Essays on Asset Pricing Factor Models

Der Wirtschaftswissenschaftlichen Fakultät der
Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades

Doktor(in) der Wirtschaftswissenschaften
- Doctor rerum politicarum -

genehmigte Dissertation

von

M.Com. Thuy Duong Dang

geboren am 11.01.1992 in Haiphong

2023

Referent: Prof. Dr. Marcel Prokopczuk, Leibniz Universität Hannover

Korreferent: Prof. Dr. Maik Dierkes, Leibniz Universität Hannover

Tag der Promotion: 11.12.2023

Eigenständigkeitserklärung – Declaration of Original Authorship

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbständig ohne Hilfe Dritter verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, wurden unter Angabe der Quelle kenntlich gemacht.

I hereby confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Thuy Duong Dang

Abstract

This dissertation investigates the utilization of factor models to measure performance in corporate bond markets, identifies an optimal factor model for corporate bond returns, and finally provides a comprehensive analysis of factor pricing and market integration across asset classes. Chapter 1 introduces the main concepts and delivers an overview of the following chapters.

Chapter 2 seeks to answer the question "which factor model do investors in corporate bonds use?" by tracking investors' decisions to invest in actively managed corporate bond mutual funds with a revealed preference approach. The main result is that *all* bond factor models are dominated by the simple Sharpe ratio and Morningstar ratings. For all major corporate bond mutual fund styles, the Sharpe ratio explains fund flows better than alphas from bond factor models. Since the Sharpe ratio (and to some extent also Morningstar ratings) can be easily manipulated in bond markets, these findings have potentially severe implications for all market participants.

Going a step further, Chapter 3 addresses the following important questions, from both an academic and a practitioner's perspective: What are important drivers of corporate bond returns? What should be a benchmark model for pricing and investing in corporate bond markets? The central finding is that factors related to carry, duration, equity momentum, and the term structure are the most important risk factors in corporate bond markets. From a large set of factor candidates for corporate bond returns, we condense an optimal model with a two-step approach. First, we filter out factors that do not systematically move bond prices. Second, we use a Bayesian model selection approach to determine the optimal, parsimonious model. Many prominent factors do not move prices, or are redundant. We document the new model's good performance compared to that of existing models in time-series and cross-sectional tests and analyze the economic drivers of the factors.

While Chapter 2 and Chapter 3 focus on corporate bonds, the study conducted in Chapter 4 extends the understanding to a bigger picture of factor pricing and market integration across asset classes. Factor models specializing in one asset class have limited pricing power for other asset classes. Thus, we reject perfect market integration. However, an optimal integrated factor model across asset classes can effectively characterize the returns of multiple asset classes and provide a useful benchmark for multi-asset, multi-factor investing. The optimal model includes several equity and corporate bond factors, suggesting the presence of multiple systematic return drivers. Despite this, there appears to be some degree of cross-market linkages, as the optimal model does not require factors from all asset classes.

Finally, Chapter 5 concludes and outlines possible future directions for research.

Keywords: Bond factor models, Sharpe ratio, bond mutual funds, investor flows, performance evaluation, flow—performance sensitivity, corporate bonds, risk factors, model comparison, factor models, asset classes, market integration.

Zusammenfassung

Diese Dissertation untersucht die Verwendung von Faktormodellen zur Messung der Performance in Unternehmensanleihen, identifiziert ein optimales Faktormodell für die Renditen von Unternehmensanleihen und liefert schließlich eine umfassende Analyse der Faktorpreisbildung und der Marktintegration über verschiedene Anlageklassen hinweg. Kapitel 1 führt in die wichtigsten Konzepte ein und gibt einen Überblick über die folgenden Kapitel.

In Kapitel 2 wird versucht, die folgende Frage zu beantworten: Welches Faktormodell verwenden Investoren in Unternehmensanleihen? Dazu werden die Entscheidungen der Investoren in aktiv verwaltete Investmentfonds für Unternehmensanleihen zu investieren, mit einem Ansatz der offengelegten Präferenz verfolgt. Unser Hauptergebnis ist, dass Anleihefaktormodelle von der einfachen Sharpe Ratio und Morningstar Ratings dominiert werden. Für alle wichtigen Arten von Investmentfonds für Unternehmensanleihen erklärt die Sharpe Ratio die Fondsströme besser als die Alphas von Anleihefaktormodellen. Da die Sharpe Ratio (und bis zu einem gewissen Grad auch die Morningstar-Ratings) auf Anleihemärkten leicht manipuliert werden können, haben unsere Ergebnisse potenziell weitreichende Folgen für alle Marktteilnehmer.

In einem weiteren Schritt werden in Kapitel 3 die folgenden wichtigen

Fragen sowohl aus akademischer als auch aus praktischer Sicht behandelt: Was sind wichtige Faktoren für die Rendite von Unternehmensanleihen? Wie sollte ein Benchmark-Modell für die Preisbildung und Investitionen auf den Märkten für Unternehmensanleihen aussehen? Das zentrale Ergebnis ist, dass die wichtigsten Risikofaktoren auf den Märkten für Unternehmensanleihen die Carry, Duration, Equity-Momentum und Laufzeitstruktur Faktoren sind. Aus einer großen Anzahl von Faktoren, die für die Renditen von Unternehmensanleihen in Frage kommen, ermitteln wir ein optimales Modell mit einem zweistufigen Ansatz. Zunächst filtern wir Faktoren heraus, die die Anleihekurse nicht systematisch beeinflussen. Als Zweites verwenden wir einen Bayes'schen Modellauswahlansatz, um das optimale Modell mit einer geringen Anzahl an Faktoren zu bestimmen. Viele prominente Faktoren beeinflussen die Kurse nicht oder sind redundant. Wir dokumentieren die gute Performance des neuen Modells im Vergleich zu den bestehenden Modellen in Zeitreihen- und Querschnittstests und analysieren die wirtschaftlichen Triebkräfte der Faktoren.

Während sich Kapitel 2 und Kapitel 3 auf Unternehmensanleihen konzentrieren, erweitert die in Kapitel 4 durchgeführte Studie das Verständnis mit einer breiteren Analyse der Faktoren und der Marktintegration über verschiedene Anlageklassen hinweg. Faktormodelle, die für eine Anlageklasse spezialisiert sind, haben einen begrenzten Erklärgehalt für Renditen anderer Anlageklassen. Daher lehnen wir die Hypothese ab, dass die Märkte perfekt integriert sind. Ein optimales integriertes Faktormodell über alle Anlageklassen hinweg kann jedoch die Renditen mehrerer Anlageklassen effektiv charakterisieren und eine nützliche Benchmark für Multi-Asset- und Multi-Faktor-Investitionen bieten. Das optimale Modell enthält mehrere Faktoren für Aktien und Unternehmensanleihen, was auf das Vorhandensein mehrerer systematischer Renditetreiber hindeutet. Dennoch scheint es ein gewisses Maß an marktübergreifenden Verknüpfungen zu geben, da das optimale Modell nicht Faktoren aus allen Anlageklassen erfordert.

In Kapitel 5 werden schließlich Schlussfolgerungen gezogen und mögliche zukünftige Forschungsrichtungen skizziert.

Schlüsselwörter: Anleihefaktormodelle, Sharpe Ratio, Anleihenfonds, Anlegerströme, Leistungsbewertung, Sensitivität zwischen Strom und Leistung, Unternehmensanleihen, Risikofaktoren, Faktormodelle, verschiedene Anlageklassen, Modellvergleich, Marktintegration.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In finance, factor models are well-established with a wide range of applications in both theory and practice. They are used for a variety of purposes, including return prediction, performance evaluation, anomaly assessment, and portfolio construction, among others.

Asset pricing theories have developed from a simple one-factor model to more comprehensive multi-factor models. In the early 1960s, the Capital Asset Pricing Model (CAPM) is one cornerstone laid by Treynor (1961), Sharpe (1964), Lintner (1965), and Mossin (1966). The key insight of the model is that an asset's expected return is determined by how much it is exposed to one fundamental risk factor: the market factor. An asset that has a large exposure to this factor is risky, because it performs poorly when the market goes down and well when the market goes up, resulting in high systematic variation in future payoffs. Conversely, an asset with low exposure to the market factor is less risky because its future payoffs will be less systematically volatile. Market participants typically demand a higher risk premium as compensation for investing in high-beta assets than low-beta assets.

To resolve the theoretical critiques and poor empirical performance of the CAPM, in the 1970s, more general models, evolving from a one-factor to a multi-factor setting, came up in the literature. The most prominent examples are the Intertemporal Capital Asset Pricing Model (ICAPM) of Merton (1973) and the Arbitrage Pricing Theory (APT) of Ross (1976), which allow to include other sources of risk in addition to the market risk. Building on these studies, Fama & French (1993) specify a reduced-form three-factor model for U.S. equities, which became very celebrated. In Fama & French (2015), they extend their model to five factors. For equities alone, hundreds of factors have been proposed (Harvey, Liu, & Zhu, 2016).

As a result, factor-based investing has become increasingly popular among practitioners such as asset managers and has been widely adopted in equity markets. The underlying idea is to capture equity risk factors, such as size, value, etc., and to earn the corresponding risk premiums. Therefore, when considering any active strategy, investors should have a clear understanding of the sources of expected returns and the underlying risk exposures of the investment portfolios.

Investors should pay attention to factor-related returns when assessing managers' skill. Active fund managers should be credited not only for offering exposure to common risk factors to harvest factor premiums that could be achieved through passive investments, but also for their skills to actively seek to produce alpha.

Despite the size and the importance of the corporate bond market, it has received far less research attention compared to that for equities. Factor models and factor-based investing should not only be relevant for stocks, but for bonds as well.

Chapter 2 investigates how investors evaluate the performance of corporate bond mutual funds. More specifically: Which factor model do they use to measure performance? Do investors even use factor models, or do they rely on simpler performance measures?

To address these questions, we employ a revealed preference approach as in Barber, Huang, & Odean (2016) and Berk & Van Binsbergen (2016). Technically, we conduct a flow–performance horse race test to infer which performance measures corporate bond mutual fund investors use when making investment decisions. Performance measures used to compare funds can vary widely from the simple raw returns to the alphas from a complex multi-factor model for returns. In our analysis, we consider a range of single- and multi-factor models as well as ratios that are popular in the asset pricing literature, all of which investors might reasonably use.

The central finding is that the Sharpe ratio and Morningstar ratings consistently explain corporate bond fund flows significantly better than the raw return or any alpha from single- or multi-factor models. It thus seems that most investors do not use any factor model at all.

The main contribution of Chapter 2 is thus a systematic analysis of which measures investors in corporate bond markets use to assess performance. To the best of my knowledge, this chapter is the first to comprehensively analyze this and the questions stated above.

The finding of investors' reliance on Sharpe ratio and Morningstar ratings to assess funds' performance carries important implications for both fund managers and investors. Fund managers have more incentives to manipulate these measures (e.g. smoothing returns by holding illiquid assets) to mislead investors' fund selections. This opportunistic behavior of corporate bond fund manager can create trading opportunities for active traders by exploiting net asset value misvaluations. Due to the mismatch between the illiquidity of the corporate bond funds' underlying assets and the liquidity they offer to investors, the gains of the active traders are matched with the losses suffered by buy-and-hold fund investors and poses a potentially serious threat of fund runs.

The finding that investors tend to use simple measures instead of factor models in Chapter 2 may be not surprising because one can argue that corporate bond investors are likely much less aware of factor models than investors in equity markets given that research in factors and factor-investing for bonds is still in the nascent stage. Which set of factors can span the efficient frontier of corporate bonds and thus should serve as the benchmark model for investors in corporate bond markets is still an open question.

Both stocks and corporate bonds are contingent claims on the value of the same underlying firm. However, bond markets have some distinguishable features from equity markets, indicating potential market segmentation. In fact, Chordia, Goyal, Nozowa, Subrahmanyam, & Tong (2017) and Choi & Kim (2018) report evidence of discrepancies in risk premiums between corporate bonds and stocks. Therefore, it is essential to investigate the cross-section of corporate bond returns by also using factors constructed based on corporate bond characteristics, rather than only relying on the available commonly used factors from the equity market. Recently, inspired by the way characteristics have been used for constructing equity factors, there is a rapidly growing number of studies devoted to discover factors and factor models for corporate bond returns, paving the way for factor-based investing in corporate bond markets.

Given the proliferation of factors, it is crucial from both an academic and a practitioner's perspective to know which are genuine risk factors in corporate bond markets, and which factors do not provide incremental information about returns, and thus are redundant. Chapter 3 addresses the following questions: Do we really need all factors proposed in the corporate bond literature to explain the cross-section of returns? Which factors move corporate bond prices systematically? What set of factors overall best describes corporate bond returns? Are some factors redundant relative to others? To what extent does each needed factor play a role in explaining time-series and cross-sectional variation in corporate bond returns? Which economic forces drive the factors?

From a collection of, from our point of view, the 23 most prominent factor candidates for corporate bond returns in the literature, we use a two-step approach to uncover an optimal model. First, we screen out factors that do not systematically move bond prices by checking the necessary condition of the factor identification protocol proposed by Pukthuanthong, Roll, & Subrahmanyam (2019). Only the viable factors that pass the identification protocol are eligible for model selection in the next step. Second, we employ the Bayesian marginal likelihood model comparison approach recently developed by Barillas & Shanken (2018) and Chib, Zeng, & Zhao (2020) (BS-CZZ) to identify the optimal, parsimonious model from all the models that are possible combinations of these factors.

The main finding that emerges from our analysis is that the best factor model for corporate bond returns is based on the combination of carry, duration, stock momentum, and term structure factors. We show that many prominent factors do not move prices, or are redundant. The optimal model outperforms the existing models proposed in the literature and has good performance in both time-series and cross-sectional tests with test assets.

The main contribution of Chapter 3 is a systematic analysis of the factors proposed in the corporate bond pricing literature. This chapter helps academics and practitioners separate useful factors from redundant ones and search the growing list of bond factors for a set that collectively best explains the differences in corporate bond returns. Based on this, we can build an "optimal" corporate bond factor model. The winning factor model can be used as a benchmark model for future research, for investors in corporate bond markets to implement factor-investing strategies, and to evaluate performance. To the best of my knowledge, this chapter is the first to systematically compare a broad set of common and recently proposed factors and factor models for corporate bonds.

So far, Chapter 2 and Chapter 3 focus on the corporate bond market. The findings in these first two studies of the dissertation contribute to the growing literature for corporate bonds in an attempt to fill the gaps and catch up with the rich literature for equities. However, corporate bonds are just one type of assets. The last chapter looks at a bigger picture of factor pricing and market integration across multiple major asset classes.

Under the law of one price (LOOP), it is possible to construct a single-factor or multi-factor proxy for the stochastic discount factor (SDF) that prices all assets (Hansen & Jagannathan, 1991; Cochrane, 2009; Kozak, Nagel, & Santosh, 2018). In a factor model, the SDF is expressed as a linear function of a small number of dominant drivers of returns. Factor models are prevalent in empirical asset pricing because they are a useful way to provide a concise summary of the cross-section of asset returns. Models for equities such as the three-factor of Fama & French (1993) and various extensions (e.g. Carhart, 1997; Fama & French, 2015) are popular among academics and practitioners.

To construct well-diversified portfolios, investors should consider as broad a range of assets as possible when allocating capital. Factor investing simplifies the problem of portfolio construction by shrinking the asset space. Factor models narrow the search for investment strategies among the universe of individual assets to the more manageable task of finding the optimal risk-return trade-off among a handful number of factors.

Despite the conceptual elegance of factor models, two factual issues exist. First, there is a plethora of factors. To describe the hundreds of factors discovered in equity research, Cochrane (2011) coined the term “factor zoo”. The second problem is that most academic studies conventionally examine different markets in isolation, and develop asset-class-specific factor models for stocks, bonds, commodities, and the like. Therefore, from both a theoretical and a practical perspective, it is worthwhile to search for a handful of factors that span the Markowitz (1952) mean-variance-efficient frontier and capture the returns of all assets.

Important questions related to this endeavor, to the best of our knowledge, have not yet been fully resolved, for example: To what extent can the factors of different asset classes also price the assets of others? What is the degree of integration between the different asset classes? What does an optimal (empirical) stochastic discount factor across all asset classes look like? What is an appropriate benchmark model for portfolios of global securities across asset classes?

The objectives of Chapter 4 are therefore twofold. First, we comprehensively examine the prominent traded factors proposed in the asset pricing literature for various individual asset classes and investigate the extent of market integration based on their explanatory power across other asset classes. Second, we attempt to identify an integrated empirical model based on a sparse number of risk factors that spans and explains returns across multiple asset classes.

The first main goal of this chapter is to investigate the extent of market integration. We examine this through the lens of the pricing power of factor models from one asset class for others. We find that factor models that specialize in one asset class typically have difficulty pricing the factors from other asset classes. We therefore reject perfect integration. There appear to be multiple underlying systematic risk drivers across asset classes and markets. However, we also detect some cross-market linkages.

These findings further motivate us to pursue the second major goal of this chapter, which is to find an optimal integrated factor model that can describe returns across asset classes. To avoid creating high-dimensional factor models, we focus on the best factors for each of seven major asset classes when building the combined model by again using the BS–CZZ method. The optimal model consists of a total of eight factors, including the U.S. equity market, the size, management, and quality-minus-junk factors for international equities, the carry and equity momentum factors for corporate bonds, the currency momentum factor, and the equity index carry factor.

The optimal unified model performs quite well across asset classes. It subsumes a long list of factors and performs on a par with the best single-asset-class-specialized models in pricing assets across different asset classes. The top integrated model achieves high in-sample and out-of-sample Sharpe ratios. Factors from equities and corporate bonds prove to be the most important. In addition, the fact that not all asset classes from which factors are needed to build the integrated model suggests the presence of some degree of cross-market linkages.

The major contribution of Chapter 4 is to provide an integrated view of asset pricing. Moving beyond the common practice of analyzing individual asset classes in isolation, this chapter uncovers an optimal factor model that can span the multi-asset return space. The proposed integrated model can serve as a benchmark for future research in pricing securities across different asset classes, and as a useful guide for investors to exploit factor-based investing through a multi-asset, multi-factor lens.

This dissertation proceeds as follows. Chapter 2 investigates the use of factor models to measure performance in corporate bond markets by inferring from the sensitivity of investors flows of actively managed corporate bond mutual funds to different performance measures. Chapter 3 identifies an optimal factor model for corporate bond returns. Chapter 4 provides a comprehensive analysis of factor pricing and market integration across asset classes. Finally, Chapter 5 summarizes the main findings of this dissertation and suggests several lines for future research.

