.43	18.75	26.58	18.75
% +ve (sig)	25.34	3.48	6.25	3.80	0.00
% -ve	36.99	69.57	81.25	73.42	81.25
% -ve (sig)	11.64	13.91	12.50	5.06	18.75
R^2	0.95	0.90	0.40	0.96	0.76

again generally neutral, however, not so many negative coefficients are identified. This improvement is primarily in relation to the Corporate category. Here we see that 25.34% of Corporate funds exhibit positive timing ability. The proportion of significant negative timers has reduced from 14.38% with the Aggregate index to 11.64%. Government funds also fare slightly better, as 3.8% are now significant and positive. However, 5.06% are significant and negative. Across all categories the model fit has improved. As such, likely to be more representative of aspects of the "market" the funds are hoping to time.

Table 5.8 presents the results incorporating both the global financial crisis and subsequent low rate environment that ensued. The results here show that it is evident that it is during this time whereby the funds achieved the majority of their positive timing ability. For example, 23% of Corporate funds exhibited positive timing ability, 4.5% Diversified, 6.25% Flexible, and 2.5% Government. High Yield funds perform poorly with regards to timing their category benchmark; 46.6% of which show significant negative timing ability. Referring back to Table 5.2, these descriptive statistics showed that the High Yield benchmark return was considerably higher than the category average and the R^2 for this category the lowest. Furthermore, liquidity issues are likely to be more prevalent in this category, with stale pricing and higher autocorrelations potentially creating inference issues.

Table 5.8: Timing Ability 2007 - 2016

This table presents the results a timing study using the TM methodology. Two representative "market" indices have been used. The first is the Bloomberg Barclays Aggregate index; $(R_i - R_f) = \alpha + \beta(Agg) + \gamma(Agg)^2 + \epsilon_i$ and the second is the category benchmark as assigned by Morningstar; $(R_i - R_f) = \alpha + \beta(MS) + \gamma(MS)^2 + \epsilon_i$. Both the fund and index data is in excess return format. For the Diversifed and Flexible categories these are the same. β denotes the sensitivity to this index, γ represents the timing coefficient and $t(\gamma)$ refers to the t-statistic of this coefficient. These values are average terms. The rows below present the proportions within each category of funds which have either positive or negative exposure to these coefficients.

Aggregate Index					
2007 - 2016	**Corporate**	**Diversified**	**Flexible**	**Government**	**High Yield**
$\beta(Agg)$	0.81	0.81	0.64	1.04	0.20
$\gamma(Agg)^2$	-5.52	-1.96	-5.80	1.39	-6.02
$t(\gamma)$	-1.80	-1.61	-1.62	0.96	-1.00
% +ve	4.20	29.46	18.75	82.28	6.67
% +ve (sig)	0.00	4.46	6.25	1.27	0.00
% -ve	95.80	70.54	81.25	17.72	93.33
% -ve (sig)	20.98	14.29	31.25	1.27	13.33
R^2	0.71	0.89	0.18	0.93	-0.01
Morningstar Sector					
2007 - 2016	**Corporate**	**Diversified**	**Flexible**	**Government**	**High Yield**
$\beta(MS)$	0.81	0.81	0.64	1.00	0.79
$\gamma(MS)^2$	0.32	-1.96	-5.80	-0.56	-2.05
$t(\gamma)$	0.80	-1.61	-1.62	-0.89	-2.02
% +ve	59.44	29.46	18.75	27.85	0.00
% +ve (sig)	23.08	4.46	6.25	2.53	0.00
% -ve	40.56	70.54	81.25	72.15	100.00
% -ve (sig)	13.29	14.29	31.25	8.86	46.67
R^2	0.96	0.89	0.18	0.98	0.80

Clare et al (2019) find consistent results from the US bond market. They evaluate the performance of 884 US bond mutual funds from 1998-2017. To assess skill they apply the standard approach of TM (1966). All funds in their paper are measured relative to the Barclays US Aggregate Bond index. They conclude that bond funds lack market timing skills and highlight that the High Yield funds exhibit value decreasing ability to the greatest extent.

Timing credit spreads may be proposed by managers as a way to increase exposure to the credit risk premium as a tactical approach. However, evidence suggests this is not likely the case as there is not a notable deviation in exposure over time. (Brooks et al. 2019) Instead, this has tended to be consistent; an unfortunate result of which has been increased correlation with equity markets. Boney et al (2009) study the timing ability of US bond mutual funds. They conclude that there is minimal evidence of timing ability as a means by which to generate alpha, but note that allocation to bond funds could provide opportunities for diversification. The results of Brooks (2019) would suggest that this may not be the case.

Funds with greater flexibility in terms of restrictions regarding credit downgrades could realise increased premiums as a result. (Ng and Phelps 2011) This is also an area whereby performance deviations relative to a representative index may notably occur. For example, the rules by which such benchmark constituents are based may impose a more stringent governance. If this is the case for these indices, then they are found to suffer as a result. For example, the month in which the downgrade occurs tends to be associated with a spike in the risk premium and liquidity cost. Therefore, the spread premium is foregone due to the requirement to offload such securities. Excess Investment Grade and High Yield (relative to Treasuries) were down during 2007 and 2008, but rebounded in 2009. The paper identifies that allowing for tolerance to downgrades is a key factor that would allow investors to capture a larger spread premium. The authors illustrate this from alternative index constructions. They find that the return enhancement is due to specifics of these securities, as when the same proportion is substituted for a High Yield index the effect does not occur to the same magnitude.

Such characteristics are consistent with those observed by Dor and Xu (2011) in their study regarding 'Fallen Angel' bonds - i.e., those which have been downgraded from Investment Grade to High Yield. In the month of the downgrade, fallen angel bonds are found to underperform High Yield counterparts. However, the subsequent reversal may persist for up to two years and is relative to the initial underperformance – i.e., the larger the initial decline, the larger the reversal. Overall, fallen angels present attractive characteristics in comparison to High Yield bonds. This is not found to be isolated to specific periods and prevails even when higher trading costs are incorporated.

The observations here suggest that dynamics of credit markets during and around the crisis period are contributing to the divergence in performance observed. Further granularity regarding fund versus benchmark composition and credit shifting during this time would provide valuable insights as to what is driving the apparent timing ability and whether this leads to outperformance in the subsequent period.

5.7 Conclusions

Most bond fund performance literature finds underperform relative to the benchmark models by approximately a magnitude of costs. Thus, the main objective of this chapter is to identify if this also applies to a new sample of bond funds with a UK investment orientation. The Jensen's alpha (1968) methodology is employed here to conduct the first empirical investigation regarding the performance of UK bond mutual funds. The model used to do so was that which was found to be consistently reliable in Chapter 4 in terms of minimising alpha bias, named Maturity 5.

The results here show evidence that is generally consistent with similar studies (Blake et al (1993), Elton et al (1995), da Silva et al (2003)); that the funds struggle to outperform on a risk-adjusted basis. Over the sample period from 1999 to 2016, the equally-weighted portfolio of All funds underperforms by -0.06%. In terms of style-specific performance, the High Yield category exhibits the least negative results, with an insignificant alpha of -0.013%. On the contrary, the category of Flexible bond funds has the lowest monthly alpha over the whole period, -0.0912%. This is perhaps surprising, given that as per the Morningstar category definitions, it appears they are similar in style to “Alternative Mutual Funds” . As such, they are supposed to perform well in both market upturns and downturns.

However, the results are more promising when the sub-samples are considered. For example, it appears that during the low interest rate environment since 2009 there is evidence of better performance. Neutral performance is exhibited across the portfolio of All fund here, as opposed to significantly negative as per the full sample period from 1999-2016. The use of Wald tests indicates that the difference is significantly more positive for the Corporate, Diversified, and Government funds. Nonetheless, there is still no evidence of superior performance after costs. A further regression analysis to complete the performance evaluation in this chapter adds support to the general conclusions that bond funds underperform by approximately a magnitude of costs.

Limitations exist with regards to ascertaining the specific sources of outperformance in the absence of holdings data. Nonetheless, a judgment can be gained as to the positioning of the funds and their ability to time market movements. An application of the Treynor Mazuy (1966) model shows favourable results from the Corporate bond category. However, the definition of the ”market” portfolio plays a key role here. When evaluated relative to the Barclays Sterling Aggregate index there is no evidence of significant positive timing ability, yet using the Barclays Sterling Corporate index as assigned by Morningstar 25% of such funds exhibit this.

5.8 Appendix

Table 5.9: Morningstar Categories

This table presents descriptions of bond fund categories as defined by Morningstar. The allocation is based-upon the self-declared style of the fund. All funds selected for inclusion in this thesis have a UK investment orientation and Sterling base currency. The benchmarks listed are those assigned by Morningstar as representative of each category

Category	**Description**	**Benchmark**
GBP Corporate	Invest principally in investment grade corporate-issued securities denominated in or hedged into GBP	Barclays Sterling Agg Corp TR GBP
GBP High Yield	Invest principally in sub-investment grade securities with a credit quality equivalent to BB, or lower and denominated in or hedged into GBP	BofAML GBP HY TR GBP
GBP Flexible	Have the flexibility to invest across a range of bond types and can exhibit significant risk concentrations. Such concentrations may include, but are not limited to, large exposures to non-investment grade securities and some moderate exposure to emerging-markets debt	Barclays Sterling Agg TR GBP
GBP Government	GBP Government Bond funds invest principally in government or explicitly government-backed agency securities denominated in or hedged into GBP	Citi UK GBI TR GBP
GBP Diversified	Invest principally in investment grade corporate and government issued bonds denominated in or hedged into GBP. These funds do not focus on a single sector	Barclays Sterling Agg TR GBP

Table 5.10: Gross Results - Maturity 5 Model

This table presents the results from evaluating equally-weighted portfolios of funds using the Maturity 5 model. Funds are grouped according to Morningstar strategy definitions: High Yield, Government, Flexible, Diversifed, and Corporate. The returns are on a net basis in excess of the risk-free rate. Alphas α are monthly and in percent.T-stats are presented below each coefficient. These are based on standard errors corrected for serial correlation and conditional heteroskedasticity as per Newey and West (1987). Four time periods are considered

	January 1999 until July 2016							**January 1999 until August 2009**						
	α (%)	**Gilts 1-10**	**Gilts 10+**	**1-10yrs**	**10+ yrs**	**HY**	**adjR^2**	**α(%)**	**Gilts 1-10**	**Gilts 10+**	**1-10yrs**	**10+ yrs**	**HY**	**adjR^2**
All Funds	0.012	-0.106	0.214	0.483	0.204	0.054	0.96	-0.004	-0.095	0.184	0.520	0.226	0.053	0.96
	0.781	-2.454	5.689	9.026	4.727	7.094		-0.259	-1.338	3.312	5.406	3.258	6.511	
High Yield	0.082	-0.556	-0.153	0.630	0.254	0.432	0.82	0.077	-0.324	-0.361	0.543	0.458	0.434	0.8
	1.235	-3.025	-1.297	2.520	1.507	8.806		1.051	-1.177	-1.589	1.605	1.584	8.595	
Government	0.011	0.262	0.567	0.079	-0.022	-0.013	0.96	-0.006	0.348	0.521	0.043	0.021	-0.016	0.95
	0.616	3.989	14.583	1.275	-0.477	-1.692		-0.289	5.422	13.015	0.522	0.359	-1.645	
Flexible	0.003	-0.511	0.206	0.746	0.112	0.192	0.9	-0.014	-0.571	0.202	0.799	0.176	0.201	0.91
	0.066	-4.587	4.777	10.213	2.052	12.506		-0.322	-5.342	3.168	7.734	2.056	10.209	
Diversified	-0.003	-0.023	0.236	0.397	0.163	0.033	0.96	-0.021	-0.009	0.189	0.443	0.180	0.031	0.96
	-0.185	-0.623	7.997	7.679	4.870	3.818		-1.461	-0.185	4.976	6.465	3.820	3.426	
Corporate	0.011	-0.309	0.052	0.778	0.353	0.041	0.96	-0.002	-0.361	0.062	0.847	0.353	0.041	0.96
	0.710	-5.939	1.177	12.609	7.034	6.207		-0.157	-4.377	0.991	8.179	4.862	4.833	

	July 2007 until August 2009							**September 2009 until July 2016**						
	α (%)	**Gilts 1-10**	**Gilts 10+**	**1-10yrs**	**10+ yrs**	**HY**	**adjR^2**	**α(%)**	**Gilts 1-10**	**Gilts 10+**	**1-10yrs**	**10+ yrs**	**HY**	**adjR^2**
All Funds	-0.021	0.077	0.038	0.323	0.409	0.038	0.97	0.070	-0.038	0.199	0.333	0.227	0.061	0.97
	-0.350	0.958	0.803	3.032	6.311	1.951		2.737	-0.398	2.856	3.340	2.608	5.872	
High Yield	-0.302	0.538	-0.839	-0.317	1.053	0.515	0.79	0.144	-0.936	0.034	0.956	0.020	0.367	0.89
	-0.689	1.451	-4.600	-0.470	2.519	3.840		1.268	-3.039	0.289	3.017	0.136	5.863	
Government	0.075	0.483	0.354	-0.105	0.226	-0.063	0.97	0.050	0.107	0.619	0.102	-0.055	-0.011	0.97
	1.084	6.479	6.749	-1.384	3.518	-2.925		2.219	1.808	28.367	1.620	-2.283	-0.902	
Flexible	-0.145	-0.441	0.049	0.518	0.418	0.177	0.92	0.128	-0.386	0.184	0.694	0.040	0.154	0.96
	-0.950	-2.547	0.549	2.499	2.925	3.759		3.741	-4.716	4.856	6.788	0.866	6.920	
Diversified	-0.028	0.106	0.084	0.302	0.320	0.013	0.98	0.042	0.094	0.209	0.220	0.213	0.042	0.97
	-0.584	1.725	1.806	4.481	5.152	0.813		1.289	1.193	4.527	2.310	3.552	2.448	
Corporate	-0.022	-0.191	-0.075	0.654	0.514	0.035	0.98	0.074	-0.111	-0.005	0.517	0.407	0.064	0.96
	-0.368	-1.774	-0.903	6.328	5.847	1.508		4.929	-1.171	-0.063	4.330	4.076	5.826	

Table 5.11: Performance Results - Alternative Models

This table presents the results from evaluating equally-weighted portfolios of funds using the Maturity 5 model. Funds are grouped according to Morningstar strategy definitions: High Yield, Government, Flexible, Diversifed, and Corporate. The returns are on a net basis in excess of the risk-free rate. Alphas α are monthly and in percent.T-stats are presented below each coefficient. These are based on standard errors corrected for serial correlation and conditional heteroskedasticity as per Newey and West (1987). Four time periods are considered

	January 1999 until July 2016										**January 1999 until August 2009**									
	Aggregate			**Style**			**Risk 2**				**Aggregate**			**Style**			**Risk 2**			
Category	α	**(Agg)**	R^2	α	**(MS)**	R^2	α	**(Agg)**	**(HY)**	R^2	α	**(Agg)**	R^2	α	**(MS)**	R^2	α	**(Agg)**	**(HY)**	R^2
All	0.005	0.810	0.84	0.005	0.810	0.84	-0.069	0.790	0.120	0.92	-0.057	0.860	0.84	-0.057	0.860	0.84	-0.095	0.830	0.110	0.92
	0.100	29.830		0.100	29.830		-2.400	41.180	9.940		-0.930	21.680		-0.930	21.680		-2.800	30.260	7.230	
High Yield	0.296	0.210	0.01	-0.070	0.620	0.76	-0.086	0.090	0.610	0.76	0.108	0.360	0.04	-0.080	0.610	0.72	-0.092	0.200	0.600	0.72
	1.330	2.140		-0.640	10.020		-0.740	1.770	10.260		0.370	1.990		-0.570	7.040		-0.670	2.310	7.370	
Government	-0.064	1.060	0.89	-0.060	1.000	0.96	-0.002	1.080	-0.100	0.93	-0.053	1.040	0.87	-0.080	1.010	0.95	-0.024	1.070	-0.090	0.91
	-1.450	44.730		-6.000	65.090		-0.040	51.510	-5.610		-0.930	23.440		-8.000	87.220		-0.550	32.500	-4.950	
Flexible	0.065	0.620	0.4	0.065	0.620	0.4	-0.130	0.560	0.310	0.83	-0.066	0.770	0.42	-0.066	0.770	0.42	-0.169	0.690	0.310	0.83
	0.510	10.770		0.510	10.770		-1.710	14.270	11.950		-0.410	8.470		-0.410	8.470		-1.910	16.750	9.050	
Diversified	-0.037	0.790	0.9	-0.037	0.790	0.9	-0.088	0.770	0.080	0.94	-0.085	0.810	0.9	-0.085	0.810	0.9	-0.109	0.790	0.070	0.94
	-1.000	38.150		-1.000	38.150		-4.020	52.490	9.860		-1.890	25.810		-1.890	25.810		-4.130	36.480	6.540	
Corporate	0.016	0.790	0.62	-0.020	0.840	0.95	-0.117	0.750	0.220	0.82	-0.081	0.900	0.67	-0.050	0.860	0.94	-0.140	0.860	0.180	0.82
	0.180	15.390		-1.000	47.310		-2.260	18.980	10.010		-0.710	13.180		-2.500	30.140		-1.800	20.820	4.950	

	July 2007 until August 2009										**September 2009 until July 2016**									
	Aggregate			**Style**			**Risk 2**				**Aggregate**			**Style**			**Risk 2**			
Category	α	**(Agg)**	R^2	α	**(MS)**	R^2	α	**(Agg)**	**(HY)**	R^2	$alpha$	**(Agg)**	R^2	α	**(MS)**	R^2	α	**(Agg)**	**(HY)**	R^2
All	-0.166	0.880	0.76	-0.166	0.880	0.76	-0.237	0.850	0.160	0.94	0.135	0.750	0.86	0.135	0.750	0.86	-0.011	0.750	0.130	0.93
	-0.660	8.410		-0.660	8.410		-2.240	31.850	10.740		2.270	19.600		2.270	19.600		-0.240	26.260	5.430	
High Yield	-0.357	0.390	0.02	-0.650	0.790	0.8	-0.713	0.230	0.790	0.78	0.697	-0.010	0.00	-0.050	0.630	0.84	-0.031	-0.030	0.630	0.84
	-0.280	0.730		-1.440	5.920		-1.630	1.280	6.810		3.100	-0.080		-0.420	10.400		-0.240	-0.680	10.500	
Government	0.194	1.030	0.81	-0.090	1.010	0.97	0.262	1.060	-0.150	0.94	-0.089	1.080	0.91	-0.020	0.990	0.96	0.060	1.080	-0.130	0.94
	0.790	9.180		-3.000	37.910		2.310	21.430	-8.410		-1.830	37.790		-1.000	41.560		1.290	36.620	-4.500	
Flexible	-0.506	0.850	0.33	-0.506	0.850	0.33	-0.679	0.770	0.380	0.85	0.365	0.430	0.38	0.365	0.430	0.38	0.014	0.420	0.300	0.91
	-0.810	3.320		-0.810	3.320		-2.540	9.300	11.550		3.300	7.820		3.300	7.820		0.300	18.960	12.460	
Diversified	-0.167	0.840	0.86	-0.167	0.840	0.86	-0.212	0.810	0.100	0.95	0.055	0.750	0.91	0.055	0.750	0.91	-0.054	0.750	0.090	0.95
	-1.010	10.490		-1.010	10.490		-3.280	24.750	13.840		1.110	24.570		1.110	24.570		-1.180	31.620	5.510	
Corporate	-0.389	0.950	0.53	-0.020	0.820	0.97	-0.516	0.890	0.280	0.87	0.241	0.640	0.56	0.060	0.800	0.96	-0.084	0.630	0.280	0.86
	-0.850	5.050		-0.220	26.900		-2.730	13.070	16.430		2.190	9.290		2.000	27.450		-1.840	14.210	6.990	

Chapter 6

False Discoveries in Fixed Income

This Chapter expands upon the findings in Chapter 5 by seeking to identify if there is evidence of true outperformance, or the funds are just lucky. Various methods are applied here to do so. First, entire-cases resampling as per Fama and French (2010) then two methods to adjust for False Discovery Rates; the "classical" approach of Barras et al. (2010) and the recent empirical advancements as proposed by Ferson and Chen (2019). Suggestions are then explored as to the reasons underlying the results observed – predominantly with regards to the value that corporate bond funds appear to offer in the post-crisis period.