For reasons of improved readability, especially of the separate parts constituting the complete thesis, each chapter is self-contained. This means, variables and acronyms are redefined in each chapter. Whenever possible, notations are consistent throughout the dissertation in order to facilitate the reading.

Chapter 2

How Do Corporate Bond Investors Measure Performance? Evidence from Mutual Fund Flows*

2.1 Introduction

Corporate bonds are an important, yet underresearched asset class. While the total market is somewhat smaller than for equities, the annual issuance of corporate bonds is on a significantly larger scale (by both value and number of issues) than that of equity for U.S. corporations: for example, in 2020, there were 2,097 corporate bond issues totaling approximately $2.3

*This chapter is based on the Article "How Do Corporate Bond Investors Measure Performance? Evidence from Mutual Fund Flows" authored by Thuy Duong Dang, Fabian Hollstein and Marcel Prokopczuk, Journal of Banking and Finance, Volume 142, 106553.

trillion compared to 1,073 common stock issues totaling $343.7 billion.[1]

In this chapter, we study performance measurement in corporate bond markets. We address the following research questions: How do investors evaluate corporate bond mutual funds? Which factor model do they use to measure the performance? Do investors even use factor models, or do they rely on simpler performance measures?

We adopt a revealed preference approach as in Barber et al. (2016) and Berk & Van Binsbergen (2016). Mutual funds provide a suitable empirical setting for this research because we can observe both performance measures and aggregate investment decisions on a timely and frequent basis. Recently, the assets under management by fixed-income mutual funds have experienced significant growth. Through several market turmoil periods over most of the past decade, bond mutual funds have experienced net inflows, while equity funds have continuously experienced net outflows.[2] Bond mutual funds have thus become an important investment vehicle, in particular for individual investors, who seek exposure to bond markets.

In our empirical analysis, we perform a flow–performance horse race test to infer which performance measures corporate bond mutual fund investors use when making capital allocation decisions. At one extreme, investors may simply use raw returns to rank funds. At the other extreme, they may compare fund performance based on the alpha from a multi-factor model for returns. Given the substantial uncertainty about corporate bond factor models, we measure performance using a range of single- and multi-factor models as well as ratios that are commonly found in the asset pricing literature, all of which investors might reasonably employ. For our main

[1] SIFMA Fact Book 2021, sources: Bloomberg, Refinitiv, Dealogic. Available at https://www.sifma.org/wp-content/uploads/2021/07/CM-Fact-Book-2021-SIFMA.pdf.

[2] See Investment Company Institute Fact Book (2021). Available at https://www.ici.org/system/files/2021-05/2021_factbook.pdf.

analysis, we use the best-known and most widely used measures and models, including: the raw return, the Sharpe ratio, a single-factor model with a bond market index, a two-factor model with a stock and a bond market index, the Bekaert & De Santis (2021) three-factor model, the Elton, Gruber, & Blake (1995) four-factor model, and the Fama & French (1993) five-factor model for bonds. We conduct our tests while controlling for well-known predictor variables such as lagged fund flows, expense ratios, fund size, age, and Morningstar ratings.

Our main contribution is thus a systematic analysis of which measures investors in corporate bond markets use to assess performance. To the best of our knowledge, we are the first to comprehensively analyze this and the questions stated above.

In the first part of the empirical analysis, we use Morningstar ratings as the main control variables since Ben-David, Li, Rossi, & Song (2022) have recently shown for equity mutual funds that investors strongly react to them. We find that the Sharpe ratio consistently explains corporate bond fund flows significantly better than the raw return or any alpha from single- or multi-factor models. The differences are highly statistically significant. Compared to the raw return and other factor models, the Elton et al. (1995) four-factor model further has a significantly higher explanatory power, although it is clearly not as high as that of the Sharpe ratio.

Next, we examine to what extent the use of simple performance measures such as the Sharpe ratio is related to investor sophistication. It is likely that the least sophisticated investors are unaware of all factor models, while those with a higher sophistication may base their investment decisions on factor models to a much greater extent. We therefore perform splits of the corporate bond fund sample into (i) retail- and institutional-oriented, (ii) high-yield and investment-grade, and (iii) rear-load and non-rear-load share classes. Institutional investors, those in high-yield bond markets, and those in non-rear-load share classes may arguably be more sophisticated. However, we show that for each of these subsamples, even in those with the presumably most sophisticated investors, the Sharpe ratio explains corporate bond mutual fund flows substantially better than all factor models.

Since part of the corporate bond mutual funds have non-trivial investments also in other asset classes, we also split the sample into those funds that invest mainly in corporate bonds and those with substantial holdings in other markets. It is likely that corporate bond factor models perform better in the former group. We find that, regardless of corporate bond mutual fund holdings ratios in corporate bonds, the Sharpe ratio explains their flows significantly better than the raw return and any factor model. Additionally, fund flows in illiquid periods could be affected by payoff complementarity (Goldstein, Jiang, & Ng, 2017). Therefore, we separately examine the flows in liquid and illiquid periods, showing that the Sharpe ratio explains fund flows better than any factor model in both.

In a further step, we consider the Morningstar ratings also as a potential explanatory variable. Morningstar ratings measure the long-term performance of a fund compared to its peer group. They are virtually uncorrelated to all other performance measures used in this study, which are based mainly on short historical horizons. We find that Morningstar ratings explain corporate bond fund flows better than any other performance measure, consistent with the finding of Ben-David et al. (2022) for equity markets. However, we also show that investors at the very least tend to use the Sharpe ratio for investment and redemption decisions *within* the different Morningstar rating groups. Further dissecting the importance of both the Sharpe ratio and the Morningstar Risk-Adjusted Return (MRAR(2)) measure underlying the calculation of the Morningstar ratings, it seems that investors rely on the Morningstar ratings mainly because they are salient and easily accessible. When put on an equal footing, the Sharpe ratio generally explains fund flows substantially better than the MRAR(2).

Finally, we examine whether investors' fund flows react to the return components related to bond risk factors. Consistent with our main results, we detect stronger fund-flow sensitivities to risk-factor related returns than to the factor model alphas.

We run a battery of robustness tests. Our results are qualitatively similar (i) when using Fama & MacBeth (1973) instead of panel regressions; (ii) with quintile instead of decile sorts of the performance measures; (iii) for alternative Sharpe ratio definitions; (iv) for a longer performance evaluation horizon (one year); (v) using various alternative factor models such as the Ludvigson & Ng (2009) bond macro factor model, as well as the bond factor model recently suggested by Bai, Bali, & Wen (2019); (vi) for an analysis at the fund instead of the share-class level; (vii) when controlling for time-varying effects of Morningstar ratings and the Morningstar fixed-income style box; and (viii) for an extended corporate bond sample without size filter. Lastly, we show that the Sharpe ratio also has superior explanatory power for future fund flows compared to the manipulation-proof performance measures of Getmansky, Lo, & Makarov (2004) and Goetzmann, Ingersoll, Spiegel, & Welch (2007).

Our horse race tests are similar in spirit to recent studies for equity markets by Barber et al. (2016) and Berk & Van Binsbergen (2016). It is natural to compare our corporate bond mutual fund evidence with the findings documented for equity mutual funds and hedge funds to have an integrated view across asset classes. Both Barber et al. (2016) and Berk & Van Binsbergen (2016) for equity mutual funds, and Blocher & Molyboga (2017) and Agarwal, Green, & Ren (2018) for hedge funds, show that the CAPM alpha explains investor flows better than the raw return or alphas from any factor model. Ben-David et al. (2022), on the other hand, show that Morningstar ratings and raw returns are the first and second most important drivers of mutual fund flows in equity markets. None of these studies analyzes corporate bond markets.

Our results reveal important differences between corporate bond funds and those of other asset classes. The contrast is particularly strong compared with Barber et al. (2016), Berk & Van Binsbergen (2016), Blocher & Molyboga (2017), and Agarwal et al. (2018), but also substantial compared with Ben-David et al. (2022). We find that raw returns and all CAPM-style models explain corporate bond mutual fund flows significantly less well than other performance measures.[3] Even more strikingly, both Morningstar ratings and the simple Sharpe ratio turn out to be more important than all factor models for corporate bonds.[4]

Our results carry potentially important implications for investors and fund managers. Unlike equity funds, corporate bond funds tend to hold more illiquid assets whose prices are often stale. As Cici, Gibson, & Merrick Jr (2011) and Choi, Kronlund, & Oh (2021) show, corporate bond mutual funds have substantial discretion in the valuation of their investments and, thus, could smooth reported returns. Smoothing returns over time leaves the portfolio's mean return unchanged but reduces its variance, which biases Morningstar ratings, the Sharpe ratio, and other similar performance measures upward (Getmansky et al., 2004; Bollen &

[3]Ben-David et al. (2022) argue that the raw return is the second-most important historical performance measure while all other studies cited in this paragraph find that CAPM alphas are the most important determinants of mutual fund flows in the equity and hedge fund markets. For corporate bond mutual funds, however, we find that these performance measures cannot explain fund flows well. For example, we show that the Elton et al. (1995) model explains corporate bond mutual fund flows significantly better than raw returns and all CAPM-style models (a single-factor corporate bond market model, a two-factor corporate bond–stock CAPM, and a three-factor corporate bond–stock–government bond CAPM).

[4]Note that Barber et al. (2016) do not present their result of pairwise comparisons between the Sharpe ratio and other measures. Berk & Van Binsbergen (2016), Blocher & Molyboga (2017), Agarwal et al. (2018), and Ben-David et al. (2022) do not consider the Sharpe ratio at all in their empirical studies.

Pool, 2008, 2009).[5] If investors use Morningstar ratings and the Sharpe ratio as the main performance measures to evaluate corporate bond funds, then fund managers aware of this will have an obvious incentive to take actions that enhance these measures without adding real economic value. Funds with illiquid assets, whose prices are only reported occasionally, may benefit from greater inflows. Reported performance could therefore strongly mislead investors' decisions. Fund managers who want to manipulate their performance measures can do so by holding more illiquid assets or "mismarking" bonds. However, such mismarking provides chances for active investors. The potential impact of Sharpe ratio (and Morningstar ratings) manipulations is thus reminiscent of the stale-price mutual fund trading scandal in 2003: managers that smooth reported returns create trading opportunities for active traders at the expense of their buy-and-hold investors.

The remainder of this chapter is organized as follows. Section 2.2 describes our data and estimation methods. Section 2.3 conducts the flow–performance horse race. In Section 2.4, we show further evidence and test the robustness of our main findings. Section 2.5 discusses the implications of our results. Section 2.6 provides concluding remarks and suggestions for further research.

[5] Kim (2021) shows that fund managers also manipulate Morningstar ratings by portfolio pumping. That is, they tend to bid up the prices of their holdings shortly before new ratings are issued. This return manipulation positively affects both Morningstar ratings and the Sharpe ratio. It is arguably easier to perform for corporate bond mutual funds which generally have less liquid holdings than equity funds.

2.2 Data and Methodology

2.2.1 Data

Our data on U.S. actively managed corporate bond funds come from the Center for Research in Security Prices (CRSP) Survivorship-Bias-Free U.S. Mutual Fund Database. We use data from 1991 to 2017. Since we use an estimation window of five years in our empirical analysis, our final sample period used for testing is from 1996 to 2017.

A corporate bond mutual fund often offers various share classes to attract investors with different wealth levels and investment horizons. These share classes are designed as different combinations of front-end and/or back-end sales charges, expense ratios, management fees, minimum investment requirements, as well as restrictions on investor types. Since these fund-share level characteristics can influence the investment and redemption decisions of mutual fund investors, we follow Goldstein et al. (2017) and use individual fund share classes as our unit of observation.[6] Fund performance measures, such as Morningstar or Lipper ratings, are disseminated and displayed at the share-class level. As shown in Section 2.4.6, the results are qualitatively similar when conducting the analysis on the fund level.

We select corporate bond funds based on the objective codes provided by CRSP.[7] Since our interest is in investors that attempt to identify

[6]Performing analyses of fund flows on the share-class level is common in the literature, particularly for corporate bond mutual funds. See also, e.g., Zhao (2005), Del Guercio & Tkac (2008), Chen, Goldstein, & Jiang (2010a), Huang, Wei, & Yan (2012), Chen & Qin (2016), and Jiang & Yuksel (2017).

[7]Specifically, to be classified as a corporate bond fund, a mutual fund must have a (i) Lipper object code in the set (‘A’, ‘BBB’, ’CPB’, ‘HY’, ‘SII’, ‘SID’, ‘IID’), or a (ii) Strategic Insight objective code in the set (‘CGN’, ‘CHQ’, ‘CHY’, ‘CIM’, ‘CMQ’, ‘CPR’, ‘CSM’), or a (iii) Wiesenberger objective code in the set (‘CBD’, ‘CHY’), or (iv) ‘IC’ as the first two characters of the CRSP objective code.

managerial skill in their fund allocation decisions, we exclude index funds, exchange traded funds, and exchange traded notes. We remove fund share classes with total net assets (TNA) less than \$10 million to mitigate data biases.[8] We merge the CRSP data with the Morningstar Direct database, matching on fund CUSIPs and Tickers (Berk & Van Binsbergen, 2015 and Pástor, Stambaugh, & Taylor, 2015).

To measure the performance of corporate bond funds, for our main tests we use the raw return, the Sharpe ratio, and six different models that investors might reasonably employ for performance evaluation. The Sharpe ratio is probably the best-known and most widely used measure of portfolio performance employed in the fund industry (Goetzmann et al., 2007 and Elton & Gruber, 2013). The Sharpe ratio is used, for example, in the Schwab Select List, the Standard and Poor's Select Funds mutual fund rating, and in the Hulbert Financial Digest newsletter ratings.[9]

The factor models used include a single-factor CAPM bond model of the aggregate corporate bond market (MKT^{bond}) return (Cb), a two-factor model of both the aggregate corporate bond return and stock market (MKT^{stock}) return (Csb) following Goldstein et al. (2017), a three-factor model following Bekaert & De Santis (2021), which includes an alternative aggregate corporate bond excess return (MKT^{bond^*}), as well as those from

[8] As shown in Section 2.4.9, the results are qualitatively similar without the filter on fund TNA. Our results are not subject to incubation bias (Evans, 2010), as we use a 60-month estimation window (with a minimum requirement of 30 observations) for Equation (2.2) below before considering the fund share classes for the analysis of the flow–performance relationship.

[9] In cases when the excess returns are negative, the Sharpe ratio rankings remain valid. To see that, consider two examples. First, supposing that two fund managers delivered the same negative returns, the one with a higher volatility performed better than the other, because the same loss was achieved with a higher positive return potential. Second, suppose that fund A has a more negative return and a higher volatility than fund B, but a higher/less negative Sharpe ratio. A combined strategy of fund and Treasury bill with a targeted volatility using fund A performs better than one using fund B.

the stock and government bond (MKT^{gov}) markets (B3).[10] Additionally, we consider the Elton et al. (1995) four-factor model (E4) including the aggregate corporate bond and stock market excess returns, as well as default risk (DEF) and option (OPTION) factors and the Fama & French (1993) five-factor model (F5) including three common stock factors: MKT^{stock}, the size factor (SMB), the value factor (HML), and two bond factors: the term spread (TERM) and DEF.[11] Finally, we also use an augmented F5 model that adds the liquidity (LIQ) and momentum (MOM) factors to the F5 model (F7).[12] We provide additional details on the factors in the following paragraphs.

The excess corporate bond market return (MKT^{bond}) is proxied by the Barclays U.S. Aggregate Bond Index in excess of the one-month T-bill return. For the B3 model, we calculate ourselves a value-weighted corporate bond market excess return using data from NAIC and TRACE (MKT^{bond^*}). The main difference to the Barclays index is that our self-calculated corporate bond index also includes high-yield bonds, which Bekaert & De Santis (2021) argue is an important feature of their model specification. For the excess government bond return, we use the 7–10 year Thomson Datastream benchmark bond total price index. TERM

[10]We use a local three-factor model rather than a global one-factor CAPM also featured in Bekaert & De Santis (2021). We do so because (i) this model specification is more flexible, (ii) the global data are not available for our full sample period, (iii) we focus on U.S. corporate bond mutual funds, and (iv) Bekaert & De Santis (2021) show that (1) a global one-factor model does not perform significantly better than a local one-factor model and (2) a local three-factor model (which we use) significantly outperforms a local one-factor model.

[11]Elton, Gruber, Agrawal, & Mann (2001) show that in addition to expected default and state taxes, the Fama and French stock factors explain the rate spread on corporate bonds. Empirical evidence in Gebhardt, Hvidkjaer, & Swaminathan (2005a) and Lin, Wang, & Wu (2011) show that DEF and TERM betas are important determinants of required corporate bond returns.

[12]The factors MKT^{stock} (excess market return), SMB (small minus big), HML (high minus low), MOM (winners minus losers), and LIQ (liquidity risk) are described in and obtained from Kenneth French's and Lubos Pástor's online data libraries: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html` and `https://faculty.chicagobooth.edu/lubos.pastor/research/liq_data_1962_2017.txt`.

is defined as the difference between the monthly long-term government bond return and the one-month Treasury bill rate, which captures returns generated by increasing duration (i.e., higher interest rate risk). DEF is defined as the difference between the return on a high-yield bond index and the Barclays Intermediate Government bond return, capturing returns generated by taking on higher default risk. OPTION captures nonlinearities due to investment in mortgage-backed securities and is measured by the difference between the Barclays GNMA index and the Barclays Intermediate Government Index. We obtain monthly return data on the Barclays bond indices from Morningstar Direct.

We use equity MOM and LIQ factors in our main test instead of the Jostova, Nikolova, Philipov, & Stahel (2013) bond MOM and the Lin et al. (2011) LIQ factors, for two reasons. First, the factors are only available from the authors for a subset of our sample period, while the equity factors are available for our entire sample period. Second, the results of Lin et al. (2011) show that the coefficient of the Pástor & Stambaugh (2003) stock liquidity factor beta is significant even after incorporating bond characteristic variables, suggesting a possible cross-market liquidity risk effect. Gebhardt, Hvidkjaer, & Swaminathan (2005b) and Jostova et al. (2013) find significant evidence of a momentum spillover from equities to corporate bonds (i.e., past equity returns significantly predict future bond returns). However, in Section 2.4.5 we also conduct robustness tests using the bond LIQ and MOM factors as well as the Bai et al. (2019) factor model for a shorter sample period corresponding to the period for which those factors are available.[13] The results of these analyses are qualitatively similar to those of our main analysis.

2.2.2 Empirical Approach

Fund Flows

The key variables in our empirical analysis are mutual fund flows and the different performance measures. Following standard practice, we calculate the flows ($F_{p,t}$) of fund share class p in month t as the percentage growth of new assets:

$$F_{p,t} = \frac{TNA_{p,t}}{TNA_{p,t-1}} - (1 + R_{p,t}), \tag{2.1}$$

where $TNA_{p,t}$ is the total net assets under management of fund share class p at the end of month t, and $R_{p,t}$ is the return of fund share class p in month t. That is, the fund flows are the growth rate in total net assets minus the growth rate explainable by the return on the previous month's total net assets. Note that this approach assumes that all flows take place at the end of the month. To mitigate the influence of outliers (for example, due to fund mergers and splits), we remove extreme fund flows at the 1% and 99% levels.

Fund Performance Measures

In the following, we will outline the procedure to estimate the realized alpha in detail for the F5 model. The procedure is similar for all other

[13] Bond MOM data are provided from 1974 until June 2011 on Gergana Jostova's website `https://business.gwu.edu/gergana-jostova`. Bai et al. (2019) propose a new bond factor model which includes: downside risk (DRF), credit risk (CRF), and liquidity risk (LRF). Data for the factors are available on Turan Bali's website: `https://sites.google.com/a/georgetown.edu/turan-bali/data-working-papers`. DRF and CRF cover the period from July 2004, LRF from August 2002. Using the TRACE database to create the factors ourselves would also limit the sample period, because TRACE does not start before July 2002.

factor models. First, we obtain factor loadings by the following time-series regression using 60 months of return data for months $\tau = t-60$ until $t-1$:

$$\begin{aligned}(R_{p,\tau} - R_{f,\tau}) \quad = \quad & \alpha_{p,t} + \beta_{p,t} MKT^{stock}_{\tau} + s_{p,t} SMB_{\tau} + h_{p,t} HML_{\tau} + t_{p,t} TERM_{\tau} \\ & + d_{p,t} DEF_{\tau} + e_{p,\tau}. \end{aligned} \tag{2.2}$$

The parameters $\beta_{p,t}$, $s_{p,t}$, $h_{p,t}$, $t_{p,t}$, and $d_{p,t}$ represent the exposures to stock market, size, value, term risk, and default risk, respectively, of fund share class p at time t. $\alpha_{p,t}$ is the average return that cannot be explained by factor tilts and $e_{p\tau}$ is a mean-zero error term.

Second, we calculate the alpha for the fund share class in month t as the difference between its realized return and its model-implied return in month t:

$$\begin{aligned}\hat{\alpha}_{p,t} \quad = \quad & (R_{p,t} - R_{f,t}) - \Big[\hat{\beta}_{p,t} MKT^{stock}_{t} + \hat{s}_{p,t} SMB_{t} + \hat{h}_{p,t} HML_{t} + \hat{t}_{p,t} TERM_{t} \\ & + \hat{d}_{p,t} DEF_{t}\Big]. \end{aligned} \tag{2.3}$$

For other factor models, we adjust Equations (2) and (3) accordingly.

The Sharpe ratio of fund share class p at the end of month t is calculated as the ratio of the average excess return of fund share class p at the end of month t over the standard deviation of its monthly returns over the past year. In Section 2.4.3, we also test the robustness of our main results to alternative ways to calculate the Sharpe ratio.

Performance Evaluation Horizon

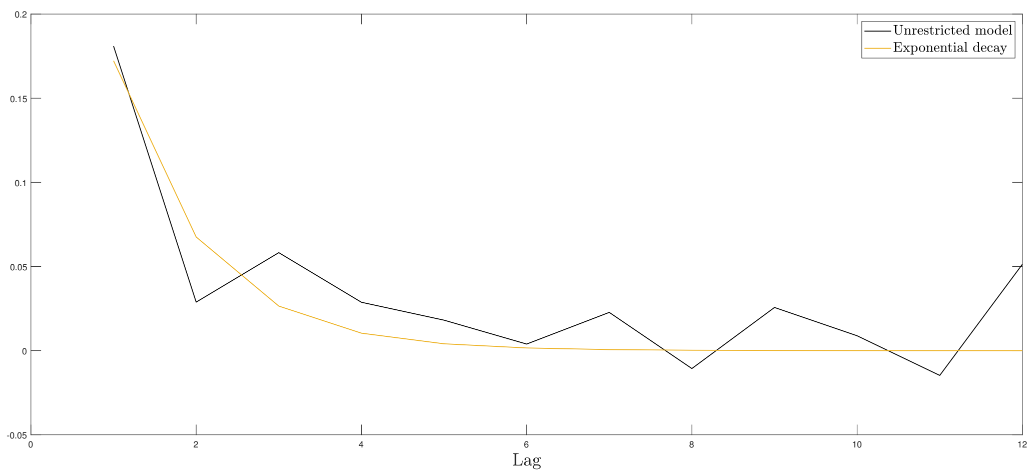
Rational investors respond to their perceptions about the skill of a fund manager. With new information, they should update this perception. To make a decision about how investors weight past returns and what performance horizon to analyze when comparing models, first, we estimate the following simple model of the flow–return relation:

$$F_{p,t} = a + \sum_{s=1}^{S} b_s R_{p,t-s} + e_{p,t}, \tag{2.4}$$

where $F_{p,t}$ are fund flows for share class p in month t and $R_{p,t-s}$ represents the lagged returns for the fund share class at lag s, where we vary the number of maximum lagged return from $S = 1$ to 48 months. The Akaike information criterion (AIC) yields a minimum for $S = 1$.[14] We thus settle on a lag length of 1 month. In Figure 2.1, the black line depicts the estimated b_s coefficients (y axis) at various lags (x axis). From this figure, it becomes clear that the first lag return is the most influential indicator about fund performance to investors, while the sensitivities to more distant returns are close to zero. Interestingly, and consistent with this result, Choi et al. (2021) document the return predictability of corporate bond mutual funds up to the one-month horizon.

Figure 2.1: Fund Flow–Past Return Relation

The figure presents the regression coefficient estimates (y axis) at various lags (x axis) for two models of monthly fund flows. The first (in black) is a simple unrestricted model, which includes twelve lags of monthly fund returns. We present the individual coefficient estimates on each lagged return. The second (smooth orange line) is an exponential decay model as in Equation (2.9) with the decay rate parameter λ.



[14] The Bayesian information criterion (BIC) also yields a minimum for $S = 1$. Both AIC and BIC also yield a minimum for $S = 1$ when we additionally include control variables and time fixed effects.

2.2.3 Descriptive Statistics

In Table 2.1, we provide summary statistics. Our final sample includes 1,591 unique funds and 3,577 unique share classes (1,042 high-yield share classes and 2,535 investment-grade share classes) with in total 356,243 share class–month observations from January 1996 to June 2017. The minimum number of unique share classes (funds) available per month during our sample period is 637 (447) in 1996. The maximum is 1,935 (852) in 2017. On average, our sample includes 1,387 fund share classes belonging to 706 funds per month. Panel A of Table 2.1 summarizes the fund characteristics. The average fund share class has total net assets of about \$609.15 million. The median is considerably smaller with \$113 million, which suggests that the fund size is skewed by large funds. The mean (median) fund age is 10.61 (8.59) years. The average annual expense ratio for our sample is 0.96%. The majority of share classes (75%) has either a front-end or back-end load. The average fund return standard deviation amounts to 1.22%, which is substantially lower compared to that of equity funds (Barber et al., 2016 report 4.92% for their sample).

Panel B of Table 2.1 reports descriptive statistics of monthly fund flows and returns, the two key variables of our analysis. Over the sample period, the mean return (the time-series average of the cross-sectional distribution of monthly fund returns) of all share classes in our fund sample is 0.42% per month (5.04% per annum). Investment-grade bond share classes yield an average return of 0.37%, while high-yield share classes yield an average return of 0.55% per month. The average (median) of the percentage fund flow is 0.58 (–0.11), with a standard deviation of 5.09% per month. The dispersion in fund flows is higher than that documented by Barber et al. (2016) for equity funds, 2.25%. High-yield bond fund returns exhibit an average first order autocorrelation of 23.07%, which is higher than that of investment-grade funds (16.15%). There is also a substantial serial correlation in the fund flows. The first-order autocorrelation is 28%, which is approximately equal for high-yield and investment-grade funds.

Table 2.1: Descriptive Statistics

This table presents summary statistics. Our sample contains 3,577 unique fund share classes (1,042 high-yield bond share classes and 2,535 investment-grade share classes) of actively managed U.S. corporate bond mutual funds from January 1996 to June 2017. Panel A summarizes the fund characteristics. SD denotes the standard deviation. P25 and P75 are the 25% and 75% quantiles, respectively. Panel B reports the time-series average of the cross-sectional distribution of fund returns and flows. $\rho 1$ is the first-order autocorrelation, reported in percentage points. Panel C summarizes the time-series of the different model alphas estimated using rolling 60-month fund past returns. Panel D presents the correlation matrix between different performance measures. The unit of observation is fund share–month.

A. Fund characteristics

	Mean	SD	P25	Median	P75
Size ($mil)	609.15	2913.23	39.10	113.00	376.70
Age (years)	10.61	8.90	4.42	8.59	14.39
Expense ratio (%)	0.96	0.47	0.62	0.85	1.22
Noload dummy	0.32	0.47	0.00	0.00	1.00
Volatility (t-12 to t-1)(%)	1.22	1.03	0.68	0.96	1.44

B. Fund return and flow

	Mean	SD	P25	Median	P75	$\rho 1$
	Fund return (% per month)					
All funds	0.42	1.65	–0.15	0.43	1.12	18.14
Investment-grade	0.37	1.22	–0.12	0.36	0.94	16.15
High-yield	0.55	2.41	–0.31	0.72	1.62	23.07
	Fund flow (% per month)					
All funds	0.58	5.09	–1.57	–0.11	1.76	28.32
Investment-grade	0.59	4.94	–1.47	–0.07	1.72	28.39
High-yield	0.57	5.44	–1.81	–0.23	1.87	28.17

C. Fund alpha

	Mean	SD	P25	Median	P75
Sharpe ratio	0.231	1.297	-0.424	0.338	0.934
Cb alpha	0.062	1.538	-0.207	0.032	0.388
Csb alpha	-0.044	1.133	-0.286	-0.031	0.221
B3 alpha	0.014	1.036	-0.266	-0.007	0.291
E4 alpha	-0.043	0.741	-0.229	-0.026	0.181
F5 alpha	-0.010	0.837	-0.303	-0.006	0.302
F7 alpha	-0.007	0.860	-0.306	-0.005	0.312

D. Correlation between different performance measures

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
(a) Raw return	1.00	0.74	0.88	0.62	0.42	0.29	0.34	0.33	0.06
(b) Sharpe ratio		1.00	0.53	0.37	0.25	0.22	0.29	0.28	0.08
(c) Cb alpha			1.00	0.69	0.48	0.34	0.36	0.35	0.05
(d) Csb alpha				1.00	0.75	0.52	0.51	0.50	0.06
(e) B3 alpha					1.00	0.65	0.59	0.57	0.06
(f) E4 alpha						1.00	0.85	0.82	0.09
(g) F5 alpha							1.00	0.97	0.08
(h) F7 alpha								1.00	0.08
(i) MS rating									1.00

As can be seen from Panel D of Table 2.1, some of the performance measures are materially correlated. Hence, the different models likely yield similar rankings in many cases. The empirical approach we use entails exploiting those cases in which rankings differ across models (Barber et al., 2016). This helps us determine the performance measure that best explains investors' capital allocation choices to actively managed corporate bond mutual funds. The correlations between Morningstar ratings and other performance measures are quite low, all below 10%.[15]

2.3 Empirical Results

2.3.1 Model Horse Race

As in Barber et al. (2016), we classify fund share class performance using decile ranks and examine in a pairwise fashion which model better explains the flows when the models yield different performance ranks.[16] To obtain the relation between fund flows and the ranking of a fund share class based on two different performance measures, we estimate the following regression:

$$F_{p,t} = a + \sum_{i} \sum_{j} b_{ij} D_{ij,p,t-1} + cX_{p,t-1} + \mu_t + e_{p,t}, \tag{2.5}$$

where the dependent variable $F_{p,t}$ is the fund flow of mutual fund share class p in month t. In each month of our test period, we assign the decile performance

[15]Del Guercio & Tkac (2008) report a correlation between Morningstar ratings and the Sharpe ratio of 0.65 for their equity fund sample from Nov 1996 to Oct 1999. Sharpe (1998) also documents a high correlation. There are three reasons for the low correlation we report. First, our sample covers both a period before and one after the 2002 major change in the methodology used by Morningstar. The two studies mentioned examine the Morningstar rating only before the methodology change in 2002. Second, Morningstar ratings are based on three-, five-, and ten-year returns while the main performance measures covered in this study are based on much more recent returns. Finally, Morningstar ratings are discrete while the Sharpe ratio is continuous. Sharpe (1998), for example, reports the percentile correlation rather than the raw one.

[16]In Section 2.4.2, we show that the results are also robust to using quintile instead of decile ranks.

rank for each fund share class based on each of the measures.[17] Decile 10 includes the best-performing fund share classes and decile 1 contains the worst fund share classes based on the performance measure. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund share class p in month $t-1$ is in the decile i based on the first model and decile j based on the second model. To estimate the model in Equation (2.5), the dummy variable for $i = 5$ and $j = 5$ is excluded. $X_{p,t-1}$ represents control variables including the lagged fund flow from month $t-1$, the lagged expense ratio, a dummy for no-load share classes, the return standard deviation estimated over the prior twelve months, the log of fund share class size, the log of fund age in month $t-1$, as well as Morningstar rating dummies in month $t-1$.[18] The studies of Del Guercio & Tkac (2008) and Ben-David et al. (2022) find that Morningstar ratings are an important determinant of equity mutual fund flows. In a first step, we use Morningstar ratings as control variables. In Section 2.3.3, we also analyze their performance as an explanatory variable of corporate bond mutual fund flows. Our results are qualitatively similar without Morningstar rating control variables. We also include time fixed effects (μ_t).

In the main analysis, we use panel regressions to estimate Equation (2.5). In Section 2.4.1, we show that the results are robust to using Fama & MacBeth (1973) regressions instead (see Ben-David et al., 2022). Following Petersen (2009) and Cameron, Gelbach, & Miller (2011), we double-cluster the standard errors by fund and month. We cluster on the fund rather than only on the share-class level. This helps address correlations in the residuals among the different share classes of a given fund, both contemporaneous and over time. Clustering by month helps address cross-sectional correlation in residuals across different funds at a given time point.

[17] We rank a fund share class based on each performance measure within its category peer group (i.e., investment-grade or high-yield). This ensures that the rankings are driven mainly by managerial skill rather than choice of investment style or systematic events that affect all share classes in a category peer group.

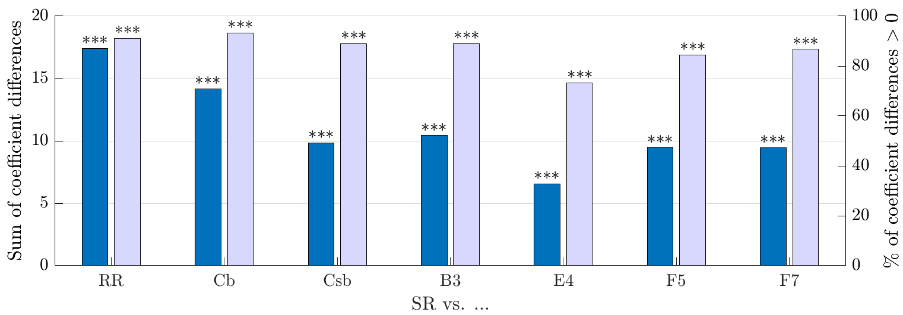
[18] Note that we expand the set of control variables in Barber et al. (2016) because Del Guercio & Tkac (2008), Evans & Sun (2021), and Ben-David et al. (2022) show that Morningstar ratings substantially influence allocation decisions, in particular for retail investors. Morningstar is the dominant information intermediary among financial advisors, being more influential than, for example, Lipper and Standard & Poor's.

The key coefficients of interest are b_{ij} ($i = 1, 2, ..., 10$ and $j = 1, 2, ..., 10$). These can be interpreted as the percentage flows received by a fund share class, which is in decile i based on the first performance measure and in decile j for the second measure relative to a corporate bond fund share class that ranks in the fifth decile for both performance measures. With each pair of coefficients b_{ij} and b_{ji}, we can determine whether investors are more sensitive to the first or to the second measure (alpha estimated using the first model or using the second model). We test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. If investors are more responsive to the first measure than the second one, we would expect to reject the null hypothesis in favor of the alternative hypothesis that b_{ij} is greater than b_{ji}. For each pairwise comparison of two performance measures, we have 45 such b coefficient comparisons. We test the null hypothesis that the summed difference across all 45 comparisons is equal to zero using a Wald test. In addition, we test the null hypothesis that the proportion of positive and negative differences equals 50% using a binomial test. We present a "winning model" if the sum of coefficient differences is significantly different from zero.

Figure 2.2 presents the main results. We compare the Sharpe ratio to all other performance measures. We find that compared to the raw return, the sum of coefficient differences amounts to 17.42, which is highly statistically significant. 91.11% of the coefficient differences are positive. It is thus clear that investors are substantially more responsive to the Sharpe ratio than they are to the raw return. That is, investors appear to risk-adjust past returns. The pairwise comparisons of the Sharpe ratio with all factor models yield similar results. The sum of coefficient differences with the bond CAPM amounts to 14.16, that with the stock and bond CAPM to 9.85, and that with the B3 model to 10.43, all of which are highly statistically significant. The differences are comparably smallest between the Sharpe ratio and the E4 model, where the sum of coefficient differences amounts to 6.55. However, this is also highly statistically significant and 73.33% of the coefficient differences are positive. For the F5 and F7 models, the results are even more pronounced and highly statistically significant. Our first main result is, thus, that investors appear to rely on the Sharpe ratio rather than raw returns or any more sophisticated factor model when making their capital allocation decisions among corporate bond mutual funds.