6.1 Introduction

The performance evaluation of the funds conducted in Chapter 5 shows results which are generally consistent with the findings of previous studies – that the funds fail to significantly outperform passive benchmarks on an after-cost basis. However, when funds are grouped into portfolios, this may mask the performance of managers who do in fact exhibit positive risk-adjusted returns. After all, it is most likely to be the case that it is the extreme performers which are of most interest to investors. Evaluating the funds on an individual basis is one step towards ascertaining a clearer picture regarding the extent of outperformance. However, limitations are imposed when a standard OLS procedure is adopted, in part due to the imposition of restrictive parametric assumptions; for example, that the returns are normally distributed. When doing so, the performance in the tails is likely to be overlooked. To make such distinction, a bootstrap simulation will be applied to identify how the actual performance observed relates to that which is hypothetical given the imposition of zero alpha.

Evaluation of performance in such a setting is otherwise known as multiple hypothesis testing. It is unfortunate that this framework is plagued by issues relating to the accuracy of inference that may be drawn from the results. For example, the decision rules for significance testing are applied across multiple funds, when they have in fact been considered with a single hypothesis in mind. In the fund performance literature, broadly speaking there are two approaches commonly applied to appease such issues – the bootstrap and adjustment for false discoveries.

There is an extensive literature that addresses many aspects of the multiple hypothesis testing framework, with constant empirical refinements being applied. However, that which relates to bond funds remains limited. As at the date of writing, to the best of my knowledge, there are no studies that investigate false discoveries in fixed income. The findings from this thesis thus far suggest an interesting and promising setting to further investigate this topic. Chapter 4 highlighted that a combination of factors allows for construction of a well specified model that would induce minimal bias. Chapter 5 identified that there appears to be scope for outperformance in the post crisis period, with the performance of Corporate, Government, and Diversified funds realising improvements here. This final empirical chapter seeks to add conviction to the results obtained throughout and by doing so, make valuable addition to the academic literature on addressing multiple testing issues with bond funds.

This chapter identifies three key findings which make a notable contribution to the existing literature in this area. (1) When applying the bootstrap method of Fama and French (2010) using net excess returns, significant outperformance across all funds is identified in the 97th percentiles. On a category specific basis, the Corporate funds fare the best. These results are more pronounced in the post crisis period from 2009 to 2016, with superior ability evident from the 95th percentile across All and Corporate funds respectively. Results are also presented to test the impact of altering the minimum return requirement. and relative to an alternative model. In both cases, the findings of superior performance are supported.

(2) Two methods of adjusting for False Discovery Rates are applied – the "classical" approach proposed by Barras, Scalliet, and Wermers (2010) (Hereafter BSW (2010)) and that of Ferson and Chen (2019) (Hereafter FC (2019)) which seeks to add further empirical refinements. There is a known bias with the BSW method whereby it lacks power to identify truly outperforming funds, hence classifying the majority as zero alpha. The results in this chapter highlight this aspect from a comparison with those obtained though both methods. In either case, superior performance is found to look most promising in the post-crisis period. Results are presented for the tails as per BSW (2010). Across All funds 22 are found to be positive, 12 of which are due to the Corporate category.

(3) Isolating what is driving this outperformance is both challenging and interesting. Post-crisis, the Corporate bond funds are more highly correlated with the High Yield index than periods otherwise and are to a greater extent than is the Corporate market, as proxied by the Barclays index. During this time, the performance of the High

Yield index was notably superior. This analysis would suggest that there may be evidence of "reaching for yield" behaviours exhibited by the funds. However, to what extent this explains the alpha observed is unclear. Existing literature suggests that any premium gained from such actions is fully explained by risk and is therefore not a determinant of true alpha.

The chapter proceeds as follows: Section 2 discusses related bootstrap literature and methodology, Section 3 outlines the same in the context of false discovery rates, Sections 4 to 6 present the data and empirical results respectively, Section 7 provides a discussion of the results, and Section 7 summarises the conclusions.

6.2 The bootstrap

6.2.1 Existing literature

An increasingly popular method to assess the uncertainty regarding estimation of the parameter of interest is the bootstrap. Statistical techniques involving resampling as such have grown in popularity, partly due to improved computational power which adds to the ease of implementation. A key attraction of this approach is the flexibility it provides. It avoids the requirement of imposing restrictive parametric assumptions with regards to the underlying mathematical or statistical process. For example, the bootstrap procedure does not depend upon the assumption of a normal distribution, nor does it require reliance upon a large sample. Therefore, the bootstrap can be considered a powerful, reliable, and flexible method of estimation for parameters of interest and their associated confidence intervals.

For a given sample size n, in the absence of such parametric assumptions the distribution of the population is unknown. From this sample, the bootstrap allows for the empirical distribution to be constructed by resampling with replacement. Each sample value is assigned a probability of $1/N$. Having derived the empirical distribution, the properties of the estimator are then determined, often by a Monte Carlo style approximation. Here follows a basic illustration of what bootstrapping is trying to achieve:

- Initial sample consists of n independent identically distributed random vectors; $S = \{X_1, X_2, ..., X_n\}$
- From this, the parameter of interest defined as $\hat{\theta}$, is estimated from the values within S
- By applying the bootstrap, it allows for an assessment as to the accuracy of $\hat{\theta}$, in accordance with the empirical distribution function, F_n.

The sample drawn helps to illustrate the relationship between the two sets of data and allows for an empirical estimate of the statistic's sampling distribution to be derived. A benefit of this approach is that it allows the characteristics to be identified from the sample drawn from the specific "population", rather than a wider population of which many may not be included in the study. However, the converse in this case can also apply and is highlighted as a limitation of the approach; i.e., the sample selected is not fully representative of the actual population in question. (Fox 2002)

Various bootstrap methods have been implemented throughout the academic literature on fund performance evaluation. One of the most common is that of Kosowski, Timmermann, Wermers, and White, (2006), hereafter KTWW (2006), which resamples the residuals from OLS regressions randomly with replacement. They apply their method to 1,788 US equity mutual funds, from 1975 to 2002. Their findings show that there is evidence of skill, most concentrated in growth-orientated managers. Many papers have since followed suit.

To apply the KTWW (2006) method for fund performance evaluation, estimates of alphas, factor loadings, and residuals are first derived from regressions of the individual funds' monthly excess returns against the factor model returns. Pseudo excess returns are created for each fund. Residuals are then randomly resampled with replacement from $\left\{\varepsilon_{i,t} t = T_{i,0}.....T_{i,1}\right\}$, and the null of zero abnormal performance is imposed:

$$(R_{i,t} - r_{f,t})^b \equiv \hat{\beta}_i(K_t).... + \hat{\varepsilon}^b_{i,t} \tag{6.1}$$

where $b = the\ b^{th}$ bootstrap iteration, $\hat{\varepsilon}_{it}$ = a drawing from the residuals estimated for each fund in the initial regression. The imposition of zero abnormal performance means that the alphas generated for this distribution can be considered "lucky". By doing so, it allows for a comparison of results of the actual sample in question, relative to what would be expected given the absence of skill. This method of resampling maintains the historical ordering of the explanatory variables. Then, for each bootstrap iteration, [b = *1,....10,000*] regressions are conducted using the zero alpha pseudo excess returns regressed against the factor model:

$$(R_{i,t} - r_{f,t})^b = \alpha_i + \beta_i(K_t).... + \hat{\varepsilon}_{i,t} \tag{6.2}$$

The alphas are saved, and the process is repeated for each fund in the sample, $i = 1, ...N$. Thus, the cross-section of bootstrapped alphas and corresponding t-stats from the sample N is derived; $\left\{\hat{\alpha}^b_i, i = 1, ...N; b = 1\right\}$, $\left\{t(\hat{\alpha}^b_i), i = 1, ...N; b = 1\right\}$ The next step is to repeat for all bootstrap iterations, [b = *1,....10,000*]. The final output from this procedure is an estimation of the cross-sectional distribution of alphas that result, given the imposition zero alpha. Therefore, this may alternatively be described as the "lucky" or "spurious" distribution and is the basis upon which the skill of the actual performance will be judged. To do so, the corresponding t-stats will be ranked from smallest to largest for both distributions and the proportion of relative outperformance at each percentile may be observed. For example, for the actual fund performance to be judged as skilful, this distribution requires to have a greater number of positive and significant t-stats at the relevant percentile than what results for the "lucky" distribution.

Here, in the method of KTWW (2006) only the residuals are resampled; the factors in this case remain time-invariant. With regards to the application to fixed income funds, this could be factor that renders this method inferior than that of Fama and French (2010). For example, given the significance of systematic factors to the performance of bonds, the time dependency is a more important consideration to make here than is the case with equities. An example of this time-ordering process is outlined in the following section.

Limitations with the KTWW (2006) method are identified by Fama and French (2010), who present an alternative known as entire-cases resampling. The KTWW (2006) bootstrap resamples with replacement the residuals of each individual fund. On the other hand, Fama and French (2010) resample with replacement over the full cross-section of returns. This gives rise to one of the key differences between the two approaches: the time ordering of the funds and risk factors. Fama and French (2010) note that an advantage of their approach is that due to the joint resampling of fund and factor returns it allows for correlated heteroskedasticity of the explanatory returns and disturbances of a benchmark model to be maintained. The details of this methodology are discussed in more detail in the subsequent section (6.2.2)

Fama and French (2010) apply this to a sample of US equity mutual funds over the period 1984-2006. They position their findings in the context of equilibrium accounting, whereby active management is considered a zero-sum game. By using both gross and net returns, they assess the extent to which costs are value reducing for the investors. An initial regression analysis using value-weighted and equally-weighted portfolios of funds shows that the actively managed funds exhibit neutral performance prior to deduction of fees. To add further conviction to the results, they next apply the simulation to 12 cross-sections: funds grouped into three assets under management (AUM) categories, evaluated using three and a four factor models, for both net and gross returns. Their findings show minimal evidence of significant positive alpha on an after-cost basis.

Extending the analysis outside of the US, Cuthbertson et al. (2008) use the KTWW (2006) method to conduct an evaluation of fund manager skill using UK data.

Likewise, however, this only includes equity mutual funds. They find some evidence of stock picking ability, but only in a few of the top performing funds. The poorest performing funds are found not just to be unlucky, but exhibit "bad skill". Furthermore, this performance is shown to persist, which is not the case for the small number of top performers. Ability is detected depending upon the style and domicile of the fund; "small stock" and "all company" funds do not show ability. In terms of the domicile, onshore funds are those for which skill is identified, but this is not the case for offshore.

Kosowski et al. (2007) conduct an evaluation of hedge funds using the KTWW (2006) bootstrap procedure. They seek to identify if there is a difference between the alphas generated using OLS regressions, relative to Bayesian performance measures. The latter is particularly applicable to a study of hedge funds as it helps to overcome some of the biases that are likely to be prevalent in the return data; such as incubation bias, backfill bias, serial correlation, or issues with only short-return histories being available. Hedge funds are found to exhibit enough skill to generate returns in excess of the multi-factor benchmark model. Furthermore, the Bayesian alphas are considered more reliable in this case than those of the OLS approach. It is highlighted by the authors that the rankings based on OLS alphas may in fact overstate the performance.

The application of bootstrapping extends across asset classes and investment vehicle; the flexibility that the approach provides allows for this to be the case, However, that which relates to bond funds remains minimal. Evidence from the Canadian market shows the difference that may be observed when using portfolios of funds relative to considering the sample on an individual basis. Ayadi and Kryzanowski (2011) form equally-weighted and size-weighted portfolios of 303 Canadian bond funds from 1984-2003. The results initially show that the funds fail to outperform a six-factor model on a net basis. However, a block bootstrap is next applied to the individual funds which yields marginally better performance. Here, there is evidence that the funds are able to return value to investors after costs.

Although the use of the bootstrap is commonly associated with the estimation of alpha, given the flexibility of its approach, it may indeed be applied when assessing various aspects of performance. For example, (Y. Chen et al. 2010), evaluate the timing ability of bond mutual funds and when doing so use the method of KTWW (2006) to add robustness to their estimates of the timing coefficients. The distribution is derived having imposed the restriction of zero timing ability by setting timing coefficient equal to zero. Again, with regards to assessing the timing ability of bond fund managers, Huang and Wang (2014) also use the KTWW (2006) method to add robustness tests to their results. An opportunity is presented for further research in this area. With no studies as yet applying the entire-cases method to bond funds, it makes for an interesting setting to do so.

Within the sample, there exists a degree of heterogeneity across the individual funds. For example, considering a comparison between Government bond funds and High Yield funds. Bootstrapping accommodates the complex, non-normal distribution of the cross-section. However, despite the many benefits, there remains a limitation. Although the bootstrap method helps to identify whether extreme performance exists, it does not help with how these funds may be selected ex ante, or what the associated characteristics of these funds may be. Nonetheless, if certain groups of funds are found to exhibit significant performance in the tails of the distribution, it helps to highlight a cross-section that is perhaps of most interest for further analysis as to characteristics driving the results.

Overall, one method of the bootstrap does not dominate within the literature. Various aspects of the data and research objectives must be considered with regards to each application. This presents a rich field of opportunity for further refinements and out-of-sample testing using new data. The purpose of this empirical chapter is to do just that. As noted, funds have been evaluated using bootstrap methods in the UK market. However, this is specific to equity funds and their bond counterparts are yet to be assessed.

6.2.2 Bootstrap methodology

The bootstrap approach used throughout the remainder of this chapter is that of Fama and French (2010), which is known as entire-cases resampling. This is proposed as an alternative to the residual only resampling approach of KTWW (2006). Again, it uses time-series regression to calculate alpha for each fund, however, the resampling is of a different form. An example of time ordering as follows:

- For bootstrap iteration $b = 1$ for the month in which this is taken, e.g. $t = 15$ zero abnormal performance is imposed for each fund i that exists in the sample at $t = 15$ creating a set of pseudo excess returns.
- The pseudo returns in this case would be calculated as: $[R_{i,15} - r_{f,15}]$ - $\alpha_{i,15}$, e.g. subtracting the fund specific alpha from the fund's excess return in month 15.
- This is repeated for $b = 10,000$ across all months in the sample period, allowing for an average estimate to be derived across the time-series; $[R_{i,1...211} - r_{f,1...211}]$ - $\alpha_{i,1...211}$
- Resampling the residuals as in KTWW (2006) does not allow for this. In each iteration, e.g. $b = 1$ drawings for each fund may relate to different time periods, e.g. the first drawing for fund 1 might be $t = 10$, whereas for fund 2, the first drawing could instead be $t = 20$

A key difference due to the time ordering is that the cross-section of fund returns is captured by the FF (2010) approach, and thus maintains the correlation structure. This is an appealing characteristic of this method. However, there are other aspects of the test design that are worth noting as potential drivers of the results observed, such as the required minimum number of observations per fund. Given that the return history will vary per fund, the choice of this means that some funds will feature more or less than others. In the original papers, FF (2010) require a minimum of 8 monthly observations, whereas in KTWW (2006) it is 36. Difficulties are present if making a direct comparison using the original papers alone, as the samples and various parameters are not exactly the same. C. Harvey and Liu (2020) control for this in order to isolate the effects of this variation. They find that due to the low minimum return requirement, the FF (2010) approach has a bias towards under-sampling; the consequence of which being a higher likelihood of failing to reject the null hypothesis of zero abnormal performance. The opposite is true of KTWW (2006).

Such findings are consistent with the assertion of FF (2010) that the KTWW (2006) method induces a bias towards identifying positive performance. This was also identified by D. Blake et al. (2017) when they applied both to UK equity mutual funds, aiming to identify if each method results in different conclusions regarding fund manager skill and value returned to investors. When net returns are used, neither bootstrap results in evidence of outperformance. However, on a before-cost basis, the method of KTWW (2006) shows that 95% of fund managers do exhibit enough skill to beat the passive benchmarks.

Overall, Harvey and Liu (2020) conclude that the FF (2010) method is optimal, however, the "truth" lies somewhere between the minimum requirement of 8 and 36 observations. Thus, the selection used throughout the remainder of this thesis is consistent with this recommendation; a minimum of 24 monthly returns will be required, with robustness tests conducted where appropriate. It is unfortunate to an extent that there is a lack of similar literature having already applied the FF (2010) method to bond funds, and hence provide a reference point for comparison. Nonetheless, it otherwise presents a case to contribute new findings to the existing literature by being the first to do so. A potential area for further research would be to conduct a comparison of alternative bootstrap methods on a new sample of bond funds. However, the focus remains on the FF (2010) approach in this thesis.

The advancements in statistical inference afforded by bootstrapping have certainly allowed for greater insights and robustness checks to be derived regarding the prevalence of alpha. However, it is relatively limited with regards to the granularity it provides in terms of identifying characteristics of funds generating any superior performance observed. Furthermore, an adjustment for the full extent of luck that may be biasing the observations remains outstanding. Overall, the bootstrap is aiming to identify if any abnormal performance exists; however, additional tests that allow for greater quantification as to the number of funds that this may be attributable prove to be valuable. Two methods that allow for this are discussed throughout the subsequent section.

6.3 False discovery rates

6.3.1 Existing literature

In the context of fund performance evaluation, the main issue that still prevails is the accuracy with which the proportion of superior performance on a risk-adjusted basis may be judged. Various papers have been credited with the development of the False Discovery Rate (FDR) approach. (Benjamin and Hochberg (1995), (Storey 2002) The distinguishing feature is the ability of this statistical method to control for the issue(s) that result from multiple testing. An attraction of the method is the relative simplicity that it provides to the researcher.

In a seminal paper, Barras, Scalliet, and Wermers (2010) (hereafter BSW(2010)) apply the method of Storey (2002) to a sample of US equity mutual funds between January 1975 and December (2006). This allows the population of actively managed funds to be divided into groups defined as zero alpha funds π_0, unskilled funds with negative alphas π_-, and skilled funds π_+ with positive alphas. Overall, the results are disappointing, with estimates of $\pi_0 = 75.4\%$, $\pi_- = 24\%$, and $\pi_+ = 0.6\%$ across the whole sample period. Unskilled funds dominate the underperformance, i.e. this is not just due to bad luck. On the other hand, there is very minimal evidence of true skill among the funds. Almost all the outperformance can be attributed to luck. There is a very notable degree of time-variation observed in this respect. The authors find that many more funds outperform in the earlier years of their sample. However, this diminishes to almost zero by the end in 2006. Reasons behind this are not explored and concluded in the paper, although various are suggested. The authors also consider usefulness of the FDR approach from the perspective of implementing a dynamic trading strategy. To do so, they set the FDR to a target proportion (e.g. 10%) and allocate funds to a portfolio that is rebalanced every year based on this criteria.

Following BSW (2010), Cuthbertson et al. (2012) implement the FDR method to a sample of all UK equity mutual funds that existed between April 1975 and December (2002). The funds were filtered to require 36 monthly observations, resulting in a total of 675 being included. Three types of models were used: unconditional (FF (1993) and Carhart (1997)), conditional-beta (FF3F), and conditional-alpha-beta (FF3F). The results from the unconditional models are presented. They find slightly more evidence of skilled funds than was the case in the US, with truly positive alphas being recognised in 1.6% of the sample. However, unskilled funds again dominate. Further to this, the authors also examine performance persistence and whether portfolio formation inspired by FDR results proves to add value. Subject to a target FDR of 10% and 20%, portfolios are formed using overlapping 5 year periods. On this basis, positive persistence is not identified as the forward-looking alphas are not statistically significant.

Performance evaluation results from the false discovery rate methodology tend to find minimal evidence of true outperformance. The studies as discussed thus far are focused on developed markets. Koutmos et al. (2020) apply the FDR approach of BSW (2010) to mutual funds invested in the Chinese stock market. In addition to evaluating the performance of fund managers, the authors extend their analysis to also assess the fund management company. Overall, their results show that 7% of their sample of Chinese mutual funds display true selection ability, which is considered high relative to their developed-world counterparts. As such, this provides further evidence of market inefficiencies and highlights there may be greater opportunity for exploiting informational advantages.

Recent literature has sought to further refine the FDR method; predominantly with regards to the power of testing that it provides. The Type II error rate is of concern here. This addresses the issue of not selecting a skilled fund as it has been incorrectly classified as unskilled - in other words, a false negative. This aspect of error rate quantification has received little coverage relative to false positives (Type I error rate). A recent critique of the BSW (2010) approach was proposed by Andrikogiannopoulou and Papakonstantinou (2019), hereafter AP (2019) Their findings contest that the statistical power of the BSW (2010) is low, and therefore biases the proportion of nonzero-alpha funds identified. As noted, the original paper finds that 75% of funds are zero alpha.The bias is due to either small alphas or due to a large amount of estimation noise.

The authors (BSW 2010) defend their initial method. In response, they update various parameter values and re-visit their empirical analysis. Using a sequential approach, they test the marginal impact that various amendments have on the results observed. Changes they apply include: use the median as opposed to mean residual volatility, account for the relation between parameters, and increase the interval from 0.5 to 0.95. BSW recognise the usefulness of the challenge by AP (2019), however, they do not find the bias as noted to be as substantial as proposed. On the basis of their analysis, they conclude that the FDR method is still appropriate for mutual fund evaluation, despite the suggestion otherwise by AP(2019).

An important point which is highlighted by BSW is that there is a trade-off with regards to complexity and accuracy, when selecting the method to minimise multiple testing issues. BSW also note that their method used to contest AP (2019) is not prone to the same errors as is the case for more parametric approach in which the shape of the distribution is specified. For example, the selection of positive/negative alpha values. The reply is not reliant upon simulations. BSW use a simple analytical approach, using the student t-distributions. A benefit of this is that it does not require all the parameters in the data generating process to be specified. Therefore, the method is free from noise as would otherwise be inherent in more complex applications.

The refinements and critiques of the FDR methodology as outlined thus far relate to adjusting the original set-up. For example, redefining the inherent parameters. Alternatively, recent research has sought to adapt the empirical methods using a parametric approach. Andrikogiannopoulou and Papakonstantinou (2016) Ferson and Chen (2019) attempt to overcome two limitations imposed by the BSW (2010) methodology. The first, like AP (2019), relates to the power of the test. FC (2019) allow this to be below 100%, i.e., it is possible to make an incorrect decision regarding the rejection (or not) of the null hypothesis. Additionally, the empirical method is further refined to incorporate confusion parameters. For example, when a good fund is classified as a bad fund and vice versa. BSW (2010) by default assign a confusion value of 0%, whereas FC (2019) allow this to be greater than zero. The authors use a bootstrap simulation, based on FF (2010). (methodology discussed in detail in Section 6.2.2). Their findings are consistent with BSW with regards to the lack of truly good funds identified. However, having allowed for imperfect power, the proportion of zero alpha funds identified is greatly reduced in favour of a higher proportion of truly underperforming funds.

FC (2019) also study FDRs in a sample of hedge funds. Characteristics observed for the hedge fund returns are noted to make them worthy candidates for the application of this method. This is an alternative method to an earlier study of hedge funds. Y. Chen et al. (2017) present a new method with which to assess skill. When conducting assessments of managerial skill, the most common approach is to adjust for luck among the skilled funds. However, the approach of this paper also considers the luck of the unskilled funds. As such, the power of testing is considered enhanced. Their findings show that the performance of the sample of hedge funds can be grouped accordingly: 9% as excellent, 38% good, 43% neutral, and 9% bad. These results provide quite a contrast to those of mutual funds, whereby evidence of skill is minimal.

Overall, the False Discovery Rate method does help to address some of the issues identified with the bootstrap. For example, the extent to which performance in the tails of the distribution can be attributed to luck makes for a more insightful analysis.

An important advantage that it provides is the ability to narrow down characteristics of the extreme performing funds. For example, highlighting the characteristics synonymous with the left and right tails, or using the associated FDR as a control variable in portfolio formation. With this field of literature continuing to expand, the lack of bond specific research is again highlighted as an interesting area for many further developments.

6.3.2 Barras, Scalliet, and Wermers (2010) methodology

When applying the FDR method, a general perspective is adopted. As opposed to isolating individual funds in the population, the objectives centre around grouping the funds according to performance proportions; π_0, π_b, and π_g. By doing so, it attempts to counter the limitations imposed from the multiple hypothesis testing framework. Such issues arise when hypotheses designed to be applied to a single observation are used to conduct inferences across multiple. An example of which would be to count the number of high/significant alphas from the individual funds in a population, as a means by which to obtain conclusions regarding the extent of true skill among them. To improve accuracy of the insights gained, the extent of luck must be adjusted for - i.e., take account of the false discoveries. These may be defined as the incidents whereby the null hypothesis of zero alpha has been rejected in favour of identifying skill, when in fact the null is true.

For example, the proportion of significant positive alpha funds, $E(S_{\gamma}^{+})$, must be adjusted to reflect the expected proportion of lucky funds, as denoted $E(F_{\gamma}^{+})$. The proportion of lucky funds is defined as follows:

$$E(F_{\gamma}^{+}) = \pi_0 \,.\, \gamma/2 \tag{6.3}$$

whereby the fraction of zero-alpha funds α_0 in the sample is multiplied by the probability in each tail $\gamma/2$. Therefore, an estimate of the extent of true outperformance is derived from:

$$E(T_{\gamma}^{+}) = E(S_{\gamma}^{+}) - E(F_{\gamma}^{+}) = E(S_{\gamma}^{+}) - \pi_0 \, . \, \gamma/2 \tag{6.4}$$

For unlucky funds the same adjustment applies, but to the left tail:

$$E(T_{\gamma}^{-}) = E(S_{\gamma}^{-}) - E(F_{\gamma}^{-}) = E(S_{\gamma}^{-}) - \pi_0 \, . \, \gamma/2 \tag{6.5}$$

The researcher is afforded flexibility when selecting the significance level, γ. For higher specifications, the closer to the true values of unskilled and skilled funds the estimates converge. In this case, the Type II error rate is minimised at the expense of the Type I rate. Therefore, if the objective is to gain insights across the whole population, then large values of γ make sense. On the contrary, if the focus primarily relates to the more extreme locations in the tails, then a small test size would be sufficient. By varying γ insights may be derived as to the location of the truly skilled funds. The more concentrated in the right tail the truly skilled funds are, the smaller the increase in $E(T_{\gamma}^{+})$ because the others that were appearing as also skilled, as denoted $E(S_{\gamma}^{+})$, would in fact be lucky. The converse also applies.

A further appealing characteristic of the FDR approach is that it is relatively simple. For example, in order to estimate the central parameter π_0, the p-value of the individual estimated alphas are required. BSW use a simulation method to estimate optimal lambda λ^*, based on the method of Storey (2002). The authors note, however, that the results are not particularly sensitive to the value selected. For example, BSW (2010) find that minimal difference in the estimate of $\hat{\pi}_0$ is observed when an intermediate value of lambda is selected, such as 0.5 or 0.6. Therefore, estimating the optimal lambda by means of simulation is not essential for an effective test design. Pre-selected values are also appropriate.

$$\hat{\pi}_0(\lambda^*) = \frac{\hat{W}(\lambda^*)}{M} \quad . \quad \frac{1}{(1 \, - \, \lambda^*)'} \tag{6.6}$$

Here, $\hat{W}(\lambda^*)$ denotes the number of funds with p-values exceeding λ^*. Thus, having estimated $\hat{\pi}_0$, the proportions of unlucky and lucky funds in the left and right tails respectively can be easily estimated:

$$\hat{F}_{\gamma}^{-} = \hat{F}_{\gamma}^{+} = \hat{\pi}_0 \, . \, \gamma/2 \tag{6.7}$$

The significant performance observed can then be adjusted to derive the proportions of unskilled and skilled funds at a given level of significance:

$$\begin{aligned} \hat{T}_{\gamma}^{-} &= \hat{S}_{\gamma}^{-} - \hat{F}_{\gamma}^{-} = \hat{S}_{\gamma}^{-} - \hat{\pi}_0 \, . \, \gamma/2 \\ \hat{T}_{\gamma}^{+} &= \hat{S}_{\gamma}^{+} - \hat{F}_{\gamma}^{+} = \hat{S}_{\gamma}^{+} - \hat{\pi}_0 \, . \, \gamma/2 \end{aligned} \tag{6.8}$$

By solving these equations we can estimate the proportion of skilled and unskilled funds in the population from:

$$\hat{\pi}_A^{-} = \hat{T}_{\gamma^*}^{-}, \quad \hat{\pi}_A^{+} = \hat{T}_{\gamma^*}^{+}, \tag{6.9}$$

To allow estimates to be more representative of the whole sample, a relatively high level of significance (test size) should be selected. Again, a bootstrap may be used to estimate the optimal value subject to minimising the estimated MSE of the true proportions of unskilled and skilled funds, i.e. $\hat{\pi}_A^{-}$ and $\hat{\pi}_A^{+}$ Values of between 0.35 and 0.40 are found to be optimal in BSW (2010), i.e., the estimates are close to those as per MSE.

6.3.3 Ferson and Chen (2019) methodology

The FC (2019) method is based upon this same approach and therefore allows an intuitive extension for further analysis. Here, rather than only impose the restriction that funds with no skill have zero alpha, FC (2019) draw further insights as to the locations of outperforming funds. They do this by conducting three simulations – deriving not only a no skill distribution (zero alpha), but also those to represent what would be expected under "good" and "bad" performance. The procedure to do so is as follows:

1. The bootstrap methodology as per section 6.2.2 applies. This first simulation derives the no skill distribution by setting the alpha value to zero. For a given significance level, e.g. 10% in each tail, it can be inferred at what location the funds could be determined as "good" or "bad". If $\gamma/2$ is 10% in each tail, then t_g is the value above which 10% of the simulated t-statistics lie and t_b is the value below which 10% of the simulated statistics lie. The α_g and α_b values correspond to these critical values.

2. Having identified the α_g value, the same simulation method is applied but instead of forcing alpha of the distribution to zero, α_g is added back to each fund. In other words, this is setting a lower bound at which a fund can be considered "good". The fraction of simulated t-statistics above is the power of the test for good funds β_g. The fraction of simulated t-statistics below t_b is the probability of confusing a good fund with a bad fund δ_g.

3. The third simulation generates the distribution under the imposition of bad performance. This time, the alpha of each fund is negatively adjusted by α_b. The fraction of simulated t-statistics below t_b is the power of the test for bad funds β_b. The fraction of simulated t-statistics above t_g is the probability of confusing a bad fund with a good fund δ_b.

$$\begin{aligned} E(F_g) &= P(\text{reject at } t_g|H_0)\pi_0 + P(\text{reject at } t_g|\text{Bad})\pi_b + P(\text{reject at } t_g|Good)\pi_g \\ &= (\gamma/2)\pi_0 + \delta_g\pi_b + \beta_g\pi_g, \text{and similarly,} \end{aligned} \tag{6.10}$$

$$\begin{aligned} E(F_b) &= P(\text{reject at } t_b|H_0)\pi_0 + P(\text{reject at } t_b|\text{Bad})\pi_b + P(\text{reject at } t_b|Good)\pi_g \\ &= (\gamma/2)\pi_0 + \beta_b\pi_b + \delta_b\pi_g. \end{aligned} \tag{6.11}$$

Find the unknown values of π_b and π_g from the given values of $\beta_g, \beta_b, \delta_g, \delta_b$ "The solution to this problem is found numerically by minimizing the errors of these equations subject to the Kuhn-Tucker conditions for the constraints that $\pi_{\geq}0$, $\pi_g \geq 0$ and $\pi_b + \pi_g \leq 0$. The parameters $E(F_g)$ and $E(F_b)$ are estimated by the fractions rejected in the actual data and the parameters $[\beta_g, \beta_b, \delta_g, \delta_b]$ are calibrated from the simulations.[1]

A criticism of the BSW method is the assumption of perfect power. When $\beta_g = \beta_b = 1$there is an upward bias in the number of zero-skill funds identified. In the case of the confusion parameters, if it is assumed that these equal zero then there is a downwards bias on the number of zero funds observed. However, the impact of the power parameter is found to dominate. The authors note that the impact of varying these parameters has very little impact on the overall results. Nonetheless, they find that for a large enough number of simulation trials, the values can be estimated with a minimal error. This highlights the difference relative to the classical approach - the incorporation of power and confusion parameters. If each of these was assumed to be perfect, then the calculation reduces to the classical estimation.

Another advancement proposed in the paper is to use a grid search to optimally locate the alpha values. This is outside the scope of this thesis but is acknowledged as a potential refinement in future work. The results observed are interesting to note. FC find that the α_g value obtained from the grid search is far below that of the pre-selected value of 0.317%, it is in fact -0.034%. This shows that even the "good" funds have a negative lower bound of performance. As such, FC redefine the classifications of funds in their paper to be "good" α_0, "bad" $\alpha_{<0}$, and "ugly" $\alpha_{<<0}$.[2]

The pre-selected alphas that define the locations correspond to those as per BSW (2010). For example, to be classified as significant at 10% in each tail $t_+ = 1.65$ and t_1 =-1.65 the corresponding alpha values are α_g= 0.317% and α_b= -0.267%. By a process of interpolation it is evident that this also fits the distribution of FC. However,

[1] The Matlab function fmincon has been used for this purpose

[2] Whereby there are no positive alpha funds identified. Good would be classified as zero alpha, with bad and ugly defining those progressively more negative.

the same alpha values are applied to the hedge fund data. Here, they do not correspond to 10%.

An attraction of the BSW (2010) classical method is the relative simplicity and ease of application that it provides. There has been a wealth of literature recently directed towards addressing the issues inherent in a multiple hypothesis testing framework. As a result, many empirical refinements have been presented. The application of FC (2019) FDR methodology is a natural progression to add further robustness to the results observed from the FF (2010) simulation. However, the lack of bond fund performance literature in this area must be acknowledged. With this in mind, it has been considered logical to first present results for the BSW (2010) method. There is a rich field of open research questions to be explained with regards to bond funds; the initial application of both these methods is the fist step to do so.

6.4 Data description

6.4.1 Individual fund returns

The same data set will be used throughout the remainder of this chapter, for both the bootstrap and false discovery rate methods. An adjustment has been made, however, relative to that which was used in Chapter 5. The minimum return requirement has been increased from 12 to 24 for each fund in the sample. By doing so, a number of funds have been lost; only three over the whole period, and sixty-nine in the most recent.

Descriptive statistics for the individual fund returns are presented in Table 6.1 for the two periods under analysis; the whole period from January 1999 until July 2016 and the recent sub-period from September 2009 until July 2016. Given the results from Chapter 5 whereby the improved performance was observed for the Corporate, Diversified, and Government categories, the initial analysis in this chapter begins with them. The

High Yield and Flexible funds comprise only a minimal proportion of the sample, therefore limiting the reliability of results for these specific categories. However, they have still been included in results pertaining to All funds.

Table 6.1: Descriptive Statistics of Sample

This table provides descriptive statistics of the individual funds used in the simulation. A minimum of 24 monthly return observations are required for each fund. N refers to the number of funds in each category, Obs is the average number of monthly return observations for each fund. The mean, Std (standard deviation) min, and max values are all monthly and in percent (%). The average skewness and kurtosis values are also presented as are the proportions of significant positive and negative alphas when evaluated using the Maturity 5 model

1999 - 2016	**N**	**Obs**	**Mean**	**Std**	**Min**	**Max**	**Skew**	**Kurt**	**Mean**	**Median**	**Pos**	**Neg**
Corp.	144	137	0.267	1.710	-5.708	5.205	-0.221	4.897	-0.051	-0.044	4.00	30.00
Gov.	79	144	0.295	1.863	-5.313	6.292	0.407	4.859	-0.042	-0.029	4.00	33.00
Div.	115	125	0.148	1.741	-6.495	4.870	-0.406	7.195	-0.106	-0.070	2.00	43.00
All	369	135	0.242	1.798	-6.374	5.531	-0.202	6.039	-0.066	-0.045	4.00	32.00
2009 - 2016												
Corp.	121	72	0.591	1.577	-3.924	4.528	-0.110	3.650	-0.007	-0.011	10.00	12.00
Gov.	70	73	0.513	1.935	-4.091	6.177	0.353	4.217	-0.019	-0.011	7.00	14.00
Div.	85	68	0.482	1.683	-4.774	4.459	-0.187	4.640	-0.033	-0.025	11.00	24.00
All	304	71	0.548	1.717	-4.258	4.993	-0.035	4.102	-0.015	-0.012	9.00	15.00

Diversified funds show the lowest mean (0.148%) and Government the highest mean return (0.295%) and highest standard deviation (1.863%). The Diversified funds are the most negatively skewed (-0.406) and exhibit the highest kurtosis. Government funds in contrast have a positive skew (0.407). With the kurtosis of a normal distribution equal to three, the evidence of excess kurtosis across the sample provides further indication of non-normal return distributions. As such, the use of standard OLS regression to determine the extent of alpha across the sample is likely to lead to unreliable results in terms of identifying the extreme performance. Given the flexibility that the bootstrap procedure affords with respect to relaxing distributional assumptions, its application is warranted to help examine the performance in the tails.