Figure 2.2: Model Horse Race – Full Sample

The figure presents the results of pairwise comparisons of different performance measures to explain fund flows using the full corporate bond fund sample. Candidate performance measures are: the Sharpe ratio (SR), the raw return (RR), and the alphas of the single-factor model with bond market factor (Cb), the two-factor model with both bond and stock market factors (Csb), the Bekaert & De Santis (2021) three-factor model (B3), the Elton et al. (1995) four-factor model (E4), the Fama & French (2015) five-factor model for bonds (F5), and an augmented F5 model with liquidity and momentum factors (F7). We present the sum of coefficient differences (dark blue, left axis) as well as the share of positive coefficient differences (light blue, right axis). Values greater than zero indicate that the Sharpe ratio outperforms other performance measures. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.



We report the full details underlying Figure 2.2 as well as the comparisons between all other performance measures in Table A.1 in the Appendix A.2. We find that corporate bond mutual fund investors are substantially and significantly more responsive to alphas relative to the Elton et al. (1995) E4 model than to the raw return and those based on the other competing models. The sums of coefficients are significantly larger for the E4 model than for all others in the pairwise comparisons. However, even the E4 model explains fund flows substantially less well than the simple Sharpe ratio.

2.3.2 Tests on Subsamples of Corporate Bond Fund Share Classes

In our main analysis, we treat all corporate bond mutual fund investors as one homogeneous group. However, there are likely vast differences across investors. First and foremost, they differ in their degree of sophistication, ranging from unsophisticated retail investors to professionals. It appears reasonable to assume that more sophisticated investors may use different, more complex methods to assess the performance of corporate bond mutual funds than do the most unsophisticated retail investors. While these may be unaware of multi-factor models and hence have to rely on simple, readily available performance measures, the most sophisticated investors may be more likely to rely on complex factor models for their investment decisions. In this section, we thus test whether there are differences between investor flows to funds with different characteristics that may attract differing investor clienteles.

First, we directly split the fund share classes into those catering to retail and institutional investors. The latter group of investors are arguably more sophisticated and may be more likely to use factor models than are the former. Second, we split high-yield and investment-grade funds. High-yield and investment-grade markets are somewhat segmented. It is therefore interesting to look at each market on its own. On top of that, the results of Ben-Rephael, Choi, & Goldstein (2021) indicate that high-yield bond markets might contain more sophisticated investors. Third, we split the sample into rear-load and non-rear-load funds. Sophisticated investors likely prefer funds without rear-load expenses that substantially reduce short-term trading profits.

In an additional step, we split the corporate bond funds into those that hold mainly corporate bonds and those that also have non-trivial holdings in other asset classes. The corporate bond factor models arguably work better for the former and, hence, may be more widely used for these. Finally, we separate the sample into liquid and illiquid periods to analyze whether the flow–performance relationship differs in different aggregate illiquidity regimes.[19]

Retail and Institutional-Oriented Corporate Bond Fund Share Classes

While the corporate bond market is largely populated by institutions, retail investors make up a substantial share in corporate bond mutual fund holders. It is thus important to check whether there are differences in the performance measures used by institutional and retail investors. We hypothesize that institutional investors should be generally expected to use more sophisticated performance measures, while retail investors have limited information access and knowledge (Del Guercio & Tkac, 2002; Chen et al., 2010a; Evans & Fahlenbrach, 2012). Hence, they may rely on simpler measures when evaluating fund performance.

We split the sample into retail- and institutional-investor-oriented fund share classes by the classification provided by the CRSP Mutual Fund Database. From December 1999, CRSP assigns each fund share class a dummy for institutional share class and a dummy for retail share class.[20] The main classification criteria used are the minimum investment requirement and the distribution channel.[21] The two dummies are not mutually exclusive. Therefore, we set a fund share class as institutional-oriented if the CRSP institutional share class dummy is one and the CRSP retail share class dummy is zero.

Our results, presented in Panel A of Figure 2.3 (the full details are in Table A.2 in the Appendix A.2), show that retail investors in corporate bond mutual funds are most responsive to the Sharpe ratio among all performance measures. For retail-oriented mutual funds, the Sharpe ratio explains investor

[19] We thank two anonymous referees for suggesting that we should pursue these analyses in Sections 2.3.2.

[20] We backfill the CRSP investor-oriented classification for the share classes for which this information becomes available in 1999.

[21] According to the ICI Fact Book, institutional accounts include accounts direct-sold or purchased by an institution, such as business or financial organizations. Accounts of individuals are issued by a broker–dealer. Morningstar classifies as institutional fund share classes those typically purchased by large institutional buyers such as pension funds. These share classes are only offered to investors who invest $1 million or more, with the lowest expenses in the mutual fund universe.

flows strongly and highly significantly better than any other performance measure. For institutional-oriented mutual funds, we find that the Sharpe ratio also explains investor flows better than any other model. The sum of coefficient differences is positive in every case. However, since institutional-oriented funds only make up little more than one third on average of our entire sample, these tests are somewhat less powerful. The differences are only statistically significant when compared to the raw return and alphas of the CAPM bond, F5, and F7 models. Moreover, further analysis reveals that for institutional-oriented mutual funds, factor models generally explain investor flows better than the simple raw return. These results are provided in the Appendix A.2, Table A.3.

On the one hand, although we use a different method, these results are consistent with James & Karceski (2006). They find that institutional fund flows are more sensitive to risk-adjusted performance measures than retail fund flows, which show no significant difference in sensitivity to raw return and factor models. On the other hand, our finding that the preference for the Sharpe ratio as a performance evaluation measure does not depend on bond mutual fund investors' sophistication lines up with a similar result of Agarwal et al. (2018) for hedge fund investors with the CAPM alpha.

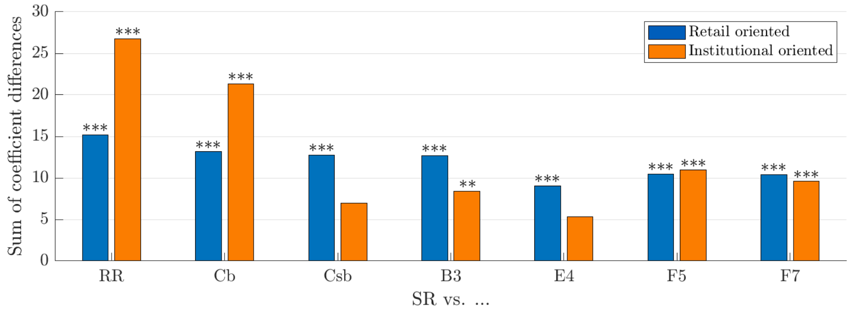
High-Yield and Investment-Grade Corporate Bond Fund Share Classes

The corporate bond market is divided into two main segments: investment-grade and high-yield bonds. Many market participants treat them as two separate asset classes (Ambastha, Dor, Dynkin, Hyman, & Konstantinovsky, 2010; Chen, Lookman, Schürhoff, & Seppi, 2014). It is thus possible that participants in both markets differ. Plausibly then, these different investors in the two market segments may employ different performance metrics. We thus test whether investors in investment-grade and high-yield bond funds differ in the way they evaluate performance of the fund when making capital allocation decisions. To perform this analysis, we split the entire sample into separate high-yield and investment-grade groups. Following Chen & Qin (2016), we categorize funds with Lipper object

Figure 2.3: Model Horse Race – Corporate Bond Fund Subsamples

The figure presents the results of pairwise comparisons of different performance measures to explain fund flows using several subsamples of corporate bond fund share classes. We separate the corporate bond mutual fund share classes into those oriented to retail and institutional investors (Panel A), those investing into high-yield and investment-grade segments (Panel B), those with and without rear-load fees (Panel C), and those with mainly corporate bond holdings and non-trivial holdings in other asset classes (Panel D). Candidate performance measures are: the Sharpe ratio (SR), the raw return (RR), and the alphas of the single-factor model with bond market factor (Cb), the two-factor model with both bond and stock market factors (Csb), the Bekaert & De Santis (2021) three-factor model (B3), the Elton et al. (1995) four-factor model (E4), the Fama & French (2015) five-factor model for bonds (F5), and an augmented F5 model with liquidity and momentum factors (F7). We present the sum of coefficient differences. Values greater than zero indicate that the Sharpe ratio outperforms other performance measures. The dark blue bars indicate the value for the respective first subset while the orange bars are for the respective second subset. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

A. Retail- and institutional-oriented bond fund share classes



B. High-yield and investment-grade bond fund share classes

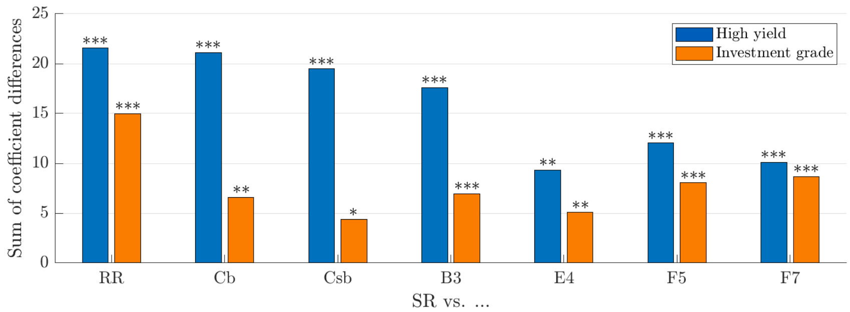
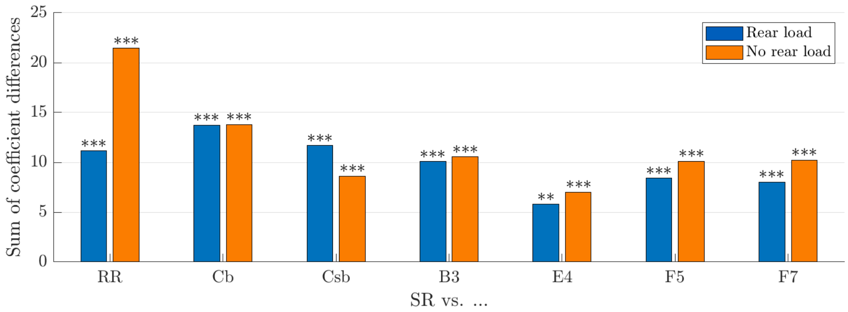
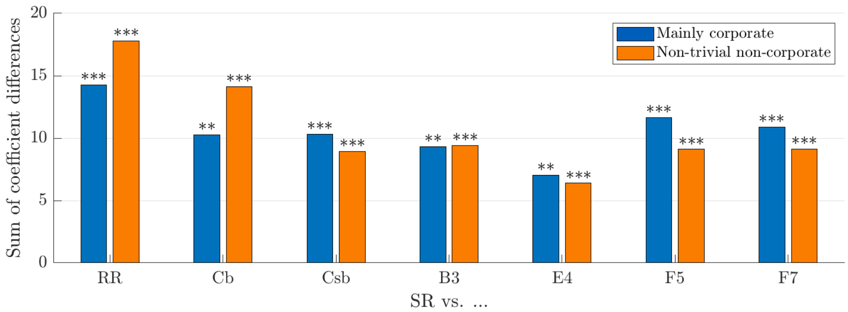


Figure 2.3: Model Horse Race – Corporate Bond Fund Subsamples (continued)

C. Rear-load and non-rear-load bond fund share classes



D. Mainly corporate bond holdings and non-trivial holdings in other asset classes



code 'HY', Strategic Insight objective code 'CHY' or Wiesenberger objective code 'CHY' as high-yield bond funds. Funds with all other objective codes are classified as investment-grade corporate bond funds.

We present the results for these two corporate bond market segments in Panel B of Figure 2.3. One can observe that the Sharpe ratio is able to best explain variation in flows across both high-yield and investment-grade bond mutual funds. Thus, investors in both classes appear to rely on the Sharpe ratio more than on the raw return or any factor model for their capital allocation decisions. Regarding factor models, Table A.3 in the Appendix A.2 reveals that investors of high-yield bond funds are somewhat more sensitive to alphas of models that include the default risk factor than factor models which do not include this factor. Investors of investment-grade bond funds seem to be more sensitive to the abnormal return relative to the bond (and stock) CAPM than to that of more complex models.

Rear-Load and Non-Rear-Load Corporate Bond Fund Share Classes

Equity and corporate bond mutual funds also differ substantially in the liquidity of the assets they are invested in. Price staleness in the net asset values (NAVs) of corporate bond mutual funds may present sophisticated investors with "hit-and-run" market timing opportunities. That is, the own private research of a sophisticated investor may indicate that the values of some of the holdings should be substantially lower than those recorded in the fund's NAV. In this case, she can still redeem her investments at the price recorded in the NAV until it is updated. Such updating is difficult and often takes time in corporate bond markets where the assets are rather illiquid and their prices can be stale, often lacking recorded market prices; a problem that is largely absent in equity markets. To deter these sophisticated investors, management companies' response has primarily been to use prohibitively high rear-load fees in order to prevent such short-term trading opportunities. Following this logic, sophisticated investors likely prefer funds without rear-load fees (front loads, on the other hand, are less likely to be an issue as long as these sophisticated investors have sufficient capital). Therefore, we split the sample into share classes with and without rear-load fees.

The results are in Panel C of Figure 2.3. We find that the Sharpe ratio explains fund flows significantly better for both classes. The sums of coefficient differences do not vary materially across the two corporate bond fund share class subsets. Interestingly, Table A.3 in the Appendix A.2 shows that the factor models generally explain fund flows better than raw returns for non-rear-load bond funds. This finding is consistent with the investors in non-rear-load asset classes being somewhat more sophisticated. However, the simple Sharpe ratio still explains their flows significantly better than each of the factor models. Thus, in summary this and the previous two subsections indicate that the Sharpe ratio is used for investment decisions largely independently of the degree of investor sophistication within the corporate bond mutual fund market.

Mainly Corporate Bond Fund Share Classes and Those with non-trivial Investments in Other Asset Classes

In a next step, we separate the funds into two groups based on whether they have exposure mainly to the corporate bond market or also non-trivially to stocks, Treasuries, or asset-backed securities. As shown by Choi, Hoseinzade, Shin, & Tehranian (2020) and Jiang, Li, & Wang (2021), some funds classified as corporate bond funds have substantial investments in these other asset classes. It is possible that investors use corporate bond factor models for funds mainly invested in corporate bonds. Recognizing that these factor models may be less suitable when managers also have non-trivial investments in other asset classes, they may resort to simpler performance measures for these.

For the classification, we define refined corporate bond fund share classes as those with at least 70% holdings in corporate bonds during more than 90% of their reportings (Chen & Qin, 2016). Approximately one third of the share classes satisfies these criteria. This subsample holds on average 89% of their assets in corporate bonds. The remaining share classes are categorized as an impure subsample of corporate bond mutual funds with non-trivial investments in other asset classes.

We present the results in Panel D of Figure 2.3. We find that for both subsamples the Sharpe ratio explains fund flows significantly better than any factor model. Thus, investors' use of the Sharpe ratio to measure the performance of corporate bond mutual funds does not seem to depend on the holdings structures of the funds.[22]

Aggregate Illiquidity

An important defining feature of corporate bond mutual funds is the existence of payoff complementarity (Chen et al., 2010a; Goldstein et al., 2017). That is, investors in mutual funds that hold particularly illiquid assets have incentives to

[22] Interestingly, from Table A.3 in the Appendix A.2 we can see that in the sample of mainly corporate bond holdings, none of the factor models explains investor flows better than the raw return.

redeem their shares first. Following substantial redemptions, the fund may have to sell its illiquid holdings at a loss. Thus, fund flows may behave substantially differently in times of market stress than in calm periods. Therefore, we form subsamples based on two indicators of the aggregate liquidity state: the S&P 500 volatility index (VIX) and the TED spread, i.e., the difference between the three-month LIBOR and the three-month Treasury bill rate. The high-VIX and high-TED-spread subsamples are defined based on whether the corresponding time-series variables exceed their sample means.

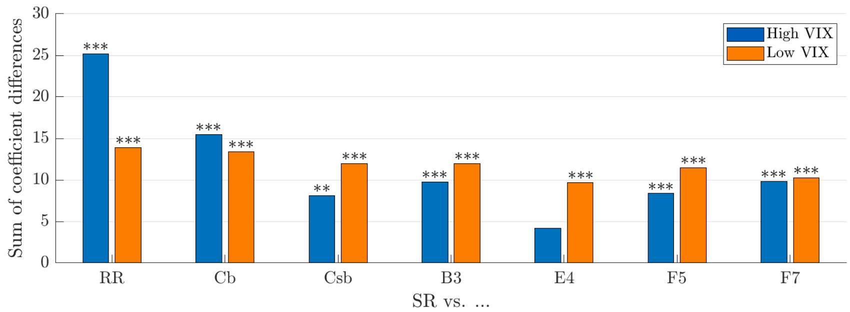
The results of this analysis are in Figure 2.4 (the full details are in Table A.4 in the Appendix A.2). We find that the Sharpe ratio explains mutual fund flows better than the raw return and all factor models in both liquid and illiquid periods. The differences are statistically significant in every case except for the comparison with the E4 model in illiquid periods. However, even in these cases the sums of coefficient differences are positive and the binomial test rejects its null hypothesis. Thus, in both liquid and illiquid periods the investors appear to rely on the simple Sharpe ratio rather than a factor model.[23]

[23] Table A.5 in the Appendix A.2 shows that investors tend to follow factor models more than the raw return in high-VIX and low-TED-spread regimes.