The most notable observation when comparing across periods is that the funds are less skewed post-crisis. Diversified funds are again the most negatively skewed by approximately half the magnitude (-0.19), and again only the Government category shows a positive skew. Every category exhibits lower kurtosis, with the Diversified funds realising the greatest reduction. Thus, from September 2009 until July 2016, it appears that the sample of funds are less likely to experience large losses, given the lower skewness and kurtosis across the categories.

Throughout this chapter, the results from evaluating the performance of the individual funds will be discussed in greater detail. Nonetheless, some initial observations are worth mentioning here. The funds have been evaluated using the Maturity 5 model. Table 6.1 presents the mean and median alphas for each category, with the proportion of significant (at 5%) positive and negative alphas also shown. Consistent with the skew and kurtosis statistics as noted, the mean (-0.106%) and median (-0.070%) alpha for the Diversified category are the lowest during 1999-2016, with the highest proportion of significant negative alphas and the lowest of positive alphas observed; 43% and 2% respectively. Although the other funds exhibit marginally more appealing alpha characteristics, there is still minimal evidence of any superior outperformance. For example, only 4% of the Corporate category (approx. 6 funds) appear to be outperform, with almost a third underperforming (approx. 43 funds). The post-crisis results are more promising in terms of these proportions. Improvements are apparent across all categories from both a reduction in the left tail and an increase in the right, most notably so for the Diversified funds. However, this method of concluding the prevalence of skill is an example of the multiple testing framework that the subsequent analysis seeks to address.

In terms of comparable studies, (Ayadi and Kryzanowski 2011) show that the average mean return across the whole sample (N=303) is 0.550%, the standard deviation is 0.275%, minimum is -1.217%, maximum is 2.236%, skewness is -0.240, and kurtosis is 11.20. This sample incorporates data from 1984 until 2003, across categories defined as Canadian bond (N=216), Short-term Canadian bond (N=38), and High Yield bond (N=29). Similarities can be seen between the Canadian and UK samples. For example, the Canadian Government bonds have a positive skew while the other categories have negative. However, the standard deviation across the whole sample is much lower at only 0.275%. The standard deviation of UK mutual bond funds was also higher than other European countries, as identified by da Silva et al (2003); for example UK All funds was

1.437% compared to the next highest which was 0.833% for Germany. The UK standard deviation is also higher than the 1.20% in US samples as per Ferson et al (2006). When evaluating the Canadian bond funds Ayadi & Kryzanowski, (2011) split the data into three subsamples – 1984 to 2003, 1984 to 1993, and 1994 to 2003. However, there is no mention of why these periods were chosen or the characteristics of the environment at such times. This is unfortunate as it would have been useful to compare with the basis upon which the sub-samples have been selected for the UK market in this study.

6.4.2 Individual fund alphas

Table 6.2 presents the alphas and associated t-stats for the categories of funds considered. Only selected percentiles are shown here, with a focus on the distribution tails. Alphas are presented monthly and in percent from an evaluation of the individual funds using the Maturity 5 model.

Table 6.2: Alpha Distributions

Alpha distributions for the categories of funds evaluated using the Maturity 5 model are presented in this table. The category All also includes the High Yield and Flexible funds as per the previous chapter, however, the number of these is insufficient to consider on a category-specific basis. The alpha values are monthly and in percent. Results from two periods have been presented; 1999 to 2016, and post-crisis 2009 to 2016

1999 - 2016								
	α_{All}	t_{All}	α_{Corp}	t_{Corp}	α_{Div}	t_{Div}	α_{Gov}	t_{Gov}
1	-0.783	-5.790	-0.420	-5.470	-0.993	-8.290	-0.666	-5.610
5	-0.263	-4.600	-0.226	-5.010	-0.395	-4.660	-0.201	-4.120
10	-0.200	-3.750	-0.164	-3.460	-0.263	-4.100	-0.122	-3.130
90	0.046	0.690	0.041	0.740	0.045	0.630	0.032	0.820
95	0.088	1.220	0.095	1.230	0.085	0.890	0.067	1.460
99	0.230	3.370	0.188	4.820	0.261	3.210	0.246	2.670
2009 - 2016								
	α_{All}	t_{All}	α_{Corp}	t_{Corp}	α_{Div}	t_{Div}	α_{Gov}	t_{Gov}
1	-0.463	-5.970	-0.318	-4.080	-0.778	-7.190	-0.701	-3.520
5	-0.191	-2.810	-0.162	-2.270	-0.241	-4.710	-0.118	-2.860
10	-0.124	-2.010	-0.097	-1.470	-0.176	-2.780	-0.075	-1.670
90	0.103	1.240	0.095	1.430	0.129	1.450	0.070	0.920
95	0.150	1.720	0.124	1.820	0.206	2.300	0.109	1.120
99	0.355	3.400	0.345	4.900	0.508	3.230	0.258	1.480

The alpha distributions play an important role in the application of the FC (2019) method that follows in section 6.6. As noted, FC (2019) use the pre-defined alpha values from BSW (2010), which in the original paper correspond to the locations at +1.65 and -1.65 (i.e. 10% level of significance); 0.317% and -0.267% on a monthly basis. [3] .

It can be seen here the importance of considering the sample-specific properties to minimise bias or error that may be induced in the parametric method. For example, if a "good" fund were to be defined by the alpha value of 0.317%, this would in fact fall out of scope of any of these alpha distributions that are presented for the whole period. Considering the post-crisis period, where the bond fund alphas are indeed higher, an alpha of 0.317% as per BSW and FC would only relate to the very top performing quartile. By using a method of interpolation the alpha values corresponding to the location at -1.65 and +1.65 can be derived. If considering the whole funds, this would equate to 0.12% and -0.07%.

6.5 Empirical Results - Bootstrap

This section presents the results from the entire-cases resampling method of Fama and French (2010). Although the alphas have been discussed as above, the t-stat of alpha is more appropriate to use when making judgements regarding the prevalence of any outperformance. This is due to certain data characteristics that may make funds appear to have larger alphas, but incorrectly so. For example, those with short return histories or those with larger tracking errors. The t-stat of alpha controls for different residual volatilities and numbers of observations.(KTWW 2006) The results of the individual alphas in Tables 6.1 and 6.1 indicate that there is scope for outperformance. The use of the bootstrap will allow for more robust conclusions to be drawn as by comparing the actual performance to the "lucky" distribution.

[3] As noted, the alternative method that is presented in FC uses a grid search to locate the optimal alpha parameters, however, this is left as an option for future research.

Table 6.3 presents the results of the bootstrap from January 1999 until July 2016, using both net and gross excess returns. Table 6.4 presents the same but for the post-crisis period from 2009 to 2016. All individual funds have been evaluated using the Maturity 5 model. Category-specific results are presented for the Corporate, Diversified, and Government funds.

6.5.1 January 1999 until July 2016

Performance is poor when considering All the funds in the sample. In the far left of the distribution, the actual fund returns always underperform the simulated results. It is not until the 98th and 99th percentiles where there appears to be any evidence of significant outperformance. The t-stats of actual and simulated returns are very close in the 98th percentile and the likelihood statistic shows that in almost half of the simulation runs the actual performance is as large as the simulated. The top 1% shows that almost 91% of the actual alphas are greater than the simulated. As such, this indicates evidence of superior performance; that in this case it is not just due to luck. As noted, no studies have been identified that use UK bond mutual fund data. Therefore, a direct comparison with previous findings has not been possible on this basis. However, the results here are more positive than those from UK equity mutual funds. Blake et al., (2017) find no evidence of skill when using the bootstrap approach of Fama and French (2010), based on net returns, regardless of the selection criteria applied to the sample. Findings from the Canadian market are more consistent with the results identified here. Ayadi and Kryzanowski (2011) also find that bond funds show evidence of superior performance, although the block-bootstrap approach is used rather than entire cases resampling as per Fama and French (2010).

The best performance is found for the Corporate bond fund category. Evidence of skill shows in the 97th, 98th, and 99th percentiles. The likelihoods are also very high; all are at least 90%. However, below the 97th percentile, the actual performance is worse than the simulated. Diversified funds show the least evidence of skill. The t-stats in the

Table 6.3: Simulation Results - 1999 to 2016

This table presents the results from the FF(2010) bootsrap method for All funds. The minimum number of return observations for each fund is 24 months. The Maturity 5 model has been used in each case. Sim refers to the t-statistics derived from the zero-alpha distribution and Act refers to the t-statistics from the actual funds in the sample. Newey and West (1987) adjustment for serial correlation and heteroskedasticty has been applied. LR % refers to the likelihood ratio statistics indicating the proportion of simulated t-stats below the t-stats of actual fund performance at each percentile

	All			**Corporate**			**Diversified**			**Government**		
	Sim	**Act**	**LR %**	**Sim**	**Act**	**LR %**	**Sim**	**Act**	**LR %**	**Sim**	**Act**	**LR %**
1	-2.58	-5.79	0	-2.51	-5.47	0	-2.63	-8.29	0	-2.76	-5.61	1.91
2	-2.20	-5.37	0	-2.17	-5.34	0	-2.22	-6.58	0	-2.24	-5.04	0.28
3	-1.99	-5.18	0	-1.96	-5.25	0	-2.00	-5.67	0	-2.01	-4.69	0.09
4	-1.84	-4.97	0	-1.81	-5.14	0	-1.84	-5.14	0	-1.84	-4.61	0.03
5	-1.71	-4.60	0	-1.69	-5.01	0	-1.72	-4.66	0	-1.71	-4.12	0.06
10	-1.31	-3.75	0	-1.30	-3.46	0.01	-1.32	-4.10	0	-1.28	-3.13	0.21
20	-0.86	-2.61	0.01	-0.84	-2.36	0.15	-0.86	-3.27	0	-0.82	-2.48	0.38
30	-0.53	-1.95	0.06	-0.53	-1.83	0.27	-0.53	-2.62	0	-0.51	-1.77	1.52
40	-0.26	-1.59	0.07	-0.26	-1.45	0.48	-0.25	-2.20	0	-0.26	-1.23	4.67
50	0.00	-1.08	0.15	-0.01	-0.96	1.30	0.02	-1.61	0	-0.03	-0.71	11.28
60	0.26	-0.56	1.01	0.24	-0.34	8.32	0.28	-1.2	0	0.2	-0.4	13.45
70	0.54	-0.05	4.46	0.51	0.05	14.11	0.57	-0.55	0	0.44	-0.02	19.06
80	0.86	0.33	5.83	0.83	0.38	14.69	0.92	0.13	0.37	0.74	0.28	18.35
90	1.34	0.69	2.22	1.28	0.74	9.53	1.42	0.63	0.51	1.17	0.82	24.9
95	1.76	1.22	7.05	1.68	1.23	15.83	1.88	0.89	0.21	1.54	1.46	47.48
96	1.88	1.50	16.23	1.8	1.68	42.23	2.03	1.07	0.53	1.65	1.6	50.33
97	2.05	1.80	29.22	1.95	2.66	92.33	2.22	1.23	0.68	1.79	1.66	44.72
98	2.28	2.56	75.96	2.18	3.36	98.03	2.51	2.02	19.38	1.97	1.67	33.47
99	2.69	3.37	90.36	2.56	4.82	99.7	3.22	3.21	59.24	2.33	2.67	73.32
	Gross Returns - January 1999 until July 2016											
	Sim	**Act**	**LR %**	**Sim**	**Act**	**LR %**	**Sim**	**Act**	**LR %**	**Sim**	**Act**	**LR %**
1	-2.58	-4.58	0.2	-2.51	-4.75	0.43	-2.63	-5.48	0.16	-2.76	-4.12	9.51
2	-2.21	-3.66	0.36	-2.17	-4.53	0.05	-2.22	-3.25	2.88	-2.24	-3.62	4.16
3	-1.99	-3.28	0.57	-1.96	-3.88	0.11	-2.00	-3.15	1.03	-2.01	-3.43	2.26
4	-1.84	-2.92	1.11	-1.81	-3.46	0.24	-1.84	-2.93	1.02	-1.84	-3.07	2.72
5	-1.71	-2.8	0.92	-1.69	-2.82	1.88	-1.72	-2.91	0.53	-1.71	-2.7	4.58
10	-1.31	-1.72	14.1	-1.3	-1.45	34.84	-1.32	-2.14	1.74	-1.28	-1.38	39.61
20	-0.86	-0.84	49.17	-0.84	-0.7	61.09	-0.86	-1.44	4.07	-0.82	-0.26	86.77
30	-0.53	-0.3	72.43	-0.52	-0.25	72.95	-0.53	-0.85	14.61	-0.51	0.12	88.31
40	-0.26	0.11	83.65	-0.26	0.09	78.81	-0.25	-0.33	38.21	-0.26	0.33	85.71
50	0.00	0.51	91.31	-0.01	0.55	90.12	0.02	0.25	78.75	-0.03	0.65	89.37
60	0.26	0.87	94.67	0.24	0.82	90.58	0.28	0.63	87.08	0.2	1.09	94.78
70	0.53	1.26	96.72	0.51	1.35	96.81	0.57	0.97	89.59	0.44	1.27	93.51
80	0.86	1.58	96.45	0.83	1.76	97.83	0.92	1.45	94.18	0.74	1.59	94.48
90	1.34	2.11	97.04	1.28	2.66	99.69	1.42	1.97	92.8	1.17	1.73	87.1
95	1.75	2.79	99.01	1.68	3.48	99.93	1.88	2.45	90.93	1.54	2.00	82.17
96	1.88	2.98	99.15	1.80	3.58	99.91	2.03	2.8	94.73	1.65	2.09	80.8
97	2.05	3.22	99.28	1.95	3.81	99.9	2.22	2.94	92.09	1.79	2.14	76.41
98	2.28	3.60	99.46	2.18	4.22	99.89	2.51	3.32	91.06	1.97	2.17	67.53
99	2.68	4.27	99.31	2.55	5.36	99.86	3.22	4.59	90.90	2.32	2.84	79.88

Table 6.4: Simulation Results - 2009 to 2016

This table presents the results from the FF(2010) bootsrap method for All funds. The minimum number of return observations for each fund is 24 months. The Maturity 5 model has been used in each case. Sim refers to the t-statistics derived from the zero-alpha distribution and Act refers to the t-statistics from the actual funds in the sample. Newey and West (1987) adjustment for serial correlation and heteroskedasticty has been applied. LR % refers to the likelihood ratio statistics indicating the proportion of simulated t-stats below the t-stats of actual fund performance at each percentile.

	All			**Corporate**			**Diversified**			**Government**		
	Sim	**Act**	**LR %**	**Sim**	**Act**	**LR %**	**Sim**	**Act**	**LR %**	**Sim**	**Act**	**LR %**
1	-2.84	-5.97	0.17	-2.52	-4.08	5.56	-2.84	-7.19	0.24	-3.13	-3.52	31.09
2	-2.41	-3.97	2.57	-2.13	-2.97	14.08	-2.38	-6.81	0.01	-2.52	-3.47	16.04
3	-2.16	-3.20	6.79	-1.91	-2.76	12.63	-2.15	-5.56	0.01	-2.22	-3.10	15.11
4	-1.99	-3.00	6.75	-1.76	-2.48	15.32	-1.98	-5.17	0.04	-2.01	-2.86	14.93
5	-1.85	-2.81	6.90	-1.63	-2.27	16.80	-1.86	-4.71	0.09	-1.86	-2.86	11.65
10	-1.39	-2.01	12.70	-1.23	-1.47	31.93	-1.43	-2.78	2.12	-1.40	-1.67	30.63
20	-0.89	-1.28	21.10	-0.78	-1.09	30.55	-0.93	-2.01	3.24	-0.86	-1.11	29.88
30	-0.55	-0.87	25.74	-0.48	-0.75	33.38	-0.58	-1.42	5.29	-0.51	-0.77	30.83
40	-0.26	-0.51	32.08	-0.23	-0.45	37.05	-0.28	-0.75	14.92	-0.25	-0.60	30.05
50	-0.01	-0.25	33.81	0.00	-0.18	40.28	0.00	-0.46	15.35	-0.04	-0.27	37.44
60	0.24	0.07	40.21	0.23	0.13	45.97	0.28	-0.13	18.07	0.17	0.02	43.30
70	0.52	0.44	49.31	0.48	0.48	53.31	0.58	0.35	33.23	0.42	0.22	42.89
80	0.85	0.78	51.82	0.78	0.87	59.10	0.93	0.78	41.21	0.74	0.63	47.52
90	1.34	1.24	49.24	1.23	1.43	65.92	1.44	1.45	55.57	1.22	0.92	34.48
95	1.78	1.72	51.62	1.64	1.82	64.50	1.91	2.30	77.82	1.62	1.12	23.71
96	1.92	1.91	54.94	1.78	2.02	67.16	2.06	2.50	78.77	1.75	1.31	27.39
97	2.09	2.19	62.09	1.95	2.15	65.01	2.25	2.74	79.07	1.91	1.42	25.02
98	2.34	2.83	81.01	2.20	3.45	93.76	2.55	2.89	71.29	2.14	1.46	18.10
99	2.79	3.40	82.32	2.69	4.90	98.38	3.23	3.23	58.03	2.53	1.48	9.55

Gross Returns - September 2009 until July 2016

	All			**Corporate**			**Diversified**			**Government**		
	Sim	**Act**	**LR %**	**Sim**	**Act**	**LR %**	**Sim**	**Act**	**LR %**	**Sim**	**Act**	**LR %**
1	-2.84	-3.04	34.15	-2.52	-2.92	27.34	-2.84	-4.93	3.00	-3.13	-3.06	43.79
2	-2.41	-2.25	54.17	-2.13	-1.50	79.26	-2.38	-2.66	29.69	-2.53	-1.89	71.23
3	-2.16	-1.99	56.00	-1.91	-1.35	76.62	-2.14	-2.33	33.33	-2.22	-1.59	73.99
4	-1.99	-1.72	63.22	-1.76	-1.24	75.21	-1.98	-2.21	30.48	-2.01	-1.33	78.39
5	-1.85	-1.43	76.05	-1.63	-1.18	71.19	-1.86	-2.16	25.50	-1.87	-0.91	90.33
10	-1.39	-0.92	81.74	-1.23	-0.72	74.62	-1.43	-1.73	23.19	-1.40	-0.42	94.49
20	-0.89	-0.10	96.79	-0.78	0.09	89.56	-0.93	-0.49	84.36	-0.86	-0.20	85.03
30	-0.55	0.34	96.79	-0.48	0.40	88.68	-0.58	0.17	97.73	-0.51	0.29	88.13
40	-0.26	0.66	95.17	-0.22	0.78	91.40	-0.28	0.42	95.47	-0.25	0.60	87.34
50	-0.01	0.94	94.13	0.00	1.03	91.50	0.00	0.87	97.18	-0.04	0.82	86.62
60	0.24	1.18	93.41	0.23	1.29	91.98	0.28	1.14	95.60	0.17	0.96	85.09
70	0.52	1.39	92.21	0.48	1.64	93.50	0.58	1.48	95.72	0.42	1.18	86.28
80	0.85	1.79	93.70	0.78	1.96	93.94	0.93	2.01	97.24	0.74	1.32	83.30
90	1.34	2.26	94.00	1.23	2.48	94.75	1.44	2.70	98.11	1.22	1.65	77.91
95	1.78	2.88	96.36	1.64	3.31	97.91	1.91	3.21	97.73	1.61	1.87	69.15
96	1.92	3.12	97.01	1.78	3.42	97.81	2.06	3.71	98.89	1.74	1.97	67.50
97	2.09	3.50	98.19	1.95	3.57	97.35	2.26	3.78	97.62	1.91	2.06	62.85
98	2.34	3.91	98.41	2.20	4.18	98.58	2.55	4.09	96.20	2.14	2.15	56.02
99	2.79	4.36	97.55	2.69	5.72	99.56	3.24	4.44	86.07	2.52	2.42	50.20

99th percentile are very similar between actual and simulated results. Likelihood ratios show that almost 60% of the actual alphas are greater than the simulated ones. The category of Government bond funds shows very minimal evidence of outperformance. Again, the t-stats are very similar in the top 1%, with just over 70% of the actual alphas likely to be greater than the simulated ones.

When using gross returns, the category that incorporates All funds shows that the worst managers still underperform the passive benchmarks, not just by chance. The left tail of the actual fund performance is to the left of that of the simulated. This is evidenced also by the Likelihood statistics, which show that up until the 20th percentile there is a higher proportion of actual performance that is worse than the simulated. In the 90th percentile, the average simulated t-stat is 1.3408 and the average actual is 2.0169, with a likelihood of 95.47%. The results are similar across each of the separate categories of funds. Poor performance shows for the Corporate bond funds in the left tail of the distribution, whereby until the 20th percentile the actual performance is worse than the simulated. Significant skill is evident from the 90th percentile onwards; simulated vs actual t-stats are 1.285 and 2.664 respectively. The likelihoods show that almost 100% of the actual performance is above the simulated, on average. Similar is shown for the Diversified category. The left tail is slightly longer, with the actual performance not beating the simulated until the 50th percentile. It is the 90th percentile again where the significant skill begins to show. The highest proportion of actual t-stats greater than the simulated is in the 96th percentile, whereby this is so in almost 95% of the cases. Government bond funds have “bad skill”, significantly until the 5th percentile. Like the Corporate funds, the proportion of actual performance greater than the simulated is evident from the 20th percentile. Significant skill begins to show in the 95th percentile; average simulated t-stat of 1.538 vs actual of 2.003, with a likelihood ratio of 82.17.

6.5.2 September 2009 until July 2016

Table 6.4 presents the results from the same analysis, this time from September 2009 until July 2016. The application of the bootstrap here adds further conviction to the time-variation observed in the preceding chapter, whereby Government, Corporate, and Diversified funds were all found to perform significantly better relative to the whole period. When considering All funds, the actual performance is better than that of the simulated in the 95th to the 99th percentiles, rather than only in the top 1% as was the case over the whole sample. In the left tail, it is not until the 60th percentile that the proportion of positive actual alphas begins to exceed the number of those simulated. For the Corporate funds, the actual performance begins to exceed the simulated in the 70th percentile. From the 96th to 99th percentiles there is more convincing evidence of positive net alpha as the significant actual t-stats exceed the simulated in most instances; notably so in the top 2%, with likelihoods of almost 94% and 98% for the 98th and 99th percentiles respectively.

There is more evidence of superior performance from the Diversified funds during this sub-sample; in the top 10% rather than just the top 1%. The Government bond funds, however, struggle to show any evidence of outperformance in these recent years. In the 80th percentile there is the most evidence that the alphas in this category are at least as positive as the simulated ones; the t-stats are similar although not significant, and the likelihood ratio is almost 50%. In the worst 1%, the t-stats are very similar between the actual and simulated values, with a likelihood of 31.09%. This indicates that the basis of performance improvement for the Government funds was due to a reduction in the left tail, as opposed to a notably greater proportion of funds realising positive alphas.

Using gross returns, evidence of skill is again observed for all the categories. Significant actual t-stats exceed the simulated from the 80th percentile for the corporate and diversified categories, the 90th across the group of all funds, and from the 96th for the Government bond funds. In the left tail, the worst managers underperform not just by chance, but do appear to exhibit significant "bad skill"; in the case of the All, Corporate, and Diversified categories. The likelihoods here show that in the bottom percentile for each of these categories, there is a higher proportion of average actual performance that is worse than that of the average simulated. Government bond funds do not appear to perform as bad. Even in the worst percentile, both the actual and simulated t-stats are very similar; -3.06 and -3.13 respectively. The likelihood ratio is almost 44%.

6.5.3 Robustness tests

Table 6.5 shows the results from a comparison of observations across All funds. For both periods, the results are generally consistent whether 24 or 60 observations are used, albeit a slight positive bias with the latter.

Table 6.5: Comparison of observations

This table presents the results from the FF(2010) bootsrap method for All funds. The minimum number of return observations for each fund is 24 months. The Maturity 5 model has been used in each case. Sim refers to the t-statistics derived from the zero-alpha distribution and Act refers to the t-statistics from the actual funds in the sample. Newey and West (1987) adjustment for serial correlation and heteroskedasticty has been applied. LR % refers to the likelihood ratio statistics indicating the proportion of simulated t-stats below the t-stats of actual fund performance at each percentile.

	1999 -2016						2009 - 2016					
	24 observations			60 observations			24 observations			60 observations		
	Sim	Act	LR %	Sim	Act	LR %	Sim	Act	LR %	Sim	Act	LR %
1	-2.58	-5.79	0	-2.42	-5.99	0.00	-2.84	-5.97	0.17	-2.60	-5.84	0.13
2	-2.20	-5.37	0	-2.10	-5.47	0.00	-2.41	-3.97	2.57	-2.24	-4.39	0.95
3	-1.99	-5.18	0	-1.91	-5.22	0.00	-2.16	-3.20	6.79	-2.04	-3.10	7.62
4	-1.84	-4.97	0	-1.76	-5.07	0.00	-1.99	-3.00	6.75	-1.88	-2.88	8.10
5	-1.71	-4.60	0	-1.65	-4.73	0.00	-1.85	-2.81	6.90	-1.75	-2.79	7.27
10	-1.31	-3.75	0	-1.27	-3.79	0.00	-1.39	-2.01	12.70	-1.32	-1.96	13.92
20	-0.86	-2.61	0.01	-0.83	-2.75	0.00	-0.89	-1.28	21.10	-0.83	-1.17	26.23
30	-0.53	-1.95	0.06	-0.52	-2.15	0.05	-0.55	-0.87	25.74	-0.51	-0.82	29.46
40	-0.26	-1.59	0.07	-0.25	-1.62	0.09	-0.26	-0.51	32.08	-0.25	-0.45	36.41
50	0.00	-1.08	0.15	0.00	-1.23	0.10	-0.01	-0.25	33.81	-0.02	-0.17	40.88
60	0.26	-0.56	1.01	0.25	-0.68	0.85	0.24	0.07	40.21	0.21	0.18	49.63
70	0.54	-0.05	4.46	0.52	-0.10	4.98	0.52	0.44	49.31	0.46	0.54	58.70
80	0.86	0.33	5.83	0.84	0.33	9.02	0.85	0.78	51.82	0.77	0.94	66.30
90	1.34	0.69	2.22	1.29	0.77	8.80	1.34	1.24	49.24	1.22	1.37	67.39
95	1.76	1.22	7.05	1.68	1.28	16.75	1.78	1.72	51.62	1.61	1.82	70.14
96	1.88	1.50	16.23	1.80	1.66	39.89	1.92	1.91	54.94	1.73	2.12	78.49
97	2.05	1.80	29.22	1.94	1.82	41.78	2.09	2.19	62.09	1.87	2.62	89.08
98	2.28	2.56	75.96	2.15	2.93	94.70	2.34	2.83	81.01	2.07	3.27	95.20
99	2.69	3.37	90.36	2.49	3.47	96.04	2.79	3.40	82.32	2.40	3.65	94.96

To determine the extent of sensitivity to test design, various inputs have been amended and the results derived. The number of required observations for each fund varies between studies. Harvey and Liu (2020b) conduct a comparison of the 8 observations required by FF (2010) and 60 by KTWW (2006). It is more difficult to isolate the impact of this difference in the original papers, as the data is not otherwise like-for-like. The findings from HL suggest that the FF (2010) method is optimal, subject to increasing the minimum return requirement, yet keeping it lower than the more extreme requirement of 60 observations.

The impact of model specification is highlighted below in Table 6.6. The results from the Maturity 5 model are compared against those derived using the style-specific benchmark as defined by Morningstar, the Barclays Sterling Agg Corporate index.

Table 6.6: Comparison of Models

This table presents the results from the FF(2010) bootsrap method for All funds. The minimum number of return observations for each fund is 24 months. The Maturity 5 model has been used in each case. Sim refers to the t-statistics derived from the zero-alpha distribution and Act refers to the t-statistics from the actual funds in the sample. Newey and West (1987) adjustment for serial correlation and heteroskedasticty has been applied. LR % refers to the likelihood ratio statistics indicating the proportion of simulated t-stats below the t-stats of actual fund performance at each percentile.

	1999 - 2016						**2009 - 2016**					
	Maturity 5			**Corporate Index**			**Maturity 5**			**Corporate Index**		
	Sim	**Act**	**LR %**	**Sim**	**Act**	**LR %**	**Sim**	**Act**	**LR %**	**Sim**	**Act**	**LR %**
1	-2.51	-5.47	0.07	-2.31	-4.18	0.84	-2.52	-4.08	5.56	-2.23	-1.68	74.86
2	-2.17	-5.34	0.00	-2.03	-3.86	0.58	-2.13	-2.97	14.08	-1.95	-1.11	89.90
3	-1.96	-5.25	0.00	-1.86	-3.45	0.75	-1.91	-2.76	12.63	-1.78	-1.03	87.53
4	-1.81	-5.14	0.00	-1.73	-3.14	1.10	-1.76	-2.48	15.32	-1.66	-0.97	86.14
5	-1.69	-5.01	0.00	-1.62	-2.79	2.25	-1.63	-2.27	16.80	-1.57	-0.94	83.73
10	-1.3	-3.46	0.01	-1.27	-1.78	13.75	-1.23	-1.47	31.93	-1.23	-0.77	75.64
20	-0.84	-2.36	0.15	-0.84	-1.02	31.81	-0.78	-1.09	30.55	-0.82	-0.20	85.97
30	-0.53	-1.83	0.27	-0.53	-0.52	48.72	-0.48	-0.75	33.38	-0.51	0.15	88.49
40	-0.26	-1.45	0.48	-0.27	-0.17	57.27	-0.23	-0.45	37.05	-0.25	0.54	92.73
50	-0.01	-0.96	1.30	-0.02	0.11	61.25	0.00	-0.18	40.28	-0.01	0.82	93.69
60	0.24	-0.34	8.32	0.23	0.45	70.54	0.23	0.13	45.97	0.23	1.07	93.79
70	0.51	0.05	14.11	0.49	0.81	77.80	0.48	0.48	53.31	0.48	1.43	95.26
80	0.83	0.38	14.69	0.79	1.20	83.58	0.78	0.87	59.10	0.77	1.81	96.03
90	1.28	0.74	9.53	1.21	1.84	91.77	1.23	1.43	65.92	1.15	2.40	97.66
95	1.68	1.23	15.83	1.56	1.99	82.64	1.64	1.82	64.50	1.48	2.91	98.14
96	1.8	1.68	42.23	1.67	2.20	87.14	1.78	2.02	67.16	1.57	3.10	98.38
97	1.95	2.66	92.33	1.79	2.55	93.00	1.95	2.15	65.01	1.69	3.27	98.31
98	2.18	3.36	98.03	1.97	3.45	99.48	2.20	3.45	93.76	1.85	3.54	98.51
99	2.56	4.82	99.70	2.24	3.70	98.61	2.69	4.90	98.38	2.13	4.00	98.70

Here, the most striking results are observed in the recent period. For both models, there is evidence of superior performance, although to a far greater extent using the Corporate index. In this case, across the entire distribution actual performance exceeds the simulated. A higher hurdle is clearly presented by the multi-factor model. Such a comparison again highlights the benefit of conducting model testing to identify the likelihood of biases that may prevail. Furthermore, the results here cast doubt on the proposition that bond funds are generally not as sensitive to the model, be it single or multi-factor, as is the case for equity funds.

The Corporate category has been selected to highlight the impact of model specification. One reason for this is that it represents the largest proportion of the sample. Furthermore, it is this category for which there was any indication of timing ability in the previous chapter. Performance attribution of credit funds is highlighted as an interesting area for future research, with recent literature exploring the determinants of the credit risk premium and identifying sources of alpha. (Houweling and Van Zundert 2017)(Frieda and Richardson 2016)

6.5.4 Summary of bootstrap results

The findings from this chapter generally support the observation that Corporate, Diversified, and Government funds perform better in the post-crisis period. Improvements come in two respects; a higher proportion of outperforming funds in the right tail, but also less extreme values in the left tail are identified. A review of the descriptive statistics suggested that this was likely to be the case, given that the negative skew and excess kurtosis did not prevail to such an extent from 2009 to 2016. A comparison of the cumulative distribution function figures highlights the extent of this difference between periods.

As discussed in Section 6.3 various methods of bootstrapping exist and have been applied to performance evaluation studies. The FF approach was considered optimal to use here. It has not yet otherwise been applied to a sample of bond funds. Furthermore, C. Harvey and Liu (2020) highlight that the extent of outperformance observed is likely to be biased downwards through this approach, therefore suggesting that the actual extent of positive alpha is likely higher than can be inferred from here.

In a similar context, the comparison of models that was conducted for the Corporate funds shows the extent to which the multi-factor model presents a higher hurdle for the funds. As such, various factors have already been presented thus far that support the evidence of outperformance.

6.6 Empirical results - FDR methods

6.6.1 BSW (2010) results

As at the date of writing, to the best of my knowledge, the false discovery rate method has not been applied to bond fund performance evaluation. This chapter adopts both the classical approach of BSW (2010) and the recent advancements of Ferson and Chen (2019). By doing so it allows for an evaluation of the differences that are observed and helps provide a basis for future refinements regarding optimal methods for bond fund evaluation. Application of the FC (2019) method is a natural extension of the bootstrap already applied in section 6.5. However, a limitation is the additional complexity it presents. This is an attraction of BSW (2010) – the relative ease of implementation that it allows.

A criticism of the bootstrap approach to overcoming multiple testing issues is that it does not adjust for false positives in the tails. For example, as per Table 6.4, across All funds the FF bootstrap shows that over half of the actual performance outperforms what would be expected from the simulated no skill distribution in the top 5%. Although this is more robust than the point estimate methods used in Chapter 5, it does not adequately control for luck. Table 6.7 presents the results for the BSW method of adjusting for false discovery rates. Both the whole period and post-crisis period have been considered separately. The performance in the left and right tails is shown for various levels of significance; 5%, 10%, 15%, and 20%. As the Corporate bond category appears to have thus far shown the greatest potential for achieving significant outperformance, the results from the FDR approach specific to these funds are presented in Table 6.8.

An initial evaluation of the p-value distribution allows for an indication as to the extent of non-zero alphas performance that may be expected. Figures 6.3 and 6.4 illustrate this for all funds in the sample and more specifically the Corporate funds during

Table 6.7: FDR Results - BSW Method - All Funds

This table presents the results from the BSW (2010) False Discovery Rate method applied to all bond funds in the sample. The performance has been measured using the Maturity 5 model. For various levels of significance ($\gamma = 0.05, 0.10, 0.15, 0.20$), the proportions of significant, lucky/unlucky funds are presented. The figures are in percentages (%) throughout. The standard errors of the estimates are presented in parenthesis on each line below. Characteristics of the funds in each tail have also been included, such as the total expense ratio (Exp.) and the average fund alpha. Both of which are on an annual basis and in percent. The number of funds active during 1999-2016 is 369, and during 2009-2016 is 304. Simulations have been used to estimate λ^* for each period, the value of which being 0.5051 and 0.5076 respectively. The corresponding values of π_0 are 57% and 73%.