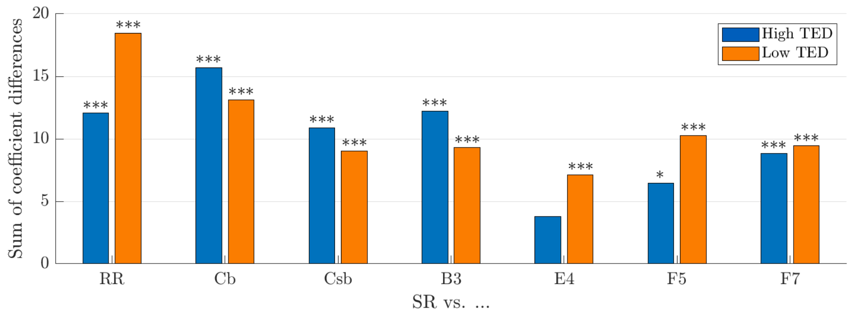
Figure 2.4: Model Horse Race – Aggregate Illiquidity Regimes

The figure presents the results of pairwise comparisons of different performance measures to explain fund flows during different aggregate illiquidity regimes. We use the VIX (Panel A) and the TED spread (Panel B) to capture the aggregate illiquidity. Candidate performance measures are: the Sharpe ratio (SR), the raw return (RR), and the alphas of the single-factor model with bond market factor (Cb), the two-factor model with both bond and stock market factors (Csb), the Bekaert & De Santis (2021) three-factor model (B3), the Elton et al. (1995) four-factor model (E4), the Fama & French (2015) five-factor model for bonds (F5), and an augmented F5 model with liquidity and momentum factors (F7). We present the sum of coefficient differences. Values greater than zero indicate that the Sharpe ratio outperforms other performance measures. The dark blue bars indicate the value for the respective first subset while the orange bars are for the respective second subset. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

A. High-VIX and low-VIX samples



B. High-TED-spread and low-TED-spread samples



2.3.3 Morningstar Ratings

Having documented that the Sharpe ratio explains investor flows in corporate bond mutual funds substantially better than all factor models, we next turn to the ability of the Morningstar ratings to explain corporate bond mutual fund flows. Responding to the studies of Barber et al. (2016) and Berk & Van Binsbergen (2016), Ben-David et al. (2022) show that the Morningstar ratings can explain equity mutual fund flows substantially better than CAPM alphas. Thus, we seek to examine whether Morningstar ratings are also the main drivers of corporate bond fund flows.

In doing so, we proceed in two steps. First, we include the Morningstar ratings both in the horse race tests conducted and in the alternative test of Berk & Van Binsbergen (2016). We have to adjust the test described in Equation (2.5) to quintiles because Morningstar only issues five different ratings (1 star up to 5 stars). However, even then the comparison is imperfect as the Morningstar ratings do not reflect actual quintiles of the distribution (see below). The Berk & Van Binsbergen (2016) methodology, which we describe in detail in Appendix A.1, relies only on the signs of flows and performance measures. This method is also imperfect as we have to define strategies based on the non-continuous Morningstar ratings variables. As Artavanis, Eksi, & Kadlec (2019) note, strategies that only use five-star funds end up selecting substantially fewer and only the best-rated funds, whereas the other performance measures use more. Therefore, to have a full picture, we consider three strategies: one that buys funds with 5 stars, one that buys all fund share classes with 4 or more, and one that buys all funds with at least 3 stars. The strategies simultaneously sell those fund share classes with less stars.

Despite these shortcomings, we perform the two tests to get an idea whether the Morningstar ratings are also a more important driver of flows in corporate bond mutual funds than raw returns, factor model alphas, or the Sharpe ratio.

The results for the model horse race are in Table A.6 in the Appendix A.2. We find that the Morningstar ratings explain fund flows better than any other performance measure considered in this study, including the Sharpe ratio. In Tables A.7 and A.8 in the Appendix A.2, we present the results for the Berk & Van Binsbergen (2016) test. Consistent with the other test, we find that the Morningstar rating strategies have higher flow–performance sensitivity probability estimates than all alternatives, with the differences being statistically significant in every case.[24] At the same time, these results confirm that the Sharpe ratio can explain fund flows better than the raw return or any factor model.[25]

Second, we perform a conditional double-sort based on Morningstar ratings and the Sharpe ratio. In Table 2.2, we present the associated equally and value-weighted percentage fund flows of the 25 double-sorted quintile portfolios. Consistent with our previous results, we find that both the Morningstar ratings *and* the Sharpe ratio are important drivers of fund flows. The smallest fund flows occur in the intersection of both the lowest Morningstar rating quintiles and the lowest Sharpe ratio quintiles. Similarly, the largest fund flows can be observed in the intersection of the highest Morningstar rating quintiles and the highest Sharpe ratio quintiles. Thus, while the Morningstar ratings have an important role for determining investors' fund flows, the Sharpe ratio clearly also has an important part, at the very least in determining in *which* of the many funds with the same Morningstar rating they allocate their flows. This is underlined by the highly statistically significant differences in investor flows between the high- and low-Sharpe-ratio portfolios within each Morningstar rating group.

[24] To keep the presentation of the tables manageable, we only present the results for a subset of models. Those for the other models considered in this study are qualitatively similar.

[25] Furthermore, we see that all of the flow–performance sensitivity probability estimates, $(\beta_{flow,performance} + 1)/2$, are greater than 50%, implying that a positive flow–performance relation exists for all performance measures. It is also noticeable that a significant fraction of flows remains unexplained when only considering past performance measures and Morningstar ratings, as none of the measures can explain more than 63%. Naturally, we cannot make out a significant difference between the Sharpe ratio and the raw return in Table A.8 in the Appendix A.2. This is because the Berk & Van Binsbergen (2016) approach only considers the sign of the performance measure. The sign of the Sharpe ratio is, to a large extent, defined by the sign of the return. However, the Sharpe ratio and the raw return may not have exactly identical signs because the risk-free rate is subtracted when computing the Sharpe ratio.

Table 2.2: Double-Sorts on the Sharpe Ratio within Morningstar Ratings Groups

This table reports the average percentage fund flows for 25 equally (Panel A) and value-weighted (Panel B) portfolios of fund share classes sorted into Sharpe ratio quintiles within each Morningstar rating category. We use the lagged TNA for value weighting. The final column presents the spread in average flows between the high- and low-Sharpe-ratio quintiles within the same Morningstar rating group. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

A. Equally weighted

	SR1 (low)	SR2	SR3	SR4	SR5 (high)	SR5 – SR1
1-star	-1.53	-1.21	-0.96	-0.98	-0.63	0.89***
2-star	-1.17	-0.73	-0.52	-0.52	-0.41	0.76***
3-star	-0.36	-0.08	0.00	-0.03	0.16	0.52***
4-star	0.40	0.67	0.62	0.82	0.92	0.53***
5-star	1.37	1.57	1.73	1.78	2.00	0.63***

B. Value-weighted

	SR1 (low)	SR2	SR3	SR4	SR5 (high)	SR5 – SR1
1-star	-1.44	-1.10	-0.88	-0.94	-0.54	0.90***
2-star	-1.15	-0.75	-0.51	-0.49	-0.40	0.76***
3-star	-0.45	-0.06	0.03	-0.04	0.10	0.55***
4-star	0.27	0.47	0.48	0.63	0.66	0.39***
5-star	0.91	1.26	1.24	1.34	1.55	0.64***

Having documented that investors respond both to the Sharpe ratio and to Morningstar ratings, we next examine the relation between the two more closely. For each share class p, Morningstar first computes the Morningstar risk-adjusted return ($\text{MRAR}(\gamma)_{p,t}$), which is essentially the expected utility of an investor with power utility and a relative risk aversion coefficient of $\gamma = 2$:

$$\text{MRAR}(\gamma)_{p,t} = \left[\frac{1}{T} \sum_{j=1}^{T} (1 + R_{p,t-j} - R_{f,t-j})^{-\gamma} \right]^{-\frac{12}{\gamma}} - 1. \tag{2.6}$$

Morningstar calculates the MRAR(2) measure each month t for each share class p over 3-, 5-, and 10-year horizons (assuming the share class has sufficient continuous monthly returns over the respective periods). Each MRAR(2) measure is further adjusted for sales charges, loads, and redemption fees. Subsequently, for each horizon, Morningstar allocates the star ratings among the category peer groups. Finally, the horizon-specific star ratings are consolidated to form an overall star rating.[26]

Hence, both the Sharpe ratio and the Morningstar rating are risk-adjusted performance measures. However, they differ in several important aspects: (i) the exact risk-adjustment method (scale-independent versus utility-based), (ii) the horizon for performance measurement, and (iii) the adjustment for category peers. On top of that, Morningstar ratings are very salient, generally being directly publicly disseminated along with information on a share class. On the other hand, there is no consensus about the historical horizon to use for computing the Sharpe ratio. Thus, while there is only one Morningstar rating for each share class, there are potentially many different Sharpe ratios investors may consider for their capital allocation decisions.

In a next step, we thus aim to dissect the reasons for the different flow sensitivities to these measures. We calculate the Sharpe ratio and the MRAR(2) of each fund for the 12-, 36-, 60-, and 120-month horizons. Then we compute

[26]Share classes with less than 3 years of historical return data are not rated. Share classes younger than 5 years receive the 3-year rating. Those between 5 and 10 years receive a weighted average rating with 60% weight on the 5-year MRAR(2). For those older than 10 years, the 10-year rating has a weight of 50%, the 5-year rating one of 30%, and the 3-year rating one of 20%.

hypothetical Sharpe-ratio-based and MRAR(2)-based ratings applying different parts of the Morningstar ratings calculation methodology. That is, we rank the Sharpe ratio and MRAR(2) for each horizon the same way Morningstar would do, but within a single group (i.e., top 10% – 5 stars, next 22.5% – 4 stars, next 35% – 3 stars, next 22.5% – 2 stars, and the bottom 10% – 1 star). We obtain the long-horizon rating by assigning different weights on the 3-, 5-, 10- year ratings of these two risk-adjusted return measures. To be able to perform a clean analysis, we focus on the period starting from July 2002, after the Morningstar rating methodology change.

Table 2.3 reports the results of this additional analysis on Morningstar ratings and the Sharpe ratio. We focus on the major differences between the two measures to find out which is the main driving force. First, in Panel A, we conduct pairwise comparisons between Sharpe ratio ranks and MRAR(2) ranks for different horizons. In this setting, we plainly analyze whether investors tend to use the Sharpe ratio or rather the MRAR(2) for performance measurement. When both measures are calculated based on a one-year horizon, the simple Sharpe ratio explains fund flows significantly better than the MRAR(2). For longer horizons (3 and 5 years), the outperformance of the Sharpe ratio is somewhat weaker but still statistically significant at the 10% level. When combining different horizons including the 10-year one, the Sharpe ratio still explains fund flows marginally better than the MRAR(2), but the difference is not statistically significant. Hence, we can conclude from this analysis that, when having to calculate the measures themselves, investors appear to prefer the Sharpe ratio over the utility-based MRAR(2).

Second, in Panel B of Table 2.3, we examine the impact of the historical horizon used to calculate the performance measures. That is, we perform pairwise comparisons between the one-year Sharpe ratio ranking and an aggregated ranking using 3-, 5-, and 10-year horizons. We find that the 1-year Sharpe ratio explains fund flows significantly better than the long-term aggregated rankings based both on the Sharpe ratio and the MRAR(2). Thus, when having to calculate the

Table 2.3: Dissecting the Impact of Morningstar Ratings and the Sharpe Ratio

This table reports further pairwise comparisons between the Sharpe ratio and the Morningstar risk adjusted return (MRAR(2)) in various specifications to explain fund flows using the full corporate bond mutual fund sample.

We first calculate the funds' Sharpe ratios and the MRAR(2) for different horizons (12, 36, 60, and 120 months). Subsequently, we rank them in the same way Morningstar does, except from the peer-group adjustment (i.e., the top 10% of funds receive five stars, the next 22.5% receive four stars, the next 35% three stars, the next 22.5% two stars, and the bottom 10% one star). We obtain the long-horizon rating by assigning different weights on the 3-, 5-, 10- year ratings of these two risk adjusted return measures. Finally, in Panel C, we compare these measures to the original star ratings disseminated by Morningstar (orgMS). We employ the Barber et al. (2016) approach with some adjustments. We estimate the relation between fund flow and its ranking based on different performance measures by running the regression:

$$F_{p,t} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{p,t-1} + \mu_t + e_{p,t}.$$

$F_{p,t}$ is the fund flow of mutual fund share class p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund share class p in month $t-1$ receives the star rating i based on the first measure and star rating j based on the second measure (excluding the dummy variable for $i = 3$ and $j = 3$). $X_{p,t-1}$ contains the following control variables (observed at the end of month $t-1$): lagged fund flow, lagged expense ratio, a dummy for no-load share classes, return standard deviation estimated over the prior twelve months, log of fund share class size, and log fund age. We also include time fixed effects (μ_t).

Panel A presents the comparisons of the sensitivity of fund flows to the two risk-adjusted return measures for different horizons. Panel B analyzes the horizon over which bond investors seem to care more when evaluating fund performance. Panel C examines the preferred setting (simply within a broad single category or based on the narrow peer-group adjustment embedded in Morningstar ratings) that bond investors use to rank fund performance when making investment decision.

For each pairwise comparison, we have 10 b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji}, we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. The table reports the results of two hypothesis tests: (1) H_0: The summed difference across all 10 comparisons is equal to zero, (2) H_0: The proportion of positive and negative differences equals 50%. We test the first hypothesis with a Wald test and the second with a Binomial test. We present a "winning model" if the sum of coefficient differences is significantly different from zero. "–" indicates that there is no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 2.3: Dissecting the Impact of Morningstar Ratings and the Sharpe Ratio (continued)

Panel A: Sharpe ratio vs. MRAR(2)

Winning measure	**SR**	**SR**	**SR**	–
Losing measure	**MRAR(2)**	**MRAR(2)**	**MRAR(2)**	–
	1-year	3-year	5-year	long horizon
Sum of coeff. Diff.	4.466***	1.839*	1.753*	1.499
p-Value	0.000	0.060	0.082	0.230
% of coeff. Diff. >0	100.00***	80.00*	90.00**	80.00*
Binomial p-Value	0.001	0.055	0.011	0.055

Panel B: Short vs. long horizon (3, 5, and 10 years)

Winning measure	**1-year SR**	**1-year SR**
Losing measure	**long-horizon SR**	**long-horizon MRAR(2)**
Sum of coeff. Diff.	2.206***	3.548***
p-Value	0.003	0.000
% of coeff. Diff. >0	90.00**	100.00***
Binomial p-Value	0.011	0.001

Panel C: Broad single category vs. peer-group adjustment

Winning setting	**Peer-group adjustment**	**Peer-group adjustment**
Losing setting	**Broad single category**	**Broad single category**
	orgMS vs. long-horizon SR	orgMS vs. long-horizon MRAR(2)
Sum of coeff. Diff.	5.468***	7.815***
p-Value	0.000	0.000
% of coeff. Diff. >0	100.00***	100.00***
Binomial p-Value	0.001	0.001

measures themselves, investors seem to prefer a shorter historical horizon to that used by Morningstar.

Finally, in Panel C of Table 2.3, we examine the impact of the peer-group adjustment and the salience of the reported measures. That is, we compare the simple rankings without adjustment for the investment style to those disseminated by Morningstar, which perform this adjustment. We find that the disseminated Morningstar ratings with the peer-group adjustment explain fund flows substantially better than both the Sharpe ratio and the MRAR(2) without this adjustment.

Overall, Morningstar ratings appear to cause fund flows mainly because of their salience and easy availability. Investors are not per se more sensitive to the underlying MRAR(2) base measure. On the contrary, when put on an equal footing, the simple Sharpe ratio explains fund flows generally better than the MRAR(2) measure. Only when directly using the exact measure actually disseminated by Morningstar does the MRAR(2) outperform the Sharpe ratio. These findings are consistent with the results of Evans & Sun (2021) and Ben-David et al. (2022) for equity mutual funds.

2.3.4 Response of Investor Flows to Components of Fund Returns

The preceding analysis indicates that the Sharpe ratio explains fund flows better than any factor model overall and also for every single market segment considered. In addition, Morningstar ratings appear to explain fund flows even better than does the Sharpe ratio. Thus, investors seem to react to very simple risk-adjusted return measures. It is likely that they do not fully account for the fact that a fund's performance in part depends on its passive exposure to systematic risk factors. Thus, we expect that investors also react to returns due to such exposure. In this section, we therefore examine to what extent investors consider factor-related returns when evaluating fund performance.

We decompose each fund's excess return into its alpha and factor-related returns by rearranging Equation (2.3). We conduct the return decomposition analysis for a seven-factor model, which is an augmented Fama & French (1993) five-factor model with added momentum and liquidity risk factors:

$$\begin{aligned}(R_{p,t} - R_{f,t}) \quad = \quad & \hat{\alpha}_{p,t} + \Big[\hat{\beta}_{p,t} MKT_t^{stock} + \hat{s}_{p,t} SMB_t + \hat{h}_{p,t} HML_t + \hat{t}_{p,t} TERM_t \\ & + \hat{d}_{p,t} DEF_t + \hat{m}_{p,t} MOM_t + \hat{l}_{p,t} LIQ_t\Big]. \qquad (2.7)\end{aligned}$$

In this return decomposition, the return of a fund share class is due to eight components: the fund's seven-factor alpha as well as its exposure to stock market risk, size, value, term risk, default risk, momentum, and liquidity risk. We calculate, for example, the portion of the return related to term risk as: $TERMRET_{p,t-1} = \hat{t}_{p,t-1} TERM_{t-1}$. Using this return decomposition, we estimate the following regression across p fund share classes and t months to test how investors react to different return components:

$$\begin{aligned} F_{p,t} &= \gamma_0 + \gamma_1 ALPHA_{p,t-1} + \gamma_2 MKTRET^{stock}_{p,t-1} + \gamma_3 SIZRET_{p,t-1} \\ &\quad + \gamma_4 VALRET_{p,t-1} + \gamma_5 TERMRET_{p,t-1} + \gamma_6 DEFRET_{p,t-1} \\ &\quad + \gamma_7 MOMRET_{p,t-1} + \gamma_8 LIQRET_{p,t-1} + cX_{p,t-1} + e_{p,t}, \end{aligned} \tag{2.8}$$

where the control variables, $X_{p,t-1}$ and the month fixed effect μ_t, are defined as before. In Equation (2.8), we are interested in the parameter estimates γ_i, $i = 1$, ..., 8. If investors react to returns from fund exposure to any specific factor, we expect the γ coefficient estimate corresponding to that factor to be significantly greater than 0. Based on Ben-David et al. (2022), we use the Fama & MacBeth (1973) method rather than panel regressions to estimate Equation (2.8).

The results are presented in Table 2.4. We observe that the sensitivities of investor flows to the unsystematic return part are significantly positive. However, the investor fund flows also respond significantly positively to returns due to size risk exposure. The coefficients on the size return components even exceed those on the factor model alpha. Thus, the investors appear to react to return components that are entirely unrelated to a fund manager's skill and can simply be obtained by following a size strategy. For exposures to other factors, the average coefficients are in part also large, although not statistically significant.