Left-Tail	**0.05**	**0.10**	**0.15**	**0.20**	**0.20**	**0.15**	**0.10**	**0.05**	**Right-Tail**
1999 - 2016									
Signif. $\hat{S}_\gamma^-$	31.71	39.84	45.53	47.70	6.23	5.15	4.07	3.25	Signif. $\hat{S}_\gamma^+$
	2.42	2.55	2.59	2.60	1.26	1.15	1.03	0.92	
Unlucky $\hat{F}_\gamma^-$	1.41	2.83	4.24	5.66	5.66	4.24	2.83	1.41	Lucky $\hat{F}_\gamma^+$
	0.11	0.23	0.34	0.46	0.46	0.34	0.23	0.11	
Unskilled $\hat{T}_\gamma^-$	30.29	37.01	41.29	42.04	0.58	0.91	1.24	1.84	Skilled $\hat{T}_\gamma^+$
	2.48	2.68	2.81	2.91	1.41	1.25	1.08	0.94	
Alpha p.a	-1.89	-1.72	-1.81	-1.77	1.52	1.74	1.80	1.71	Alpha p.a
Exp. p.a	0.96	0.94	0.95	0.96	0.34	0.36	0.38	0.31	Exp. p.a
2009 - 2016									
Signif. $\hat{S}_\gamma^-$	14.14	18.09	21.05	26.32	13.49	11.51	9.87	7.89	Signif. $\hat{S}_\gamma^+$
	2.00	2.21	2.34	2.53	1.96	1.83	1.71	1.55	
Unlucky $\hat{F}_\gamma^-$	1.82	3.64	5.45	7.27	7.27	5.45	3.64	1.82	Lucky $\hat{F}_\gamma^+$
	0.13	0.27	0.40	0.53	0.53	0.40	0.27	0.13	
Unskilled $\hat{T}_\gamma^-$	12.33	14.46	15.60	19.04	6.21	6.06	6.23	6.08	Skilled $\hat{T}_\gamma^+$
	2.04	2.32	2.53	2.81	2.18	1.98	1.80	1.58	
Alpha p.a	-1.86	-1.75	-1.67	-1.70	1.57	1.71	1.80	1.65	Alpha p.a
Exp. p.a	0.97	0.94	0.93	0.98	0.60	0.62	0.63	0.62	Exp. p.a

Table 6.8: FDR Results - BSW Method - Corporate Funds

This table presents the results from the BSW (2010) False Discovery Rate method applied to the Corporate bond fund category in the sample. The performance has been measured using the Maturity 5 model. For various levels of significance ($\gamma = 0.05, 0.10, 0.15, 0.20$), the proportions of significant, lucky/unlucky funds are presented. The figures are in percentages (%) throughout. The standard errors of the estimates are presented in parenthesis on each line below. Characteristics of the funds in each tail have also been included, such as the total expense ratio (Exp.) and the average fund alpha. Both of which are on an annual basis and in percent. The number of funds active during 1999-2016 is 144, and during 2009-2016 is 121. Simulations have been used to estimate λ^* for each period, the value of which being 0.5051 and 0.4778 respectively. The corresponding values of π_0 are 59% and 71%.

Left-Tail	**0.05**	**0.10**	**0.15**	**0.20**	**0.20**	**0.15**	**0.10**	**0.05**	**Right-Tail**
1999 - 2016									
Signif. $\hat{S}_\gamma^-$	29.17	38.89	43.06	44.44	6.25	5.56	4.17	4.17	Signif. $\hat{S}_\gamma^+$
	3.79	4.06	4.13	4.14	2.02	1.91	1.67	1.67	
Unlucky $\hat{F}_\gamma^-$	1.48	2.95	4.43	5.90	5.90	4.43	2.95	1.48	Lucky $\hat{F}_\gamma^+$
	0.19	0.37	0.56	0.75	0.75	0.56	0.37	0.19	
Unskilled $\hat{T}_\gamma^-$	27.69	35.94	38.63	38.54	0.35	1.13	1.22	2.69	Skilled $\hat{T}_\gamma^+$
	3.87	4.28	4.48	4.64	2.27	2.08	1.76	1.70	
Alpha p.a	-1.86	-1.62	-1.57	-1.54	1.43	1.59	1.58	1.58	Alpha p.a
Exp. p.a	0.98	0.91	0.90	0.90	0.29	0.33	0.41	0.41	Exp. p.a
2009 - 2016									
Signif. $\hat{S}_\gamma^-$	11.57	14.88	18.18	21.49	17.36	14.05	11.57	9.09	Signif. $\hat{S}_\gamma^+$
	2.91	3.24	3.51	3.73	3.44	3.16	2.91	2.61	
Unlucky $\hat{F}_\gamma^-$	1.83	3.65	5.48	7.30	7.30	5.48	3.65	1.83	Lucky $\hat{F}_\gamma^+$
	0.22	0.44	0.66	0.88	0.88	0.66	0.44	0.22	
Unskilled $\hat{T}_\gamma^-$	9.74	11.22	12.70	14.18	10.05	8.57	7.92	7.26	Skilled $\hat{T}_\gamma^+$
	2.98	3.40	3.79	4.16	3.84	3.42	3.06	2.67	
Alpha p.a	-1.68	-1.51	-1.44	-1.47	1.46	1.62	1.75	1.28	Alpha p.a
Exp. p.a	1.08	0.99	0.96	0.95	0.60	0.63	0.64	0.59	Exp. p.a

the post-crisis period. The empirical p-value distributions have been calculated for the individual funds, evaluated using the Maturity 5 model. The concentration of p-values to the left of the distribution indicates a high proportion of funds for which the null hypothesis of zero alpha may be rejected. What is unclear, however, is whether this will be in favour of significant positive of negative performance.

When considering All funds the dominance of the long left tail is evident. During 1999-2016, at 20% significance 47.70% of the sample underperformed, of which only 5.66% were unlucky. The remaining 42.04% (155 funds) were truly unskilled. Unfortunately there is a limited selection of fund characteristics available in Morningstar.[4] However, details of the average annual alphas and expense ratios for the funds in each tail are presented. As would be expected, higher expenses are associated with poorer performance. The right tail exhibits very minimal evidence of true ability. This is isolated in the top 5%. Even then, only marginal significance is observed and given the relatively small sample size, would only account for approximately three funds. The magnitude of annualised alphas in each tail are relatively small, with 1.80% at the upper range and -1.89% the lower.

Focusing on the post-crisis period, however, shows that the improved performance is supported. The extent of underperformance has reduced by almost half. Here, 26% are significantly poor. Adjusting for the unlucky funds leaves 19% (23 funds)as exhibiting true underperformance. More promising results are also found in the right tail, with 13.5% of funds showing significant outperformance. Adjusting for luck leaves 6.2% (7 funds) with true superior ability. The range of alphas achievable across the tails between periods remains almost unchanged.

Table 6.8 presents the results for the Corporate category. A comparison between the tables shows that these are very similar, highlighting the dominance that these funds have within the full sample. Post-crisis period results are again most interesting and where the strongest evidence is identified. Of the Corporate funds, only 14.18% (18

[4]This appears to be more limited for a UK-based sample than is the case for more commonly used US data

funds) are truly bad. The right tail shows that 10% of these funds have true ability, corresponding to 12 funds.

6.6.2 Ferson and Chen (2019) methodology - results

Given the FF method of entire-cases resampling has already been used in the chapter, it makes for an interesting extension to apply this approach to add further robustness to the results already observed. The aim here is to help provide insights and identify areas for further research in terms of addressing multiple hypothesis testing issues with regards to fixed income. With an absence of academic literature in this area and small sample size relative to that which has so far focused on equities, the results here are somewhat preliminary.

The first step to implementing the FC method is to simulate the distribution under the imposition of a zero-alpha null hypothesis. This is the same as per the initial bootstrap analysis in section 6.5. From this, the predefined α_g and α_b values are derived. For example, given a test size of 10% $t_{-} = 1.65$ and $t_{+} = 1.65$), the alphas can be deduced by a process of interpolation to determine the alpha values correspond to these critical t values in each tail. On this basis, referring to Table 6.2, α_g is equal to 0.12% and α_b is equal to -0.072% (monthly). The equivalent values in FC are 0.3170% and -0.2670% respectively.

As described in section 6.3.3, further simulations are conducted. Instead of setting the null equal to zero, one simulation adds α_g of 0.12% to each fund and another subtracts α_b of 0.072%. Table 6.9 presents the results for the FC (2019) method of adjusting for false discoveries. The size of the test γ varies from 5% to 20%, denoting the portion of the tail under consideration. A larger test size (40%) is more representative of the full sample of funds, however, the focus at present it to consider the more extreme values in the tails. The size selected has implications for the likelihood of error types that may be encountered. For example, a narrow range could lead to a higher likelihood of incurring Type II errors, i.e. failing to identify truly outperforming funds. Given the very small economic magnitude of the bond fund positive alphas that have resulted thus far, this has not been considered to be problematic in this respect.

Table 6.9: FDR Results - Ferson and Chen (2019) Method

This table presents the results from Ferson and Chen (2019) method of adjusting for false discovery rates. The Maturity 5 model has been used in each case. $\gamma/2$ denotes the size of the test. α_g and α_b refer to the locations of "good" and "bad" funds in the distribution. β_g and β_b refer to the power parameters. For example, β_g means the power to reject against the alternative hypothesis that a fund is good, and β_b relates to the converse, i.e. the power to reject against the bad alternative. δ_g and δ_b refer to the confusions. These denote the probabilities of mistaking a good fund for bad, and mistaking a bad fund for good, respectively. The corresponding empirical standard errors for the FC fractions are presented in the rows below.

January 1999 to July 2016										September 2009 to July 2016									
All Funds																			
α_g	α_b	$\gamma/2$	β_g	β_b	δ_g	δ_b	π_0	π_b	π_g	α_g	α_b	$\gamma/2$	β_g	β_b	δ_g	δ_b	π_0	π_b	π_g
0.12	-0.07	0.05	68.27	45.46	0.80	0.21	0.19	90.36	9.45	0.14	-0.10	0.05	65.65	45.37	0.71	0.30	0.00	69.15	30.84
							0.42	6.83	7.07								0.03	14.22	14.23
0.12	-0.07	0.10	77.39	57.30	1.42	0.52	0.05	88.87	11.08	0.14	-0.10	0.10	75.81	58.13	1.33	0.59	0.01	66.77	33.22
							0.15	6.74	6.81								0.06	14.18	14.20
0.12	-0.07	0.20	85.46	70.53	3.11	1.19	0.01	85.44	14.55	0.14	-0.10	0.20	85.22	72.97	2.76	1.43	0.01	62.45	37.54
							0.04	7.36	7.37								0.04	14.93	14.95
Corporate Funds																			
α_g	α_b	$\gamma/2$	β_g	β_b	δ_g	δ_b	π_0	π_b	π_g	α_g	α_b	$\gamma/2$	β_g	β_b	δ_g	δ_b	π_0	π_b	π_g
0.10	-0.08	0.05	61.59	48.56	0.43	0.21	0.15	85.25	14.60	0.11	-0.12	0.05	56.86	59.93	0.47	0.32	0.02	50.90	49.08
							0.40	9.49	9.70								0.17	19.15	19.19
0.10	-0.08	0.10	73.18	61.43	0.88	0.49	0.02	82.92	17.06	0.11	-0.12	0.10	70.81	73.39	0.96	0.72	0.01	52.28	47.72
							0.09	8.41	8.45								0.06	17.00	17.01
0.10	-0.08	0.20	83.64	75.19	1.99	1.15	0.00	79.08	20.92	0.11	-0.12	0.20	82.55	83.73	1.77	1.63	0.00	54.38	45.62
							0.02	8.58	8.58								0.03	17.51	17.51

The results derived using the Ferson and Chen (2019) method are the most positive yet in this thesis. Evidence of outperformance is evident even when the whole period is considered. Until now, this has predominantly been isolated in the post-crisis period. Given a test size of 10%, the FC method finds that 11% of funds exhibit significant outperformance. This value increases with the test size, indicating that the "good" funds are not only isolated in the most extreme regions, but distributed throughout the tail. However, despite this increased propensity for positive alpha, the same applies to the left tail; 89% are classified as"bad" funds. Again, similar values result for the Corporate funds, indicating the strong influence this category has on the overall performance observed; $\alpha_b = 83\%$ and $\alpha_g = 17\%$. The proportion of "good" funds also increases with test size here. Further improvements are realised in the post-crisis period. The results here show that 33% of All funds outperform and 48% of Corporate funds do. The proportions of "bad" performing funds are 67% and 52% respectively.

These results are striking relative to most others that seek to address the same/similar research question. For example, FC found no evidence of outperformance among their sample of US equity funds. However, the results they obtain from hedge funds resonate to a greater degree here. The proportions of hedge funds in each performance group are α_0=0%, α_g=53.2%, and α_b=46.8%. These results are derived having identified the optimal alpha locations using a grid search. For various alternative alpha values, no zero-alpha funds are identified.

To date, there are no other studies that have yet adopted this method as per Ferson and Chen (2019). Therefore, comparisons on this basis have not been possible. In future work, however, it will be interesting to see the developments and empirical refinements that continue to be made in this area.

6.7 Discussion of results

Despite this apparent outperformance from credit allocation, it must be noted that there may be a corresponding detrimental effect on diversification in a wider portfolio context. Although the UK bond mutual funds are seen to be outperforming, further analysis as to sources of return would allow for a more complete picture regarding the value that may be derived. This is likely to be of relevance given the magnitude of alphas observed for bond mutual funds. For example, relative to equities and hedge funds, they are notably smaller. As such, diversification gains could instead be of primary concern to investors

Figure 4.1: Credit correlations shows that since 2009 there have been periods whereby the Corporate funds are more highly correlated with the High Yield index. As a basis for comparison, the correlations of the Corporate index have also been presented. The higher correlation of the funds with High Yield could be indicative of active Corporate funds taking tilts towards higher yielding securities. It is during this period that the performance of the UK High Yield index (Figure 4.2) experienced favourable performance relative to both UK Investment Grade and European High Yield. Given that index composition is restricted by various rules, it is likely to be the case that UK bond funds have greater scope to be able to do this.

Figure 6.1: Credit Correlations

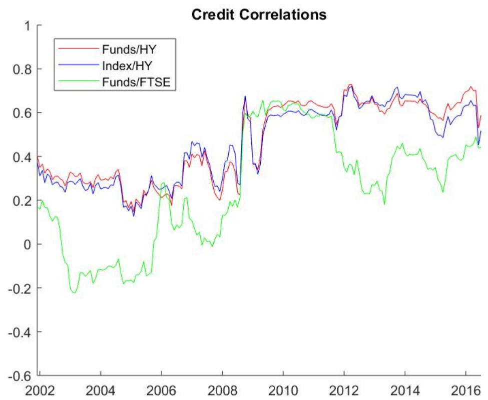
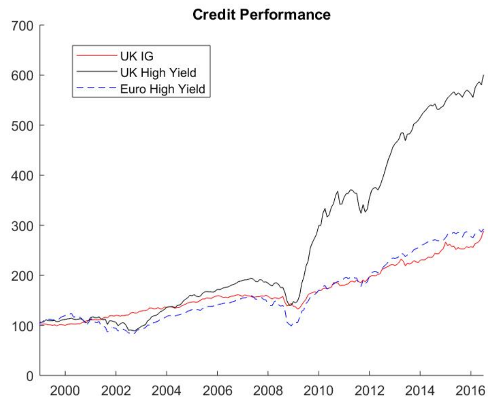


Figure 6.2: Corporate Returns



Choi and Kronlund (2018) identify "reaching for yield" tendencies in US Corporate bond mutual funds. With such activity comes both positive and negative implications. Although tilts to higher yielding securities have been found to bring higher returns and inflows, it is still often the case that negative performance prevails after risk-adjustment. Liquidity conditions tightened in corporate bond markets during the onset of the global financial crisis, and as such, these conditions were not associated with the propensity to reach for yield. However, with the sustained low yield environment being driven by easy monetary policy, the tendency to reach for yield has increased. Higher yields may be gained as a result of fund manger skill and are not always riskier, if able to identify and capitalise on these. However, the paper finds that the incremental returns from such tilts are not defined as alpha, given that they are fully explained by risk. The risk factors are defined here as stock market, term (30 year - 1 year Treasury), and default (EW Corp bond return - rf) - these factors explain the returns to RFY and as such do not indicate superior ability.

Brooks et al. (2019) find that active fixed income management has appeared to do well in terms of providing positive risk-adjusted returns over the past 20 years. As such, it may often be perceived that these managers are using their skills to do so. An important distinction to make here is whether the observed outperformance is driven by exposure to traditional premia (such as primarily credit risk in this case), or the uncorrelated returns that deliver "true alpha", that are driving the observed outperformance. It is acknowledged that isolating the determinants of excess returns in fixed income is challenging task, in part due to the size and scope of fixed income markets as noted.

An important point to consider in this respect is that the indices to which fixed income funds are benchmarked rarely capture a complete picture of the investment universe the managers have at their disposal. It may be the case, and often is, that managers are taking off-benchmark positions which are a key source of excess returns. (Fabozzi 2012) This highlights the importance of conducting model testing to ensure an appropriate performance hurdle is specified. A further point with regards to benchmarking is the time-variation of fixed income index composition. It has been identified that the proportion of credit in flagship benchmark indices, such as Barclays Aggregate, has been increasing in recent years. Overall, the findings of Brooks et al (2019) conclude that fixed income alpha is an "illusion". They raise concern with respect to the higher exposure to equities that is likely to result when managers reach for yield. For example, when they tilt towards higher risk and higher return segments of the fixed income markets, as noted above.

As discussed in Chapter 2, there has been a growing interesting in defining factors to explain the cross-section of corporate bond returns. With a wealth of literature directed at equities in this sense, it is perhaps surprising that there has been a lack of literature dedicated to doing so. Bai et al. (2019) find that stock market factors and aggregate bond market factors fare poorly in terms of explaining the cross-section of Corporate bond market returns. The authors construct four novel factors with significant risk premiums, unexplained by the common factors as noted. Specific features of the securities are required to construct these factors, which are defined as downside risk-factor, liquidity risk-factor, credit risk-factor, and a return reversal risk-factor. The authors examine the explanatory power using newly constructed test asset portfolios sorted by size, maturity, and industry. Inclusion of the four factors along with the bond market factors increases the explanatory power in relation to the portfolios as noted. Alphas are compensation for these risks. Therefore, it is suggested that these should be specified in factor models seeking to identify the true outperformance of Corporate bond portfolios.

6.8 Conclusions

A wealth of literature has been directed towards addressing multiple testing issues that plague empirical studies. By using a new sample of bond funds, it allows for further tests of methods used to assess the merits of active management. The results are promising. Regardless of the method used - bootstrapping or adjusting for false discovery rates - evidence of superior performance in the post-crisis period is identified. Various contributions to existing literature have been presented from this chapter.

First, the entire-cases resampling method is applied to a new sample of funds. To date, bond funds have not been included in similar studies. The results here add conviction to those from Chapter 5 whereby post-crisis period improvements were indicated. As such, the bootstraps show that when all funds are considered there is evidence of outperformance from the 97th percentile during 2009 - 2016. The category of Corporate bond funds shows the most potential for this, with evidence from the 95th percentile. This is also the case for Diversified funds, however, greater scope for underperformance is indicated by the longer left tail exhibited here.

To add robustness to these results, two methods to control for luck are applied; the "classical" false discovery rate method of Barras et al. (2010), and the recent empirical refinements as proposed by Ferson and Chen (2019). Results using the former indicate that any evidence of superior performance is isolated to the post-crisis period. Here, the sample of funds is grouped into performance defined by $\pi_g = 6\%$, $\pi_b = 19\%$, and the remainder unskilled, $\pi_0 = 75\%$. The results are marginally more positive when specifically considering the Corporate funds. The proportions in this case are $\pi_g = 10\%$, $\pi_b = 14\%$, and $\pi_0 = 76\%$. Chapter 5 found that expenses were a significant detriment to performance. This is again highlighted from a comparison of the tails; higher expense ratios characterising the funds in the left tail.

The BSW (2010) approach has been critiqued in recent literature regarding its low power to identify truly outperforming funds, resulting in a high proportion of π_0. This is also evident from the results in this chapter. The FC (2019) method seeks to address the lower power limitations. When applied here, the results from this method are the most favourable yet. Over the whole period and post-crisis, a 10% test size finds 10% and 33% of funds outperform respectively. In terms of the Corporate category, 17% and 48% of funds are found for these corresponding periods.

The application of both these methods to identify skill indicates that there is evidence of UK bond funds having the ability to return value to investors, particularly the Corporate category. However, it must be noted that the magnitude of alpha is small in economic terms and the funds do still exhibit notable scope for significant underperformance, as is apparent from the long left tail during 1999-2016. As such, the perspective of value may be debatable. For example, given the limited capacity for realising a high return advantage from these funds, a primary attraction may be for the purpose of diversification in a wider portfolio context. However, the findings from this chapter suggest that this sample of funds would perhaps not provide such an opportunity, given the increasing correlations with High Yield and equity markets. Further to this point, an additional use for FDR results is as a portfolio construction tool. Existing literature uses a rolling estimation to form portfolios conditional on funds achieving a target FDR (e.g. 10%). This has not been explored in this thesis, however, is acknowledged as an interesting area for further research. An additional perspective that could be taken is to assess the contribution to diversification of a typical 60/40 equity/bond portfolio if this method was applied.

Overall, the results from this empirical chapter provide support for the prevalence of outperformance among this sample of UK bond mutual funds. Various contributions to existing literature have been made, with many interesting directions in which the analysis may further progress also highlighted.

6.9 Appendix

Figure 6.3: P-values - All

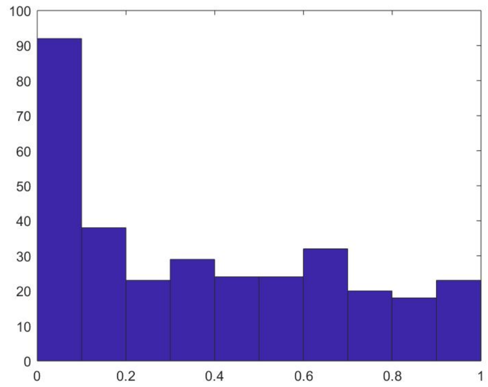


Figure 6.4: P-values - Corporate

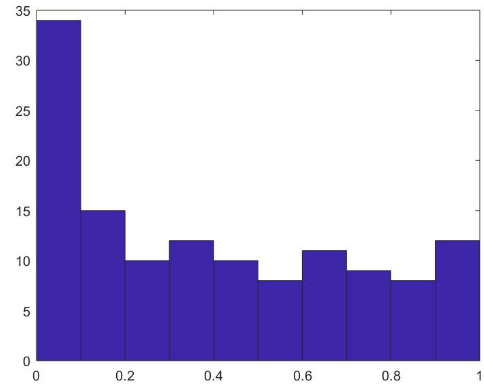


Figure 6.5: All Funds

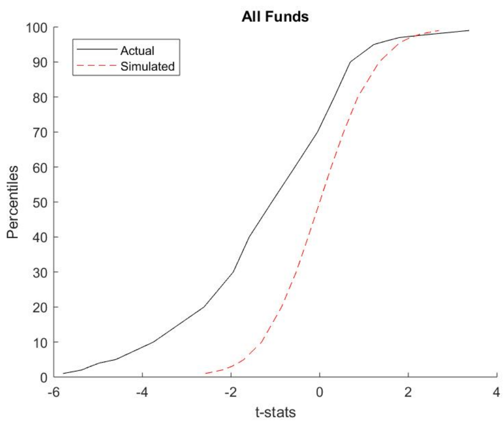


Figure 6.6: Corporate Funds

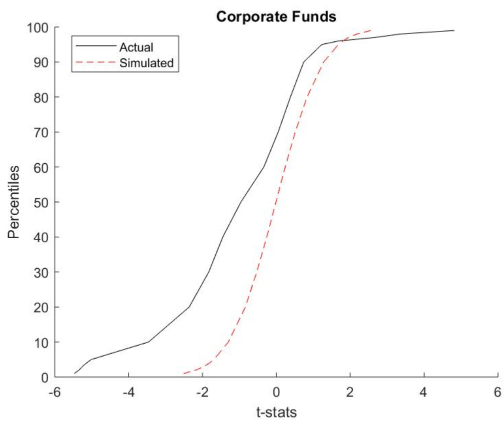


Figure 6.7: Diversified Funds

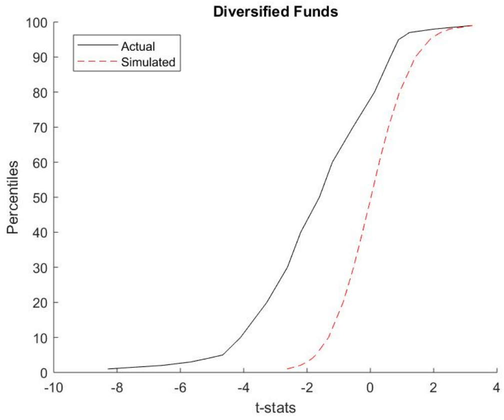


Figure 6.8: Government Funds

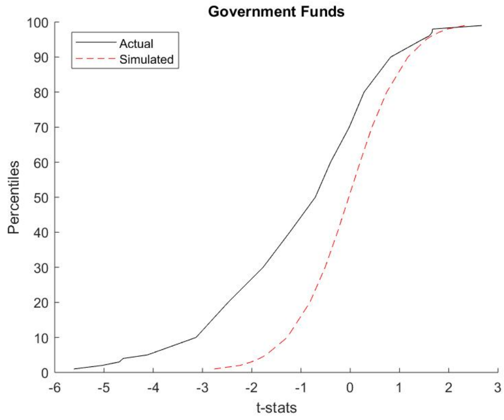


Figure 6.9: All Funds

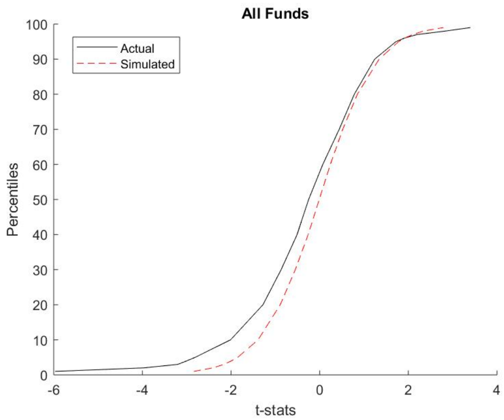


Figure 6.10: Corporate Funds

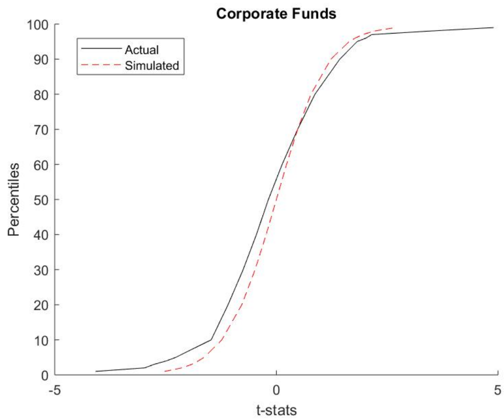


Figure 6.11: Diversified Funds

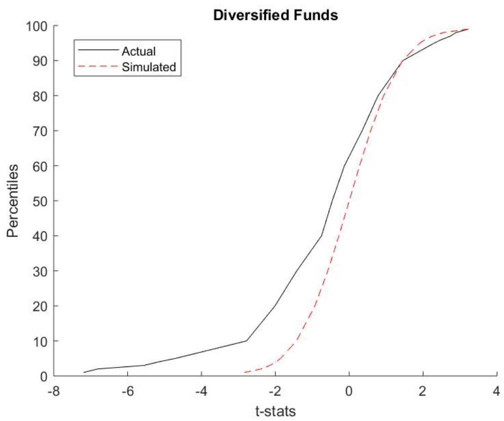


Figure 6.12: Government Funds

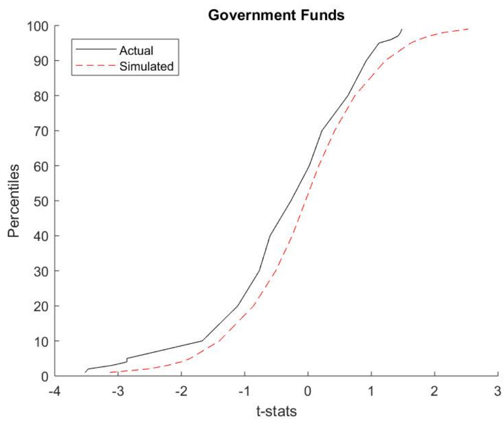


Figure 6.13: All Funds 24 obs.

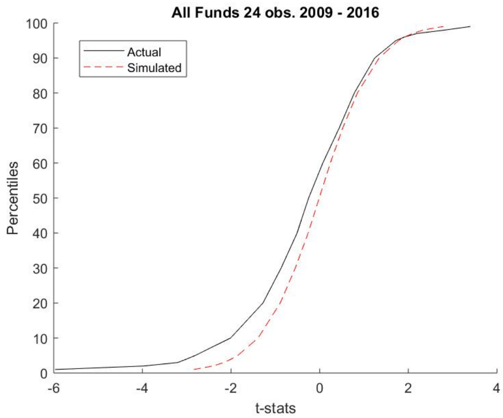


Figure 6.14: All Funds 60 obs.

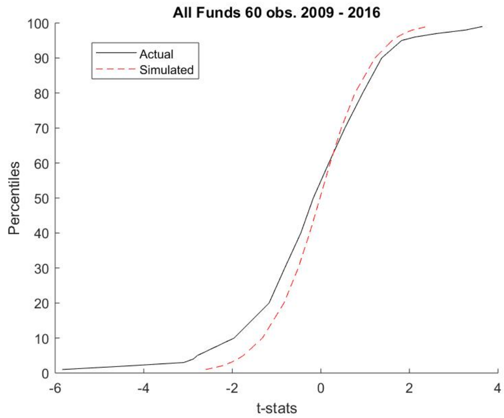


Figure 6.15: Maturity 5 Model

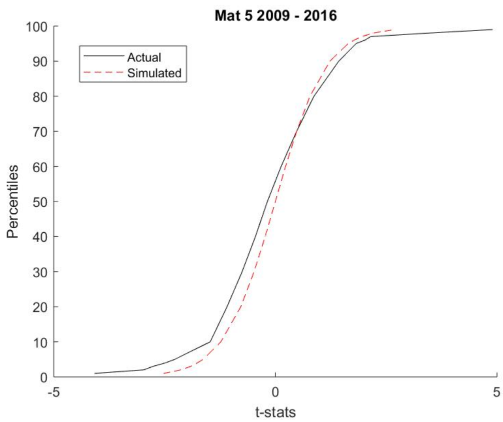
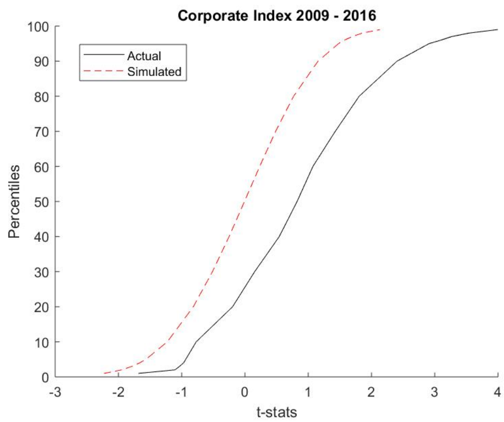


Figure 6.16: MS Corporate Index



Chapter 7

Conclusions

With a wealth of academic literature directed at evaluating the merits of active management, it is perhaps surprising there is a relative lack with regards to bond funds. This thesis identifies gaps in the literature, highlighting that the UK market has not yet been isolated in any of such studies to date. Various empirical methods are used throughout to enhance robustness and minimise bias that may be induced in conclusions regarding the performance of this sample of UK bond funds.

7.1 Summary of findings

7.1.1 Model testing

Central to the performance evaluation of actively managed funds is the specification of a model that will provide correct risk adjustment and induce minimal bias into the performance observed. The Gibbons, Ross, and Shanken (1989) test of mean-variance efficiency has been used along with the corresponding alpha statistics to assess the reliability of candidate models to be used for the performance evaluation of UK bond mutual funds. This seeks to determine if the factor(s) represent a proxy for the "market" portfolio. In the case of fixed income, as defined by term and credit. Candidate models are tested by assessing their power to explain the time-series returns of test asset passive portfolios, selected to represent the investment opportunity set available to the active managers. With a key research aim of this thesis being to evaluate a sample of actively managed funds, an examination of the potential for the alpha bias that may be induced is key. Although this chapter is framed in the context of the GRS tests, it is in fact the corresponding alpha statistics that are of most use to evaluate the suitability of the candidate models here.

Model comparison tests are limited in the academic literature in this area for fixed income funds. A reason for which is perhaps the assertion that bond funds are not as sensitive to the choice of model as is the case for equity funds, given their lower idiosyncratic risks (Ang 2014) The models proposed in this thesis are generally consistent with those applied in key papers, such as Blake et al (1995). However, they are the first to be applied to a sample of UK bond mutual funds and have been adapted accordingly; for example, with regards to geographical focus. Term and default factors define the "market" portfolio for bonds, with interest rate risk the dominant factor to which bonds are exposed. However, the cross-section of bond markets is vast; accurate benchmarking is in fact a difficult task. Thus, conducting model testing is a valuable step towards conducting fund performance evaluation with greater conviction.

Two single-factor and three multi-factor models were tested. The factors were selected to represent a range of credit and term exposures. From these, one consisting of five factors (named Maturity 5) performed the best in terms of minimising potential alpha bias in the performance evaluation. This model consistently exhibited preferential features, such as the smallest absolute alphas across the test asset portfolios, lowest standard errors of alpha, and the highest R^2s. Therefore, indicating minimum bias and high precision. In terms of passing the GRS test, evidence of this was minimal; however, this does not present a cause for concern in this situation. The sensitivity and power of the GRS test is frequently highlighted in existing literature as a limitation if basing judgment on this statistic, with models commonly failing. Given the appeal of other characteristics exhibited by the Maturity 5 model, it can be concluded as the most reliable for the purpose of fund performance evaluation.

7.1.2 Performance evaluation of UK bond mutual funds

Given the results from model testing, the actively managed bond funds were evaluated using the Maturity 5 model. Results over the whole period (1999-2016) indicate that in aggregate the funds underperform on a risk-adjusted basis. The equally-weighted portfolio of All funds realises a significant monthly alpha of -0.06%. Further regressions found that expenses were a significant determinant of the underperformance observed. However, the recent data from 2009-2016 is conducive to more promising results. Here, the overall performance is neutral. Dummy variables have been assigned to isolate the post-crisis period and identify if this difference is significant. Wald tests indicate that this is indeed the case. Corporate, Diversified, and Government funds show improvements during the low rate environment. With the model testing having shown that alphas across the test asset portfolios are tiny and generally insignificant, it can be inferred that the performance of the active funds is not being driven by exposure to passive strategies; it instead appears that there is ability beyond that this.

When bonds are concerned, it is difficult to ascertain the specific drivers of performance using a returns-based data. Nonetheless, the classical method of Treynor Mazuy (1966) was applied to the individual funds to assess the extent of any timing ability prevalent. Two market models are selected; the Barclays Sterling Aggregate index and the category-specific model as defined by Morningstar. Overall, minimum evidence is identified with respect to the Aggregate index. However, relative to their category-specific benchmark (the Barclays Sterling Corporate index), 25% of the Corporate bond funds exhibit significant and positive timing ability. This is most prominent when data from 2007-2016 is considered, as such including both the global financial crisis and the low rate environment.

7.1.3 False discoveries in fixed income

With many funds under analysis in this thesis, the conclusions are likely to be inhibited by bias inherent in the multiple hypothesis testing framework. Various methods are applied in this final empirical chapter to achieve further insights regarding the prevalence of outperformance and control for luck. The use of portfolios of funds in performance evaluation has the unfortunate implication of potentially masking the ability of managers in the tails of distributions. The first empirical analysis here applies the method of bootstrapping known as entire-cases resampling to the individual funds, allowing more granularity regarding the extreme performers (Fama and French 2010). Entire-cases resampling takes draws from the sample of both the dependent and independent variable, retaining the time-ordering. Results are supportive of the post-crisis improvements. Superior performance is identified from the 97th percentile for All funds, and from the 95th for Corporate and Diversified categories. However, the latter maintains a longer left tail than the Corporate funds, therefore indicating greater scope for significant underperformance. The results are robust to altering the minimum number of observations per fund and an alternative model.

To add robustness to the results and better control for luck, two versions of the false discovery rate (FDR) method are applied. The findings from the Barras, Scalliet, and Wermers (2010) approach find outperformance is isolated to the post-crisis period. Here, 6.2% of all funds are identified as having true ability. Furthermore, the left tail reduces from 42% over the whole period to 19% of funds being classified as exhibiting true underperformance. Corporate bond funds are the primary drivers of these results, with 10% of this category outperforming post-crisis. Characteristics attributable to the funds in each tail are also highlighted. Consistent with expectations and the results in Chapter 5, higher expense ratios are synonymous with the underperforming funds.

A commonly noted critique of the BSW (2010) method is that it has limited power to identify funds with non-zero alpha. There is an expanding volume of research in this area, aiming to add empirical refinements to address such biases. The Ferson and Chen (2019) methodology has been proposed as a means by which this can be achieved. Furthermore, the entire-cases resampling method (FF (2010)) that has already been applied in this chapter is used as a basis for their study. Therefore, it makes for an interesting and intuitive extension for this chapter. The results from doing so find the highest prevalence of good funds yet for this sample. The proportion of post-crisis period outperformance is 33% and 48% across All and Corporate funds respectively. However, a higher proportion of bad funds are identified relative to the BSW (2010) approach. The proportions in this respect are 67% across All funds and 52% for Corporate funds. Nonetheless, relative to what is more commonly observed in equities literature, the results from this sample of bond funds find suggest that there is greater scope for significant extreme performance.

7.1.4 Limitations, implications, and areas for further research

An unfortunate limitation of this research has been the availability of certain data for the full sample of UK bond mutual funds. For example, a complete time-series of expenses, total net assets, turnover, various other characteristics or portfolio holdings information would have been valuable. This would have allowed for construction of characteristics-based portfolios, in addition to those equally-weighted according to their Morningstar style-specific classifications.

Overall, the findings from this research have contributed to the fund performance literature in many respects. Gaps have been acknowledged; primarily with regards to the coverage of the UK bond market and the specification of models to conduct an evaluation of active management. From a review of the academic literature and industry developments it is evident that there are many research questions that remain open to be explored.