Table 2.4: Response of Fund Flows to Different Components of Fund Returns

This table reports the coefficient estimates from a Fama & MacBeth (1973) regression of percentage fund flows on the components of a fund's return based on a seven-factor model, which is an augmented Fama & French (1993) five-factor model with added momentum and liquidity factors. The regression also include control variables and month fixed effects. The standard errors are calculated by the Newey & West (1987) procedure with six lags (p-values are in parentheses). *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	All funds	**Investment grade**	**High yield**
ALPHA	0.331***	0.385***	0.355***
	(0.000)	(0.000)	(0.000)
MKTRET stk	-0.280	-0.102	-0.784
	(0.571)	(0.919)	(0.260)
SIZRET	11.32*	8.663*	12.97**
	(0.06)	(0.068)	(0.03)
VALRET	-2.573	-5.303	-3.855
	(0.440)	(0.296)	(0.204)
MOMRET	2.863	8.004	5.647
	(0.550)	(0.108)	(0.484)
TERMRET	5.104	1.694	0.359
	(0.499)	(0.810)	(0.950)
DEFRET	-4.298	-7.449	-7.018
	(0.351)	(0.311)	(0.377)
LIQRET	7.645	10.61	3.253
	(0.186)	(0.16)	(0.546)

2.4 Robustness

2.4.1 Fama–MacBeth Regressions

Ben-David et al. (2022) argue that panel regressions such as Equation (2.5) can be biased due to the time-varying nature of the flow–performance sensitivities.

To address this potential bias, they suggest also using Fama & MacBeth (1973) regressions. We follow their advice and test the robustness of our results to using Fama & MacBeth (1973) regressions of Equation (2.5).[27] The results are in Table A.9 in the Appendix A.2. Our conclusions are unchanged. The Sharpe ratio also explains corporate bond mutual fund flows significantly better than the raw return and any factor model when estimating Equation (2.5) with the Fama & MacBeth (1973) method.

2.4.2 Quintile Sorts

Next, we test the robustness of our results to building quintile instead of decile portfolios based on the different performance measures. Since the sample of corporate bond funds is somewhat smaller than that of equity funds, it is worth examining whether using the sorting mechanism based on deciles is a sensible choice. For this robustness check, we amend the main approach of Equation (2.5) accordingly. Consequently, instead of 45 b coefficient comparisons, we have only 10. We present the results in Table A.10 in the Appendix A.2. These are very similar to those when forming decile portfolios.

2.4.3 Alternative Sharpe Ratio Calculations

In our main analysis, we define the Sharpe ratio as the ratio of the previous one-month excess return to the one-year standard deviation so that its calculation is aligned with the estimation of all other performance measures, which are based on a one-month horizon for performance valuation (as indicated in Section 2.2.2). One possible concern is that this particular horizon may not be used by investors.

In this section, we show that our findings are robust to a conventional way to calculate the Sharpe ratio as the ratio of the monthly average twelve-month excess return to the one-year standard deviation. As can be seen from Table A.11 in the Appendix A.2, the Sharpe ratio based on a twelve-month window can explain the

[27] Naturally, we do not use the time fixed effects μ_t when performing Fama & MacBeth (1973) regressions as the intercept already captures these.

fund flows even better than that used in the main part of this study: the sum of coefficient differences and the share of positive coefficient differences are both larger than with one-month returns. Both the sums of coefficient differences and the proportions of positive coefficient differences are higher and more strongly statistically significant.

As an alternative to the simple volatility estimate, we use a volatility forecast of a GARCH(1,1) model. To estimate the model, we use 60 months of past returns, to align the estimation period with the method to estimate alphas. We then repeat our model horse race. The results, as reported in Panel B of Table A.11, are qualitatively similar to those from our main test.

2.4.4 One-Year Horizon for Performance Evaluation

To test the robustness of our main results, in this section, we examine a one-year horizon, instead of the one-month horizon for the performance evaluation as in our main analysis. While our analysis based on the AIC indicates that a one-month window is optimal, one might argue that a one-year window is also suitable because it broadly balances relevance (i.e., recent returns are likely more informative) and the signal-to-noise ratio (i.e., returns obtained over short horizons may mostly be noise and carry only little signal). Furthermore, there may be frictions such as inattention and transaction costs, which could create delays in the response of flows to fund performance.

The regression using Equation (2.4) yields a series of coefficient estimates, b_s, that represent the relation between flows in month t and the fund's return lagged s months, $s = 1, ..., 12$. Figure 2.1 shows that the most recent past return seems to be much more important to explain fund flows than more distant returns (i.e., the weights investors attach to past return quickly decay after the first previous month). Therefore, we follow Barber et al. (2016) and weight the performance measures. We empirically estimate the rate of decay λ in the flow–return relation using an exponential decay model:

$$F_{p,t} = a + b \sum_{s=1}^{12} e^{-\lambda(s-1)} R_{p,t-s} + e_{p,t}. \tag{2.9}$$

We present the results of this regression in Figure 2.1. The orange smooth line represents the estimated exponential decay function. It closely tracks the unconstrained coefficient estimates from the regression of Equation (2.4). We apply this decay function to calculate each fund's alphas as the weighted average of the prior twelve monthly alphas:

$$ALPHA_{p,t-1} = \left(\frac{\sum_{s=1}^{12} e^{-\hat{\lambda}(s-1)} \hat{\alpha}_{p,t-s}}{\sum_{s=1}^{12} e^{-\hat{\lambda}(s-1)}} \right). \tag{2.10}$$

We estimate this weighted alpha for each of the five models that we evaluate. We obtain the exponential decay rate $\hat{\lambda}$ based on the estimates from Equation (2.9). The Sharpe ratio of fund p at the end of month t is calculated as the ratio of weighted average of prior 12-month excess return of fund share class p using the decay rate λ over the one-year return standard deviation.

Table A.12 in the Appendix A.2 reports the result of the test using the one-year performance horizon. Consistent with the one-month horizon, the Sharpe ratio explains fund flows significantly better than the raw return and all factor models.

2.4.5 Alternative Factor Models

We check the robustness of our main results with a battery of alternative relevant recent factor models for corporate bonds, including:

1. An augmented Fama & French (1993) model with the Jostova et al. (2013) bond momentum factor (MOMb).
2. The Bai et al. (2019) four-factor model including a bond market factor and three new factors: downside risk, credit risk, and liquidity risk (B4).
3. An augmented Fama & French (1993) model with liquidity risk and aggregate volatility risk as in Chung, Wang, & Wu (2019) (C7).
4. The Ludvigson & Ng (2009) macro-factors for bonds (Macro).

5. An augmented Fama & French (2015) five-factor model with TERM and DEF (F7e).
6. An augmented Hou, Xue, & Zhang (2015) q-4 factor model with TERM and DEF (HXZ).
7. An augmented Stambaugh & Yuan (2017) M4 mispricing factor model with TERM and DEF (M4).

Table A.13 in the Appendix A.2 reports the results of the horse races between each alternative factor model and the other measures used in our main test. Consistent with our previous results, the Sharpe ratio also explains investor flows significantly better than each of these additional factor models.

2.4.6 Analysis on the Fund Level

In our main analysis, we follow Goldstein et al. (2017) and perform the analysis on the share-class level. This helps us to (i) have a larger sample size, (ii) account for the differences in performance and characteristics at the share-class level, and (iii) analyze different subsets of corporate bond share classes with different characteristics. We account for cross-correlations in the flows to different share classes of the same fund by clustering the standard errors by both fund and month. In this section, we test the robustness of our results to performing the analysis directly at the fund level. We aggregate the fund flows and value-weight the share-class returns and other variables to obtain those on the fund level. The results are in Table A.14 in the Appendix A.2. These are qualitatively similar to those of our main analysis on the share-class level.

2.4.7 Controlling for Time-Varying Effects of Morningstar Ratings

In our main analysis, we control for Morningstar ratings by including dummy variables. However, the Morningstar ratings may have time-varying effects. Hence, in this section, we examine the robustness of our results to controlling for

Morningstar ratings-times-month interaction fixed effects. We present the results in Table A.15 in the Appendix A.2. These are qualitatively similar to those with just Morningstar dummy variables.

2.4.8 Controlling for Morningstar Fixed-Income Style Box

We also follow Aragon, Li, & Qian (2019) and additionally control for fund-style effects. We do so by including Morningstar fixed-income style box-times-month interaction fixed effects as further control variables. The results, reported in Table A.16 in the Appendix A.2, clearly show that fixed-income styles do not account for our main result: the Sharpe ratio explains fund flows significantly better than all factor models. It is thus unlikely that unobservable time-varying variables common to Morningstar ratings or fund styles drive both investor flows and the Sharpe ratio.

2.4.9 Extended Corporate Bond Fund Sample

Lastly, we test the robustness of our results to the sample screening. That is, we drop the filter that fund share classes have to have a TNA of at least $10 million. The results, presented in Table A.17 in the Appendix A.2, are very similar to those of our main analysis.

2.5 Implications

Style- or factor-investing strategies have been implemented widely in equity markets. However, there are currently few investment vehicles for investors to harvest factor premia in the corporate bond market. Implementation of bond factor strategies in investment portfolios may not be easy because bond trading costs can be very high. This suggests that corporate bond investors may be less aware of factor models than investors in equity funds.[28] Therefore, one may argue that our finding that investors use simple measures instead of factor models to evaluate fund performance may be not that surprising.[29] On the other hand, on a broader level, our finding that investors rely on simple performance measures rather than multi-factor alphas is consistent with the findings of Ben-David et al. (2022) for equity markets.

However, if investors use Morningstar ratings and the Sharpe ratio instead of more sophisticated performance measures, this has much more severe implications for corporate bond fund managers and investors than for actors in equity mutual fund markets. Managers have strong incentives to "improve" both measures of their funds. They may be inclined to achieve this by smoothing their returns to reduce volatility instead of improving their investment decisions.[30] Some specifics of the corporate bond markets may enable them to do precisely that.

Corporate bonds are relatively more illiquid compared with government bonds and equities. The Investment Company Act of 1940 states that "A fund is generally required to price its portfolio using *readily available* market quotations" (emphasis added). However, many corporate bonds are held mainly in long-term investment portfolios of insurance companies or pension funds and are rarely traded. In case of thinly-traded bonds, the fund should value the securities at their "fair value", determined in "good faith" by, or under the direction of, the fund's board of directors. From these quotes, fund managers appear to have substantial discretion when it comes to valuing those illiquid assets on a daily basis.

[28] While the equity factor models of, e.g., Fama & French (1993), are rather well known among investors and there are numerous small-cap, value, and growth mutual funds, things are different in bond markets. Only few funds are explicitly designed to follow factor-related strategies. Thus, given the same level of sophistication, a corporate bond fund investor is likely less aware of factor models, simply lacking exposure to direct or indirect information about them.

[29] It should be noted that while the traditional players in the corporate bond markets are institutions such as insurance companies, pension funds, etc., the main investors in bond mutual funds are retail investors, who hold individual accounts sold through a broker–dealer.

[30] While reducing volatility trivially increases the Sharpe ratio by reducing its denominator, Morningstar ratings also penalize volatility (see, e.g., Barber et al., 2016; Ben-David et al., 2022).

Indeed, Cici et al. (2011) find that the valuations of *the same* bonds at *the same* point in time differ substantially across mutual funds. The authors find that the interquartile range (i.e., the width of the middle 50% of the distribution) of bond valuations is on average 0.30% of the bond face value for investment-grade bonds and 0.56% for high-yield bonds. While these numbers may seem large already, one should bear in mind that they are averages over a large number of bonds and that they do not reflect the most extreme valuations in the top and bottom quartiles. Thus, the extremest valuation differences are most likely multiples of these.

We implement two tests to check whether investors' capital allocation decisions in corporate bond mutual funds can be affected by such return smoothing manipulations. First, we examine the manipulation-proof performance measure (MPPM) of Goetzmann et al. (2007) as an alternative determinant of fund flows. The MPPM is a kind of enhanced Sharpe ratio and is defined as follows:

$$\hat{\Theta}_{p,t} = \frac{1}{(1-\rho)\Delta_\tau} \ln \left(\frac{1}{s} \sum_{\tau=t-s}^{t} \Big[(1+R_{p,\tau})/(1+R_{f,\tau}) \Big]^{1-\rho} \right), \tag{2.11}$$

where s is the length of the measurement horizon (in months) and $\Delta_\tau = 1/12$. We use relative risk aversion coefficients (ρ) of 3 and 4, as in Goetzmann et al. (2007). The higher ρ, the more heavily risk is penalized.

In Table 2.5, Panel A presents the pairwise comparison between the MPPM measure and the Sharpe ratio for different specifications. For the one-month horizon, as used in our main analysis, we find that the Sharpe ratio explains fund flows significantly better than the MPPM. The sum of coefficient differences is 17.42 and 91% of coefficient differences favor the Sharpe ratio. Also for a 12-month horizon and different values for the relative risk aversion, the results are qualitatively similar. Investors seem to rely on the simple Sharpe ratio instead of an alternative manipulation-proof measure.

Second, we conduct horse race tests between the simple Sharpe ratio and a smoothing-adjusted Sharpe ratio. Getmansky et al. (2004) propose a method to correct for the impact of return smoothing. Return smoothing implies that the reported or observed return $R_{p,t}$ of a fund in month t is a weighted average

Table 2.5: Model Horse Race – Manipulation-Proof Performance Measures

This table presents the results of pairwise comparisons between the Sharpe ratio (SR) and two performance measures that adjust for return-smoothing manipulations, including the manipulation-proof performance measure (MPPM) proposed by Goetzmann et al. (2007) and the smoothing-adjusted Sharpe ratio (adjSR) proposed by Getmansky et al. (2004) to explain fund flows using the full corporate bond fund sample. We estimate the relation between flow and the decile ranking of a fund share class based on different performance measures by running the regression:

$$F_{p,t} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{p,t-1} + \mu_t + e_{p,t}.$$

$F_{p,t}$ is the fund flow of mutual fund share class p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund share class p in month $t-1$ is in the decile i based on the first model and decile j based on the second model (excluding the dummy variable for $i = 5$ and $j = 5$). $X_{p,t-1}$ contains the following control variables (observed at the end of month $t-1$): lagged fund flow, lagged expense ratio, a dummy for no-load share classes, return standard deviation estimated over the prior twelve months, log of fund share class size, log fund age, as well as Morningstar rating dummies. We also include time fixed effects (μ_t).

We estimate the candidate performance measures for measurement horizons of one month and twelve months as well as with two different relative risk aversion coefficients for the MPPM (for the one-month horizon, the MPPM is independent of ρ).

For each pairwise comparison, we have 45 b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji}, we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. The table reports the results of two hypothesis tests: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of positive and negative differences equals 50%. We test the first hypothesis with a Wald test and the second with a Binomial test. We present a "winning model" if the sum of coefficient differences is significantly different from zero. "–" indicates that there is no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

A. Sharpe ratio vs. MPPM

Winning model	**SR**	**SR**	**SR**
Losing model	**MPPM**	**MPPM**	**MPPM**
Measurement horizon	one month	twelve months	twelve months
Risk aversion		$\rho = 3$	$\rho = 4$
	SR_MPPM	SR_MPPM	SR_MPPM
Sum of coeff. Diff.	17.42***	16.88***	17.40***
p-Value	0.000	0.000	0.000
% of coeff. Diff. >0	91.11***	88.89***	93.33***
Binomial p-Value	0.000	0.000	0.000

B. Simple vs. smoothing-adjusted Sharpe ratio

Winning model	**SR**	**SR**
Losing model	**adjSR**	**adjSR**
Measurement horizon	one month	twelve months
	SR_adjSR	SR_adjSR
Sum of coeff. Diff.	11.12***	14.46***
p-Value	0.001	0.000
% of coeff. Diff. >0	75.56***	88.89***
Binomial p-Value	0.000	0.000

of its unobservable true returns ($R^*_{p,t}$) in the current and k lagged periods. We follow Chen, Ferson, & Peters (2010b) and use a moving-average model with one lag ($k = 1$) because in our sample the first-order return autocorrelations are significant while the second-order autocorrelations are small and generally insignificant. Specifically, the observed returns of fund p are:

$$R_{p,t} = \theta_0 R^*_{p,t} + (1 - \theta_0) R^*_{p,t-1}. \tag{2.12}$$

To obtain the parameter θ_0 in Equation (2.12), we estimate a MA(1) model on the observed return series and re-scale the parameters to satisfy the equation (see Getmansky et al., 2004). Since the degree of return smoothing of a fund may be varying over time, we use rolling 12-month windows to obtain the parameters. We use the smoothing-adjusted return series $R^*_{p,t}$ to calculate a return smoothing-adjusted Sharpe ratio.

Panel B of Table 2.5 reports the results for the comparison between the simple Sharpe ratio and its smoothing-adjusted counterpart. For both evaluation horizons of one and twelve months, the response of investor flows to the simple Sharpe ratio is significantly stronger than to the smoothing-adjusted Sharpe ratio. The sum of coefficient differences is greater than 10 and more than 75% of coefficient differences favor the simple Sharpe ratio. This analysis provides further evidence, which supports the conclusion that investors appear to rely on the simple Sharpe ratio when making their investment decisions, while making no adjustment for the effect of return smoothing.

Thus, fund managers both have the means and strong incentives to manipulate their Sharpe ratios and Morningstar ratings. Cici et al. (2011) detect further evidence consistent with return smoothing behavior by corporate bond mutual fund managers. That is, the bond valuations are on average particularly high (in comparison to those of other funds) when the entire fund portfolio performs badly and low when the entire fund portfolio performs well.

Collectively, investors' reliance on Morningstar ratings and the Sharpe ratio for performance measurement can have high costs for investors. First, manipulations by fund managers may mislead them into adverse investment decisions and fund selections. Second, fund managers' return smoothing behavior can generate trading opportunities for active traders. There is a mismatch between the illiquidity of corporate bond funds' underlying assets and the liquidity they offer to investors by providing withdrawal rights on a daily basis. Similar to stale price-oriented mutual fund trading strategies (see, e.g., Chalmers, Edelen, & Kadlec, 2001; Goetzmann, Ivković, & Rouwenhorst, 2001; Boudoukh, Richardson, Subrahmanyam, & Whitelaw, 2002; Greene & Hodges, 2002; Zitzewitz, 2003, Choi et al., 2021) active investors can benefit from simple strategies based on the fund's observed valuations. This entails a further wealth transfer because buy-and-hold investors suffer from the offsetting losses and expenses (for example, from dilution effects). On top of this, fund managers may be forced to sell good securities to have enough cash for redemptions, which could further increase the adverse effects.