Although there is a wealth of literature directed towards refining methods to address the luck versus skill debate in fund performance, there is a distinct lack that specifically relates to bond funds. This thesis, to the best of my knowledge, is the first to apply the entire-cases resampling method of Fama and French (2010), and any version of the false discovery rate approach. In this case, both those attributable to Barras, Scaillet, and Wermers (2010) and Ferson and Chen (2019) have been implemented. Notable scope remains for extensions to continue in this line of research. For example, as an alternative to the method of FC (2019) of addressing bias, the refinements as proposed by BSW (2010) in response to the recent critique by Andrikogiannopoulou and Papakonstantinou (2016) would be interesting to test.

A relatively limiting factor that currently exists with regards to assessing skill in bond fund performance is the availability of models to do so. Recent literature has proposed various factors that have been found to increase explanatory power in terms of the cross-section of Corporate bond returns (Houweling (2017), Bai et al., (2019)). These are also diversifying with regards to tradition risk premia, for example equity market returns. This is highlighted as an interesting area for for continued research that going forward may help to shed further light on the sources of outperformance observed; such as that which has been evident throughout this thesis.

References

Andrikogiannopoulou, Angie and Filippos Papakonstantinou (2016). "Estimating Mutual Fund Skill: A New Approach". In: *Working Paper* August.

— (2019). "Reassessing False Discoveries in Mutual Fund Performance: Skill, Luck, or Lack of Power?" In: *Journal of Finance* 74.5, pp. 2667–2688.

Ang, Andrew and William Goetzmann (2009). *Evaluation of active management of the Norwegian government pension fund–global.* Tech. rep.

Aragon, George O and Wayne Ferson (2006). "Portfolio Performance Evaluation". In: *Foundations and Trends® in Finance* 2.2, pp. 83–190.

Ayadi, Mohamed A. and Lawrence Kryzanowski (2011). "Fixed-income fund performance: Role of luck and ability in tail membership". In: *Journal of Empirical Finance* 18.3, pp. 379–392.

Bai, Jennie, Turan G. Bali, and Quan Wen (2019). "Common risk factors in the cross-section of corporate bond returns". In: *Journal of Financial Economics* 131.3, pp. 619–642. ISSN: 0304405X. DOI: 10.1016/j.jfineco.2018.08.002. URL: https://doi.org/10.1016/j.jfineco.2018.08.002.

Barras, Laurent, Olivier Scaillet, and Russ Wermers (2010). "False discoveries in mutual fund performance: Measuring luck in estimated alphas". In: *Journal of Finance* 65.1, pp. 179–216.

Bektić, Demir, Josef-stefan Wenzler, Michael Wegener, and Dirk Schiereck (2016). "Extending Fama-French Factors to Corporate Bond Markets". In: *Working Paper*.

Berk, Jonathan and Richard Green (2004). "Mutual Fund Flows and Performance in Rational Markets". In: *Journal of Political Economy* 112.6, pp. 1269–1295.

Blake, Christopher R, Edwin J Elton, and Martin J Gruber (1993). "The Performance of Bond Mutual Funds". In: *The Journal of Business* 66.3, pp. 371–403.

Blake, David, David Blake, Allan Timmermann, and Allan Timmermann (1998). "Mutual Fund Performance: Evidence from the UK". In: *European Finance Review* 2.1995, pp. 57–77.

Blake, David, Tristan Caulfield, Christos Ioannidis, and Ian Tonks (2017). "New Evidence on Mutual Fund Performance: A Comparison of Alternative Bootstrap Methods". In: *Journal of Financial and Quantitative Analysis* 52.3, pp. 1279–1299.

Breedon, Francis, Jagjit S. Chadha, and Alex Waters (2013). "The financial market impact of UK quantitative easing". In: *Oxford Review of Economic Policy* 28.4, pp. 702–728.

Brennan, Michael and Eduardo Schwartz (1979). "A Continuous Time Approach To The Pricing of Bonds". In: *Journal of Banking & Finance* 3, pp. 133–155.

Brooks, Jordan, Tony Gould, and Scott Anthony Richardson (2019). "Active Fixed Income Illusions". In: *SSRN Electronic Journal*, pp. 1–27.

Brooks, Jordan and Scott Richardson (2018). "Style investing in fixed income". In.

Carhart, Mark M. (1997). "On persistence in mutual fund performance". In: *Journal of Finance* 52.1, pp. 57–82.

Chen, Yong, Michael Cliff, and Haibei Zhao (2017). "Hedge Funds: The Good, the Bad, and the Lucky". In: *Journal of Financial and Quantitative Analysis* 52.3, pp. 1081–1109.

Chen, Yong, Wayne Ferson, and Helen Peters (2010). "Measuring the timing ability and performance of bond mutual funds". In: *Journal of Financial Economics* 98.1, pp. 72–89.

Chen, Zhiwu and Peter Knez (1996). "Portfolio Performance Measurement: Theory and Applications". In: *The Review of Financial Studies.*

Chen., Yong and Nan Qin (2014). "The Behavior of Investor Flows in Corporate Bond Mutual Funds". In: *Working paper.*

Chordia, Tarun, Asani Sarkar, and Avanidhar Subrahmanyam (2005). "An empirical analysis of stock and bond market liquidity". In: *Review of Financial Studies* 18.1, pp. 85–129.

Cici, Gjergji, Scott Gibson, and John J Merrick (2011). "Missing the marks? Dispersion in corporate bond valuations across mutual funds $". In: *Journal of Financial Economics* 101.1, pp. 206–226.

Cici, Gjergji, Scott Gibson, Source The, Quantitative Analysis, No February, Gjergji Cici, and Scott Gibson (2012). "The Performance of Corporate Bond Mutual Funds : Evidence Based on Security-Level Holdings". In: *The Journal of Financial and Quantitative Analysis* 47.1, pp. 159–178.

Comer, George (2005). "Evaluating Bond Fund Sector Timing Skill". In: *Working Paper* August.

Comer, George and Javier Rodriguez (2013). "A comparison of corporate versus government bond funds". In: *Journal of Economics and Finance* 37, pp. 495–510.

Connor, Gregory and Robert a. Korajczyk (1986). "Performance measurement with the arbitrage pricing theory". In: *Journal of Financial Economics* 15, pp. 373–394.

Cornell, Bradford and Kevin Green (1991). "The Investment Performance of Low-Grade Bond Funds". In: *The Journal of Finance* 46.1, pp. 29–48.

Correia, Maria, Scott Richardson, and Irem Tuna (2012). "Value investing in credit markets". In: *Review of Accounting Studies* 17.3, pp. 572–609.

Cox, John C, Jonathan E Ingersoll, and Stephen A Ross (1985). "An Intertemporal General Equilibrium Model of Asset Prices". In: *Econometrica* 53.2, pp. 363–384.

Cremers, Martijn, Miguel Ferreira, Pedro Matos, and Laura Starks (2016). "Indexing and active fund management: International evidence". In: *Journal of Financial Economics* 120.3, pp. 539–560.

Cuthbertson, Keith, Dirk Nitzsche, and Niall O'Sullivan (2012). "False Discoveries in UK Mutual Fund Performance". In: *European Financial Management* 18.3, pp. 444–463.

Cuthbertson, Keith, Dirk Nitzsche, and Niall O Sullivan (2008). "UK mutual fund performance : Skill or luck?" In: 15, pp. 613–634.

Dannhauser, Caitlin D. (2017). "The impact of innovation: Evidence from corporate bond exchange-traded funds (ETFs)". In: *Journal of Financial Economics* 125.3, pp. 537–560.

Detzler, Miranda Lam (Aug. 1999). "The performance of global bond mutual funds". In: *Journal of Banking & Finance* 23.8, pp. 1195–1217.

Dietz, Peter, H Russell Fogler, and Donald J Hardy (1980). "The Challenge of Analyzing Bond Portfolio Returns". In: *Journal of Portfolio Management* 6.3, pp. 53–58.

Dietz, Peter, H Russell Fogler, and Anthony U Rivers (1981). "Duration, nonlinearity and bond portfolio performance". In: *Journal of Portfolio Management* 7(3), pp. 37–41.

Domian, Dale L and William Reichenstein (2008). "Returns-Based Style Analysis of High-Yield Bonds". In: *The Journal of Fixed Income* 17.4, pp. 72–73.

Dor, Ben Arik and Zhe Xu (2011). "Fallen Angels: Characteristics, Performance, and Implications for Investors". In: *The Journal of Fixed Income*, pp. 33–59. ISSN: 10598596. DOI: 10.3905/jfi.2011.20.4.033.

Elliott, David and Menno Middeldorp (2016). *What explains the fall in sterling corporate bond issuance?* URL: https://bankunderground.co.uk/2016/04/01/what-explains-the-fall-in-sterling-corporate-bond-issuance/. (accessed: 01.07.2020).

Elton, Edwin J, Martin J Gruber, and Christopher R Blake (1999). "Common Factors in Active and Passive Portfolios". In: *European Finance Review* 3.June 1996, pp. 53–78.

— (1995). "Fundamental Economic Variables, Expected Returns, and Bond Fund Performance". In: *The Journal of Finance* 50.4, pp. 1229–1256.

Fabozzi, Frank (2012). *The Handbook of Fixed Income Securities*. Vol. 8th. McGraw-Hill Education.

Fama, E (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work ". In: *The Journal of Finance* 25.2, pp. 383–417.

Fama, E and K French (2015). "A five-factor asset pricing model". In: *Journal of Financial Economics* 116.1, pp. 1–22.

Fama, E and J Macbeth (1973). “Risk , Return , and Equilibrium : Empirical Tests”. In: *The Journal of Political Economy* 81.3, pp. 607–636.

Fama, E. and K. French (2012). “Size , value , and momentum in international stock returns”. In: *Journal of Financial Economics* 105.3, pp. 457–472.

Fama, Eugene and Kenneth French (2010). “Luck versus Skill in the cross-section of mutual fund returns”. In: *Journal of Finance* 65.5, pp. 1915–1947.

Feroli, Michael, Anil K. Kashyap, Kermit L. Schoenholtz, and Hyun Song Shin (2014). “Market Tantrums and Monetary Policy”. In: *SSRN Electronic Journal* 1.February. ISSN: 1556-5068. DOI: 10.2139/ssrn.2409092. URL: http://papers.ssrn.com/abstract=2409092.

Ferruz Agudo, Luis and José Luis Sarto Marzal (2004). “An analysis of Spanish investment fund performance: Some considerations concerning Sharpe’s ratio”. In: *Omega* 32.4, pp. 273–284.

Ferson, Wayne and Min S Kim (2012). “The factor structure of mutual fund flows”. In: *International Journal of Portfolio Analysis and Management* 1.2, p. 112.

Ferson, Wayne, Darren Kisgen, and Tyler Henry (2006). “Evaluating government bond fund performance with stochastic discount factors”. In: *Review of Financial Studies* 19.2, pp. 423–455.

Ferson, Wayne and Meijun Qian (2010). “Conditional Performance Evaluation , Revisited”. In.

Ferson, Wayne and Rudi W Schadt (1996). “Measuring Fund Strategy and Performance in Changing Economic Conditions”. In: *The Journal of Finance* 51.2, pp. 425–461.

Fox, John (2002). “Bootstrapping Regression Models”. In: *An R and S-PLUS Companion to Applied Regression.*

Frieda, April and Scott Richardson (2016). “Systematic Credit Investing”. In: *Alternative Investment Analyst Review.*

Gebhardt, William R., Soeren Hvidkjaer, and Bhaskaran Swaminathan (2005). “The cross-section of expected corporate bond returns: Betas or characteristics?” In: *Journal of Financial Economics* 75.1, pp. 85–114.

Gibbons, M, Stephen A Ross, and Jay Shanken (1989). “A Test of the Efficiency of a Given Portfolio”. In: *Econometrica* 57.5, pp. 1121–1152.

Goldstein, Itay, Hao Jiang, and David T. Ng (2017). “Investor flows and fragility in corporate bond funds”. In: *Journal of Financial Economics* 126.3, pp. 592–613.

Grose, Chris, Apostolos Dasilas, and Christos Alexakis (2014). “Journal of International Financial Markets , Institutions & Money Performance persistence in fixed interest funds : With an eye on the post-debt crisis period”. In: *”Journal of International Financial Markets, Institutions & Money”* 33, pp. 155–182.

Gultekin, N Bulent and Richard J Rogalski (1985). “Government Bond Returns, Measurement of Interest Rate Risk, and the Arbitrage Pricing Theory”. In: *Journal of Finance* 40.1, pp. 43–61.

Gutierrez Jr, Roberto C, William F Maxwell, and Danielle Xu (2009). “On Economies of Scale and Persistent Performance in Corporate-Bond Mutual Funds”. In.

Harvey, Campbell and Yan Liu (2020). “Luck versus Skill in the Cross-Section of Mutual Fund Returns: Reexamining the Evidence”. In: *SSRN Electronic Journal*.

Harvey, Campbell R and Yan Liu (2019). “A Census of the Factor Zoo *”. In: *SSRN Working Paper*, pp. 1–7. URL: `https://ssrn.com/abstract=3341728`.

Henriksson, Roy D and Robert C Merton (1981). “On Market Timing and Investment Performance . II . Statistical Procedures for Evaluating Forecasting Skills”. In: *Journal of Business* 54.4, pp. 513–533.

Houweling, Patrick and Jeroen Van Zundert (2017). “Factor investing in the corporate bond market”. In: *Financial Analysts Journal* 73.2, pp. 100–115.

Huang, Jing-zhi and Ying Wang (2014). “Timing Ability of Government Bond Fund Managers : Evidence from Portfolio Holdings”. In: *Management Science*.

Huij, Joop and Jeroen Derwall (2008). “”Hot Hands” in bond funds”. In: *Journal of Banking and Finance* 32.4, pp. 559–572.

Ilmanen, Antti (1995). “Time-Varying Expected Returns in International Bond Markets”. In: *Journal of Finance* 50.2, pp. 481–506.

Israel, Ronen, Diogo Palhares, and Scott Richardson (2018). “Common factors in corporate bond returns”. In: *Journal of Investment Management* 16.2, pp. 17–46.

Jensen, Michael C (1968). “Problems in the selection of security Portfolios - The Performance of Mutual Funds in the Period 1945-1964”. In: *Journal of Finance* 23.2, pp. 389–416.

Jostova, Gergana, Stanislava Nikolova, Alexander Philipov, and Christof W. Stahel (2013). “Momentum in corporate bond returns”. In: *Review of Financial Studies* 26.7, pp. 1649–1693.

Koijen, Ralph S.J., Tobias J. Moskowitz, Lasse Heje Pedersen, and Evert B. Vrugt (2018). “Carry”. In: *Journal of Financial Economics* 127.2, pp. 197–225.

Korajczyk, Robert A and Claude J Viallet (1989). “An Empirical Investigation of International Asset Pricing”. In: *The Review of Financial Studies* 2.4, pp. 553–585.

Kosowski, Robert (2006). “Do Mutual Funds Perform When it Matters Most to Investors? US Mutual Fund Performance and Risk in Recessions and Expansions”. In: *SSRN Electronic Journal* August.

Kosowski, Robert, Narayan Y. Naik, and Melvyn Teo (2007). “Do hedge funds deliver alpha? A Bayesian and bootstrap analysis”. In: *Journal of Financial Economics* 84.1, pp. 229–264.

Kosowski, Robert, Allan Timmermann, Russ Wermers, and Hal White (2006). “Can mutual fund ”stars” really pick stocks? New evidence from a bootstrap analysis”. In: *Journal of Finance* 61.6, pp. 2551–2595.

Koutmos, Dimitrios, Bochen Wu, and Qi Zhang (2020). “In search of winning mutual funds in the Chinese stock market”. In: *Review of Quantitative Finance and Accounting* 54.2, pp. 589–616.

Kryzanowski, Lawrence and Mohamed A. Ayadi (2005). *Chapter 10: A Stochastic Discount Factor-based Appraoch for Fixed-income Mutual Funds*, pp. 193–226.

Leite, Paulo, Orlando Faria, and Manuel Rocha Armada (2016). “The impact of the Eurozone sovereign debt crisis on bond fund performance persistence : Evidence from a small market”. In: *Investment Analysts Journal* 3523.November, pp. 32–45.

Levy, M. and R. Roll (2010). “The Market Portfolio May Be Mean/Variance Efficient After All”. In: *Review of Financial Studies* 23.6, pp. 2464–2491.

Lintner, John (1965). “Security Prices, Risk, and Maximal Gains from Diversification”. In: *Journal of Finance* XX.4, pp. 587–615.

Lipton, Amy F and Richard J Kish (2010). “Robust performance measures for high yield bond funds”. In: *The Quarterly Review of Economics and Finance* 50, pp. 332–340. DOI: 10.1016/j.qref.2010.03.001.

Litterman, Robert B and Josè Scheinkman (1991). “Common Factors Affecting Bond Returns”. In: 1, pp. 54–61.

Lo, Andrew W, Andrew W Lo, and Harris Group (2002). “The Statistics of Sharpe Ratios”. In.

Markowitz, H (1952). “Portfolio selection*”. In: *The Journal of Finance* 7.1, pp. 77–91.

Moneta, Fabio (2015). “Measuring bond mutual fund performance with portfolio characteristics”. In: *Journal of Empirical Finance* 33, pp. 223–242.

Newey, Whitney K and Kenneth D West (1987). “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix”. In: *Econometrica* 55.3, pp. 703–708.

Ng, Kwok Yuen and Bruce D. Phelps (2011). “Capturing credit spread premium”. In: *Financial Analysts Journal* 67.3, pp. 63–75. ISSN: 0015198X. DOI: 10.2469/faj.v67.n3.4.

Palhares, Diogo and Scott Richardson (2018). “(Il) liquidity premium in credit markets : A myth?” In: *Working Paper* Il, pp. 1–45.

Rohleder, Martin, Hendrik Scholz, and Marco Wilkens (2012). “Bond Fund Disappearance: What’s Return Got to Do with it?” URL: http://www.ssrn.com/abstract=1864423.

Ross, Stephen A. (1976). “The arbitrage theory of capital asset pricing”. In: *Journal of Economic Theory* 13.3, pp. 341–360.

Sharpe, William F (1964). “Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk”. In: *Journal of Finance* XIX, pp. 425–442.

Silva, Florinda, Maria C. Cortez, and Manuel R. Armada (2003). “Conditioning information and European bond fund performance”. In: *European Financial Management* 9.2, pp. 201–230.

Storey, John D (2002). “A Direct Approach to False Discovery Rates”. In: *Journal of the Royal Statistical Society* 64.3, pp. 479–498.

Treynor, J and K Mazuy (1966). “Can mutual funds outguess the market?” In: *Harvard Business Review* 44.4, pp. 131–136.

Vasicek, Oldrich (1977). “An equilibrium characterization of the term structure”. In: *Journal of Financial Economics* 5, pp. 177–188.

Zhao, Xinge (2005). “Determinants of Flows into Retail Bond Funds Determinants of Flows into Retail Bond Funds”. In: *Financial Analysts Journal* 61.4, pp. 47–59.