2.6 Concluding Remarks

How do corporate bond mutual fund investors measure performance? To answer this question, we analyze the relation between mutual fund flows and different performance measures. We run a horse race among different performance measures, ranging from the simple raw return and the Sharpe ratio to alphas estimated by using single and different multi-factor models. Our empirical analysis reveals that the Sharpe ratio explains the net flows into actively managed U.S. corporate bond mutual funds better than any of these alternatives. Morningstar ratings appear to explain an even larger share of investor fund flows, but the Sharpe ratio has important explanatory power within the Morningstar ratings groups. It thus seems that most investors do not use any factor model at all.

The use of Morningstar ratings and the Sharpe ratio as primary performance measures is problematic for several reasons. First, it facilitates the opportunistic behavior of fund managers to boost both measures on purpose (for example by holding illiquid assets or "hard-to-mark" bonds). Therefore, our findings have potentially important implications for both investors and managers of corporate bond mutual funds. Second, mutual fund return predictability caused by inaccurate prices allows profitable active trading strategies. Gains earned by active fund traders from trades that exploit NAV misvaluations are matched with the losses suffered by buy-and-hold fund investors.

We believe that further research should be undertaken to explore bond factors and make bond factor investing strategies more feasible. This might help corporate bond mutual funds to provide vehicles for bond investors to harvest factor premiums and make more sophisticated factor-model-based investor decisions. In the meantime, investors should be cautious about manipulation of reported measures and at least rely on manipulation-proof measures rather than on the simple Sharpe ratio.

A Appendix

A.1 The Berk–Van Binsbergen Testing Approach

First, we test for a positive relation between fund flows and performance (i.e., whether the regression coefficient of the sign of the subsequent flows on the sign of the performance measure is positive). Φ is defined as a simple sign function that returns the sign of a real number, taking values of 1 for a positive number, –1 for a negative number, and 0 for zero. We test the following null hypothesis:

$$\beta_{flow,performance} = \frac{cov(\Phi(F_{p,t}), (\Phi(\alpha_{p,t-1}))}{var(\Phi(\alpha_{p,t-1}))} > 0. \tag{A.1}$$

For the ease of interpretation, Table A.7 reports $(\beta_{flow,performance} + 1)/2$ which denotes the average likelihood that the sign of the fund flow [*S(flow)*] is positive (negative) conditional on the sign of the past performance measure [*S(performance)*] being positive (negative).

Furthermore, we can consider pairwise comparisons of two performance measures (models) and test which better captures how investors assess fund performance to allocate their capital by the following equation:

$$\Phi(F_{p,t}) = a + b_1 \left(\frac{\Phi(\alpha^{m1}_{p,t-1})}{var(\Phi(\alpha^{m1}_{p,t-1}))} - \frac{\Phi(\alpha^{m2}_{p,t-1})}{var(\Phi(\alpha^{m2}_{p,t-1}))} \right) + \xi_{p,t}. \tag{A.2}$$

If the coefficient of this regression is positive (i.e., $b_1 > 0$), this implies that the flow–performance regression coefficient of model $m1$ is larger than that of model $m2$, and we can infer that model $m1$ better explains the sign of the subsequent fund flows than model $m2$.

A.2 Additional Tables

Table A.1: Model Horse Race – Full Sample

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using the full corporate bond fund sample. We estimate the relation between flow and the decile ranking of a fund share class based on different performance measures by running the regression:

$$F_{p,t} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{p,t-1} + \mu_t + e_{p,t}.$$

$F_{p,t}$ is the fund flow of mutual fund share class p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund share class p in month $t-1$ is in the decile i based on the first model and decile j based on the second model (excluding the dummy variable for $i = 5$ and $j = 5$). $X_{p,t-1}$ contains the following control variables (observed at the end of month $t-1$): lagged fund flow, lagged expense ratio, a dummy for no-load share classes, return standard deviation estimated over the prior twelve months, log of fund share class size, log fund age, as well as Morningstar rating dummies. We also include time fixed effects (μ_t).

Candidate performance measures are: the Sharpe ratio (SR), the raw return (RR), and the alphas of the single-factor model with bond market factor (Cb), the two-factor model with both bond and stock market factors (Csb), the Bekaert & De Santis (2021) three-factor model (B3), the Elton et al. (1995) four-factor model (E4), the Fama & French (2015) five-factor model for bonds (F5), and an augmented F5 model with liquidity and momentum factors (F7).

For each pairwise comparison, we have 45 b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji}, we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. The table reports the results of two hypothesis tests: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of positive and negative differences equals 50%. We test the first hypothesis with a Wald test and the second with a Binomial test. We present a "winning model" if the sum of coefficient differences is significantly different from zero. "–" indicates that there is no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table A.1: Model Horse Race – Full Sample (continued)

A. Sharpe ratio

Winning model	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**
Losing model	**RR**	**Cb**	**Csb**	**B3**	**E4**	**F5**	**F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	17.42***	14.16***	9.848***	10.43***	6.547***	9.482***	9.462***
p-Value	0.000	0.000	0.000	0.000	0.001	0.000	0.000
% of coeff. Diff. >0	91.11***	93.33***	88.89***	88.89***	73.33***	84.44***	86.67***
Binomial p-Value	0.000	0.000	0.000	0.000	0.001	0.000	0.000

B. Raw return

Winning model	–	–	–	**E4**	–	–
Losing model	–	–	–	**RR**	–	–
	RR_Cb	RR_Csb	RR_B3	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-5.286	-4.041	-1.964	-5.508**	-1.456	-0.994
p-Value	0.154	0.107	0.380	0.016	0.490	0.628
% of coeff. Diff. >0	28.89***	31.11***	40.00	15.56***	35.56**	40.00
Binomial p-Value	0.003	0.008	0.116	0.000	0.036	0.1163

C. CAPM bond

Winning model	–	–	**E4**	–	–
Losing model	–	–	**Cb**	–	–
	Cb_Csb	Cb_B3	Cb_E4	Cb_F5	Cb_F7
Sum of coeff. Diff.	-2.308	-0.176	-4.626**	0.180	0.936
p-Value	0.373	0.938	0.046	0.938	0.675
% of coeff. Diff. >0	40.00	57.78	20.00***	57.78	53.33
Binomial p-Value	0.116	0.186	0.000	0.186	0.383

D. CAPM stock + bond

Winning model	–	**E4**	–	–
Losing model	–	**Csb**	–	–
	Csb_B3	Csb_E4	Csb_F5	Csb_F7
Sum of coeff. Diff.	1.927	-4.841*	3.100	3.692
p-Value	0.463	0.089	0.235	0.139
% of coeff. Diff. >0	64.44**	42.22	62.22*	71.11***
Binomial p-Value	0.036	0.186	0.068	0.003

E. B3

Winning model	**E4**	–	–
Losing model	**B3**	–	–
	B3_E4	B3_F5	B3_F7
Sum of coeff. Diff.	-4.629*	1.458	1.561
p-Value	0.085	0.562	0.511
% of coeff. Diff. >0	24.44***	55.56	60.00
Binomial p-Value	0.000	0.276	0.116

F. E4

Winning model	**E4**	**E4**
Losing model	**F5**	**F7**
	E4_F5	E4_F7
Sum of coeff. Diff.	7.710***	7.593***
p-Value	0.006	0.002
% of coeff. Diff. >0	75.56***	86.67***
Binomial p-Value	0.000	0.000

G. F5

Winning model	–
Losing model	–
	F5_F7
Sum of coeff. Diff.	3.569
p-Value	0.456
% of coeff. Diff. >0	53.33
Binomial p-Value	0.383

Table A.2: Model Horse Race – Corporate Bond Fund Subsamples

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using several subsamples of corporate bond fund share classes. We separate the corporate bond mutual fund share classes into those oriented to retail and institutional investors (Panel A), those investing into high-yield and investment-grade segments (Panel B), those with and without rear-load fees (Panel C), and those with mainly corporate bond holdings and non-trivial holdings in other asset classes (Panel D). In each case, we perform the analysis separately for the corporate bond fund subcategory indicated in the corresponding panel headings. We estimate the relation between flow and the decile ranking of a fund share class based on different performance measures by running the regression:

$$F_{p,t} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{p,t-1} + \mu_t + e_{p,t}.$$

$F_{p,t}$ is the fund flow of mutual fund share class p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund share class p in month $t-1$ is in the decile i based on the first model and decile j based on the second model (excluding the dummy variable for $i = 5$ and $j = 5$). $X_{p,t-1}$ contains the following control variables (observed at the end of month $t-1$): lagged fund flow, lagged expense ratio, a dummy for no-load share classes, return standard deviation estimated over the prior twelve months, log of fund share class size, log fund age, as well as Morningstar rating dummies. We also include time fixed effects (μ_t).

Candidate performance measures are: the Sharpe ratio (SR), the raw return (RR), and the alphas of the single-factor model with bond market factor (Cb), the two-factor model with both bond and stock market factors (Csb), the Bekaert & De Santis (2021) three-factor model (B3), the Elton et al. (1995) four-factor model (E4), the Fama & French (2015) five-factor model for bonds (F5), and an augmented F5 model with liquidity and momentum factors (F7).

For each pairwise comparison, we have 45 b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji}, we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. The table reports the results of two hypothesis tests: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of positive and negative differences equals 50%. We test the first hypothesis with a Wald test and the second with a Binomial test. We present a "winning model" if the sum of coefficient differences is significantly different from zero. "–" indicates that there is no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table A.2: Model Horse Race – Corporate Bond Fund Subsamples (continued)

A.1 Retail-oriented bond fund share classes

Winning model	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**
Losing model	**RR**	**Cb**	**Csb**	**B3**	**E4**	**F5**	**F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	15.19***	13.16***	12.78***	12.67***	9.043***	10.48***	10.41***
p-Value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
% of coeff. Diff. >0	91.11***	93.33***	95.56***	84.44***	80.00***	86.67***	82.22***
Binomial p-Value	0.000	0.000	0.000	0.000	0.000	0.000	0.000

A.2 Institutional-oriented bond fund share classes

Winning model	**SR**	**SR**	–	**SR**	–	**SR**	**SR**
Losing model	**RR**	**Cb**	–	**B3**	–	**F5**	**F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	26.76***	21.33***	6.939	8.423**	5.324	10.94***	9.588***
p-Value	0.000	0.004	0.125	0.042	0.155	0.003	0.006
% of coeff. Diff. >0	86.67***	80.00***	60.00	55.56	62.22*	75.56***	71.11***
Binomial p-Value	0.000	0.000	0.116	0.276	0.068	0.000	0.003

B.1 High-yield bond fund share classes

Winning model	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**
Losing model	**RR**	**Cb**	**Csb**	**B3**	**E4**	**F5**	**F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	21.56***	21.12***	19.47***	17.61***	9.331**	12.05***	10.09***
p-Value	0.000	0.000	0.000	0.000	0.013	0.000	0.001
% of coeff. Diff. >0	88.89***	95.56***	95.56***	88.89***	71.11***	86.67***	73.33***
Binomial p-Value	0.000	0.000	0.000	0.000	0.003	0.000	0.001

B.2 Investment-grade bond fund share classes

Winning model	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**
Losing model	**RR**	**Cb**	**Csb**	**B3**	**E4**	**F5**	**F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	14.97***	6.599**	4.358*	6.936***	5.077**	8.062***	8.670***
p-Value	0.000	0.033	0.061	0.003	0.013	0.000	0.000
% of coeff. Diff. >0	86.67***	77.78***	62.22*	82.22***	64.44**	75.56***	77.78***
Binomial p-Value	0.000	0.000	0.068	0.000	0.036	0.000	0.000

Table A.2: Model Horse Race – Corporate Bond Fund Subsamples (continued 2)

C.1 Rear-load bond fund share classes

Winning model	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**
Losing model	**RR**	**Cb**	**Csb**	**B3**	**E4**	**F5**	**F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	11.14***	13.72***	11.67***	10.06***	5.793**	8.402***	8.034***
p-Value	0.001	0.000	0.000	0.000	0.026	0.000	0.000
% of coeff. Diff. >0	66.67**	80.00***	88.89***	75.56***	71.11***	68.89***	75.56***
Binomial p-Value	0.018	0.000	0.000	0.000	0.003	0.008	0.000

C.2 Non-rear-load bond fund share classes

Winning model	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**
Losing model	**RR**	**Cb**	**Csb**	**B3**	**E4**	**F5**	**F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	21.44***	13.78***	8.624***	10.55***	6.993***	10.10***	10.21***
p-Value	0.000	0.000	0.001	0.000	0.002	0.000	0.000
% of coeff. Diff. >0	91.11***	88.89***	80.00***	86.67***	73.33***	88.89***	77.78***
Binomial p-Value	0.000	0.000	0.000	0.000	0.001	0.000	0.000

D.1 Mainly corporate bond holdings

Winning model	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**
Losing model	**RR**	**Cb**	**Csb**	**B3**	**E4**	**F5**	**F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	14.25***	10.26**	10.30***	9.304***	7.028**	11.65***	10.89***
p-Value	0.003	0.017	0.002	0.010	0.041	0.000	0.000
% of coeff. Diff. >0	71.11***	77.78***	75.56***	71.11***	66.67**	68.89***	73.33***
Binomial p-Value	0.003	0.000	0.000	0.003	0.018	0.008	0.001

D.2 Non-trivial holdings in other asset classes

Winning model	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**
Losing model	**RR**	**Cb**	**Csb**	**B3**	**E4**	**F5**	**F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	17.79***	14.14***	8.952***	9.419***	6.381***	9.104***	9.136***
p-Value	0.000	0.000	0.001	0.000	0.002	0.000	0.000
% of coeff. Diff. >0	93.33***	88.89***	84.44***	80.00***	82.22***	80.00***	84.44***
Binomial p-Value	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table A.3: Model Horse Race – Corporate Bond Fund Subsamples (Raw Return)

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using several subsamples of corporate bond fund share classes. We separate the corporate bond mutual fund share classes into those oriented to retail and institutional investors (Panel A), those investing into high-yield and investment-grade segments (Panel B), those with and without rear-load fees (Panel C), and those with mainly corporate bond holdings and non-trivial holdings in other asset classes (Panel D). In each case, we perform the analysis separately for the corporate bond fund subcategory indicated in the corresponding panel headings. We estimate the relation between flow and the decile ranking of a fund share class based on different performance measures by running the regression:

$$F_{p,t} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{p,t-1} + \mu_t + e_{p,t}.$$

$F_{p,t}$ is the fund flow of mutual fund share class p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund share class p in month $t-1$ is in the decile i based on the first model and decile j based on the second model (excluding the dummy variable for $i = 5$ and $j = 5$). $X_{p,t-1}$ contains the following control variables (observed at the end of month $t-1$): lagged fund flow, lagged expense ratio, a dummy for no-load share classes, return standard deviation estimated over the prior twelve months, log of fund share class size, log fund age, as well as Morningstar rating dummies. We also include time fixed effects (μ_t).

Candidate performance measures are: the raw return (RR) and the alphas of the single-factor model with bond market factor (Cb), the two-factor model with both bond and stock market factors (Csb), the Bekaert & De Santis (2021) three-factor model (B3), the Elton et al. (1995) four-factor model (E4), the Fama & French (2015) five-factor model for bonds (F5), and an augmented F5 model with liquidity and momentum factors (F7).

For each pairwise comparison, we have 45 b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji}, we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. The table reports the results of two hypothesis tests: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of positive and negative differences equals 50%. We test the first hypothesis with a Wald test and the second with a Binomial test. We present a "winning model" if the sum of coefficient differences is significantly different from zero. "–" indicates that there is no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table A.3: Model Horse Race – Corporate Bond Fund Subsamples (Raw Return) (continued)

A.1 Retail-oriented bond fund share classes

Winning model	–	–	–	–	–	–
Losing model	–	–	–	–	–	–
	RR_Cb	RR_Csb	RR_B3	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-2.805	0.725	1.385	-2.070	1.086	1.508
p-Value	0.462	0.780	0.563	0.396	0.619	0.470
% of coeff. Diff. >0	44.44	55.56	57.78	37.78*	60.00	53.33
Binomial p-Value	0.276	0.276	0.186	0.068	0.116	0.383

A.2 Institutional-oriented bond fund share classes

Winning model	–	**Csb**	**B3**	**E4**	–	–
Losing model	–	**RR**	**RR**	**RR**	–	–
	RR_Cb	RR_Csb	RR_B3	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-10.65	-13.20***	-7.930*	-12.08***	-6.652	-6.554
p-Value	0.102	0.004	0.063	0.006	0.111	0.111
% of coeff. Diff. >0	40.00	24.44***	48.89	31.11***	35.56**	42.22
Binomial p-Value	0.116	0.000	0.500	0.008	0.036	0.186

B.1 High-yield bond fund share classes

Winning model	–	–	–	–	–	–
Losing model	–	–	–	–	–	–
	RR_Cb	RR_Csb	RR_B3	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	10.07	5.005	1.493	-5.820	-4.898	-4.524
p-Value	0.254	0.189	0.673	0.139	0.180	0.177
% of coeff. Diff. >0	68.89***	55.56	55.56	33.33**	35.56**	37.78*
Binomial p-Value	0.008	0.276	0.276	0.018	0.036	0.068

B.2 Investment-grade bond fund share classes

Winning model	**Cb**	**Csb**	–	**E4**	–	–
Losing model	**RR**	**RR**	–	**RR**	–	–
	RR_Cb	RR_Csb	RR_B3	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-8.554**	-6.217**	-2.812	-4.740*	-0.042	0.491
p-Value	0.018	0.023	0.262	0.055	0.985	0.830
% of coeff. Diff. >0	20.00***	26.67***	24.44***	31.11***	42.22	48.89
Binomial p-Value	0.000	0.001	0.000	0.008	0.186	0.500

Table A.3: Model Horse Race – Corporate Bond Fund Subsamples (Raw Return) (continued 2)

C.1 Rear-load bond fund share classes

Winning model	–	–	–	–	–	–
Losing model	–	–	–	–	–	–
	RR_Cb	RR_Csb	RR_B3	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	4.258	3.167	2.221	-1.349	1.129	1.555
p-Value	0.358	0.279	0.424	0.626	0.645	0.513
% of coeff. Diff. >0	71.11***	66.67**	62.22*	46.67	46.67	53.33
Binomial p-Value	0.003	0.018	0.068	0.383	0.383	0.383

C.2 Non-rear-load bond fund share classes

Winning model	**Cb**	**Csb**	**B3**	**E4**	–	–
Losing model	**RR**	**RR**	**RR**	**RR**	–	–
	RR_Cb	RR_Csb	RR_B3	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-10.40**	-8.149***	-4.395*	-7.727***	-2.796	-2.454
p-Value	0.011	0.006	0.096	0.005	0.271	0.323
% of coeff. Diff. >0	15.56***	22.22***	28.89***	17.78***	37.78*	44.44
Binomial p-Value	0.000	0.000	0.003	0.000	0.068	0.276

D.1 Mainly corporate bond holdings

Winning model	–	–	–	–	–	–
Losing model	–	–	–	–	–	–
	RR_Cb	RR_Csb	RR_B3	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-0.984	0.356	0.609	-1.964	1.623	1.878
p-Value	0.871	0.921	0.863	0.583	0.637	0.557
% of coeff. Diff. >0	48.89	51.11	53.33	46.67	51.11	53.33
Binomial p-Value	0.500	0.500	0.383	0.383	0.500	0.383

D.2 Non-trivial holdings in other asset classes

Winning model	**Cb**	**Csb**	–	**E4**	–	–
Losing model	**RR**	**RR**	–	**RR**	–	–
	RR_Cb	RR_Csb	RR_B3	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-6.977*	-5.445**	-3.165	-6.003**	-1.787	-1.500
p-Value	0.083	0.046	0.198	0.013	0.436	0.507
% of coeff. Diff. >0	28.89***	28.89***	24.44***	26.67***	42.22	44.44
Binomial p-Value	0.003	0.003	0.000	0.001	0.186	0.276

Table A.4: Model Horse Race – Aggregate Illiquidity Regimes

This table presents the results of pairwise comparisons of different performance measures to explain fund flows during different aggregate illiquidity regimes. We use the VIX (Panel A) and the TED spread (Panel B) to capture the aggregate illiquidity. In each case, we perform the analysis separately for the different illiquidity regimes as defined by whether the corresponding variables exceed their averages. We estimate the relation between flow and the decile ranking of a fund share class based on different performance measures by running the regression:

$$F_{p,t} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{p,t-1} + \mu_t + e_{p,t}.$$

$F_{p,t}$ is the fund flow of mutual fund share class p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund share class p in month $t-1$ is in the decile i based on the first model and decile j based on the second model (excluding the dummy variable for $i = 5$ and $j = 5$). $X_{p,t-1}$ contains the following control variables (observed at the end of month $t-1$): lagged fund flow, lagged expense ratio, a dummy for no-load share classes, return standard deviation estimated over the prior twelve months, log of fund share class size, log fund age, as well as Morningstar rating dummies. We also include time fixed effects (μ_t).

Candidate performance measures are: the Sharpe ratio (SR), the raw return (RR), and the alphas of the single-factor model with bond market factor (Cb), the two-factor model with both bond and stock market factors (Csb), the Bekaert & De Santis (2021) three-factor model (B3), the Elton et al. (1995) four-factor model (E4), the Fama & French (2015) five-factor model for bonds (F5), and an augmented F5 model with liquidity and momentum factors (F7).

For each pairwise comparison, we have 45 b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji}, we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. The table reports the results of two hypothesis tests: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of positive and negative differences equals 50%. We test the first hypothesis with a Wald test and the second with a Binomial test. We present a "winning model" if the sum of coefficient differences is significantly different from zero. "–" indicates that there is no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table A.4: Model Horse Race – Aggregate Illiquidity Regimes (continued)

A.1 High-VIX sample

Winning model	**SR**	**SR**	**SR**	**SR**	–	**SR**	**SR**
Losing model	**RR**	**Cb**	**Csb**	**B3**	–	**F5**	**F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	25.21***	15.46***	8.095**	9.779***	4.187	8.358***	9.821***
p-Value	0.000	0.001	0.018	0.001	0.119	0.004	0.000
% of coeff. Diff. >0	82.22***	77.78***	66.67**	84.44***	66.67**	71.11***	77.78***
Binomial p-Value	0.000	0.000	0.018	0.000	0.018	0.003	0.000

A.2 Low-VIX sample

Winning model	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**
Losing model	**RR**	**Cb**	**Csb**	**B3**	**E4**	**F5**	**F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	13.88***	13.39***	11.99***	11.98***	9.665***	11.47***	10.25***
p-Value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
% of coeff. Diff. >0	93.33***	86.67***	91.11***	82.22***	82.22***	91.11***	82.22***
Binomial p-Value	0.000	0.000	0.000	0.000	0.000	0.000	0.000

B.1 High-TED-spread sample

Winning model	**SR**	**SR**	**SR**	**SR**	–	**SR**	**SR**
Losing model	**RR**	**Cb**	**Csb**	**B3**	–	**F5**	**F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	12.05***	15.70***	10.88***	12.23***	3.774	6.465*	8.846***
p-Value	0.007	0.001	0.007	0.001	0.251	0.062	0.009
% of coeff. Diff. >0	73.33***	82.22***	71.11***	73.33***	62.22*	64.44**	66.67**
Binomial p-Value	0.001	0.000	0.003	0.001	0.068	0.036	0.018

B.2 Low-TED-spread sample

Winning model	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**
Losing model	**RR**	**Cb**	**Csb**	**B3**	**E4**	**F5**	**F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	18.47***	13.11***	9.010***	9.286***	7.139***	10.25***	9.435***
p-Value	0.000	0.000	0.001	0.000	0.001	0.000	0.000
% of coeff. Diff. >0	93.33***	93.33***	86.67***	80.00***	73.33***	84.44***	86.67***
Binomial p-Value	0.000	0.000	0.000	0.000	0.001	0.000	0.000

Table A.5: Model Horse Race – Aggregate Illiquidity Regimes (Raw Return)

This table presents the results of pairwise comparisons of different performance measures to explain fund flows during different aggregate illiquidity regimes. We use the VIX (Panel A) and the TED spread (Panel B) to capture the aggregate illiquidity regimes. In each case, we perform the analysis separately for the different illiquidity regimes as defined by whether the corresponding variables exceed their averages. We estimate the relation between flow and the decile ranking of a fund share class based on different performance measures by running the regression:

$$F_{p,t} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{p,t-1} + \mu_t + e_{p,t}.$$

$F_{p,t}$ is the fund flow of mutual fund share class p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund share class p in month $t-1$ is in the decile i based on the first model and decile j based on the second model (excluding the dummy variable for $i = 5$ and $j = 5$). $X_{p,t-1}$ contains the following control variables (observed at the end of month $t-1$): lagged fund flow, lagged expense ratio, a dummy for no-load share classes, return standard deviation estimated over the prior twelve months, log of fund share class size, log fund age, as well as Morningstar rating dummies. We also include time fixed effects (μ_t).

Candidate performance measures are: the raw return (RR) and the alphas of the single-factor model with bond market factor (Cb), the two-factor model with both bond and stock market factors (Csb), the Bekaert & De Santis (2021) three-factor model (B3), the Elton et al. (1995) four-factor model (E4), the Fama & French (2015) five-factor model for bonds (F5), and an augmented F5 model with liquidity and momentum factors (F7).

For each pairwise comparison, we have 45 b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji}, we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. The table reports the results of two hypothesis tests: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of positive and negative differences equals 50%. We test the first hypothesis with a Wald test and the second with a Binomial test. We present a "winning model" if the sum of coefficient differences is significantly different from zero. "–" indicates that there is no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table A.5: Model Horse Race – Aggregate Illiquidity Regimes (Raw Return) (continued)

A.1 High-VIX sample

Winning model	**Cb**	**Csb**	–	**E4**	–	–
Losing model	**RR**	**RR**	–	**RR**	–	–
	RR_Cb	RR_Csb	RR_B3	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-12.72**	-9.347**	-4.227	-10.27***	-5.022	-2.917
p-Value	0.029	0.014	0.194	0.003	0.129	0.362
% of coeff. Diff. >0	26.67***	26.67***	37.78*	20.00***	33.33**	37.78*
Binomial p-Value	0.001	0.001	0.068	0.000	0.018	0.068

A.2 Low-VIX sample

Winning model	–	–	–	–	–	–
Losing model	–	–	–	–	–	–
	RR_Cb	RR_Csb	RR_B3	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-1.132	-0.042	0.823	-1.220	2.034	1.212
p-Value	0.795	0.989	0.768	0.665	0.416	0.614
% of coeff. Diff. >0	42.22	60.00	46.67	42.22	55.56	68.89***
Binomial p-Value	0.186	0.116	0.383	0.186	0.276	0.008

B.1 High-TED-spread sample

Winning model	–	–	–	–	–	–
Losing model	–	–	–	–	–	–
	RR_Cb	RR_Csb	RR_B3	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	1.902	-0.223	4.101	-3.523	-0.518	1.817
p-Value	0.703	0.959	0.274	0.323	0.888	0.607
% of coeff. Diff. >0	60.00	46.67	53.33	40.00	51.11	62.22*
Binomial p-Value	0.116	0.383	0.383	0.116	0.500	0.068

B.2 Low-TED-spread sample

Winning model	–	**Csb**	–	**E4**	–	–
Losing model	–	**RR**	–	**RR**	–	–
	RR_Cb	RR_Csb	RR_B3	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-7.133	-5.224*	-4.263	-5.820**	-1.488	-1.756
p-Value	0.111	0.071	0.110	0.030	0.537	0.456
% of coeff. Diff. >0	24.44***	28.89***	26.67***	20.00***	35.56**	48.89
Binomial p-Value	0.000	0.003	0.001	0.000	0.036	0.500

Table A.6: Model Horse Race with Morningstar Ratings

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using the full corporate bond fund sample. We estimate the relation between flow and the quintile ranking of a fund share class based on different performance measures by running the regression:

$$F_{p,t} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{p,t-1} + \mu_t + e_{p,t}.$$

$F_{p,t}$ is the fund flow of mutual fund share class p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund share class p in month $t-1$ is in the quintile i based on the first model and quintile j based on the second model (excluding the dummy variable for $i = 3$ and $j = 3$). $X_{p,t-1}$ contains the following control variables (observed at the end of month $t-1$): lagged fund flow, lagged expense ratio, a dummy for no-load share classes, return standard deviation estimated over the prior twelve months, log of fund share class size, and log fund age. We also include time fixed effects (μ_t).

Candidate performance measures are: the Morningstar ratings (MS), the Sharpe ratio (SR), the raw return (RR), and the alphas of the single-factor model with bond market factor (Cb), the two-factor model with both bond and stock market factors (Csb), the Bekaert & De Santis (2021) three-factor model (B3), the Elton et al. (1995) four-factor model (E4), the Fama & French (2015) five-factor model for bonds (F5), and an augmented F5 model with liquidity and momentum factors (F7).

For each pairwise comparison, we have 10 b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji}, we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. The table reports the results of two hypothesis tests: (1) H_0: The summed difference across all 10 comparisons is equal to zero, (2) H_0: The proportion of positive and negative differences equals 50%. We test the first hypothesis with a Wald test and the second with a Binomial test. We present a "winning model" if the sum of coefficient differences is significantly different from zero. "–" indicates that there is no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Morningstar ratings

Winning model **Losing model**	**MS** **SR**	**MS** **RR**	**MS** **Cb**	**MS** **Csb**	**MS** **B3**	**MS** **E4**	**MS** **F5**	**MS** **F7**
	MS_SR	MS_RR	MS_Cb	MS_Csb	MS_B3	MS_E4	MS_F5	MS_F7
Sum of coeff. Diff.	5.867***	7.202***	6.615***	6.642***	6.757***	6.197***	6.670***	6.749***
p-Value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
% of coeff. Diff. >0	100.00***	100.00***	100.00***	100.00***	100.00***	100.00***	100.00***	100.00***
Binomial p-Value	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

Table A.7: Single Flow–Performance Sensitivity Estimations

This table reports the beta estimates from the following equation for different risk models:

$$\beta_{flow,performance} = \frac{cov(\Phi(F_{p,t}), (\Phi(\alpha_{p,t-1}))}{var(\Phi(\alpha_{p,t-1}))} > 0,$$

where Φ is a function that returns the sign of a real number, taking values of 1 for a positive number, -1 for a negative number, and 0 for zero. The sample period is from 1996 to June 2017. For the ease of interpretation, the table reports $(\beta_{flow,performance} + 1)/2$ which denotes the average probability that the sign of the fund flow [*S(flow)*] is positive (negative) conditional on the sign of the performance measure [*S(performance)*] being positive (negative). Each row corresponds to a different performance measure. p-values are based on a t-test of $\beta_{flow,performance}$ using double-clustered standard errors (by fund and month).

Candidate performance measures are: the Morningstar rating = 5, Morningstar rating ≥ 4, Morningstar rating ≥ 3, the Sharpe ratio (SR), the raw return (RR), and the alphas of the single-factor model with bond market factor (Cb) and the Elton et al. (1995) four-factor model (E4).

	S (Flow)	p-Value
S (MS rating = 5)	63.03	0.00
S (MS rating $\geq$ 4)	60.42	0.00
S (MS rating $\geq$ 3)	59.34	0.00
S (Sharpe ratio)	54.25	0.00
S (Raw return)	54.18	0.00
S (Cb alpha)	53.02	0.00
S (E4 alpha)	52.78	0.00

Table A.8: Flow–Performance Model Horse Race: Berk & Van Binsbergen (2016) Pairwise Model Comparisons

This table reports the results from pairwise comparisons of Morningstar rating dummies, the Sharpe ratio, raw returns, and different factor model alphas as in Berk & Van Binsbergen (2016). Columns (1) and (2) provide the coefficient estimates and the double-clustered (by fund and month) t-statistics of univariate regressions of signed flows on signed outperformance. Columns (3) to (9) provide the double-clustered t-statistics of the pairwise test coefficients b_1 in the following equation:

$$\Phi(F_{p,t}) = a + b_1 \left(\frac{\Phi(\alpha^{m1}_{p,t-1})}{var(\Phi(\alpha^{m1}_{p,t-1}))} - \frac{\Phi(\alpha^{m2}_{p,t-1})}{var(\Phi(\alpha^{m2}_{p,t-1}))} \right) + \xi_{p,t},$$

where we compare the flow–performance regression coefficients, $\beta_{flow,performance}$ of two models $m1$ and $m2$.

Candidate performance measures are: the Morningstar rating = 5, Morningstar rating ≥ 4, Morningstar rating ≥ 3, the Sharpe ratio (SR), the raw return (RR), and the alphas of the single-factor model with bond market factor (Cb) and the Elton et al. (1995) four-factor model (E4).

	β	Uni. t-stat	Rating = 5	Rating ≥ 4	Rating ≥ 3	Sharpe ratio	Raw return	Cb alpha	E4 alpha
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Rating = 5	0.2420	20.94	0.00	4.19	4.46	8.59	8.25	10.82	14.70
Rating ≥ 4	0.2021	23.81		0.00	2.44	7.49	6.67	10.77	16.68
Rating ≥ 3	0.1949	20.14			0.00	6.04	5.23	8.57	13.46
Sharpe ratio	0.0850	8.98				0.00	-0.37	2.11	4.67
Raw return	0.0836	8.58					0.00	1.86	4.58
Cb alpha	0.0603	7.62						0.00	4.81
E4 alpha	0.0556	9.98							0.00

Table A.9: Model Horse Race – Robustness with Fama–MacBeth Regressions

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using the full corporate bond fund sample. We estimate the relation between flow and the decile ranking of a fund share class based on different performance measures by running the Fama & MacBeth (1973) regression:

$$F_{p,t} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{p,t-1} + e_{p,t}.$$

$F_{p,t}$ is the fund flow of mutual fund share class p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund share class p in month $t-1$ is in the decile i based on the first model and decile j based on the second model (excluding the dummy variable for $i = 5$ and $j = 5$). $X_{p,t-1}$ contains the following control variables (observed at the end of month $t-1$): lagged fund flow, lagged expense ratio, a dummy for no-load share classes, return standard deviation estimated over the prior twelve months, log of fund share class size, log fund age, as well as Morningstar rating dummies.

Candidate performance measures are: the Sharpe ratio (SR), the raw return (RR), and the alphas of the single-factor model with bond market factor (Cb), the two-factor model with both bond and stock market factors (Csb), the Bekaert & De Santis (2021) three-factor model (B3), the Elton et al. (1995) four-factor model (E4), the Fama & French (2015) five-factor model for bonds (F5), and an augmented F5 model with liquidity and momentum factors (F7).

For each pairwise comparison, we have 45 b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji}, we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. The table reports the results of two hypothesis tests: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of positive and negative differences equals 50%. We test the first hypothesis with a Wald test and the second with a Binomial test. We present a "winning model" if the sum of coefficient differences is significantly different from zero. "–" indicates that there is no significant difference. The standard errors are calculated by the Newey & West (1987) procedure with six lags. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Sharpe ratio

Winning model **Losing model**	**SR** **RR**	**SR** **Cb**	**SR** **Csb**	**SR** **B3**	**SR** **E4**	**SR** **F5**	**SR** **F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	8.687***	7.138***	7.201***	8.033***	4.553***	7.806***	7.561***
p-Value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
% of coeff. Diff. >0	84.44***	84.44***	82.22***	77.78***	66.67**	86.67***	82.22***
Binomial p-Value	0.000	0.000	0.000	0.000	0.018	0.000	0.000

Table A.10: Model Horse Race with Quintile Sorts

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using the full corporate bond fund sample. We estimate the relation between flow and the quintile ranking of a fund share class based on different performance measures by running the regression:

$$F_{p,t} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{p,t-1} + \mu_t + e_{p,t}.$$

$F_{p,t}$ is the fund flow of mutual fund share class p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund share class p in month $t-1$ is in the quintile i based on the first model and quintile j based on the second model (excluding the dummy variable for $i = 3$ and $j = 3$). $X_{p,t-1}$ contains the following control variables (observed at the end of month $t-1$): lagged fund flow, lagged expense ratio, a dummy for no-load share classes, return standard deviation estimated over the prior twelve months, log of fund share class size, and log fund age, as well as Morningstar rating dummies. We also include time fixed effects (μ_t).

Candidate performance measures are: the Sharpe ratio (SR), the raw return (RR), and the alphas of the single-factor model with bond market factor (Cb), the two-factor model with both bond and stock market factors (Csb), the Bekaert & De Santis (2021) three-factor model (B3), the Elton et al. (1995) four-factor model (E4), the Fama & French (2015) five-factor model for bonds (F5), and an augmented F5 model with liquidity and momentum factors (F7).

For each pairwise comparison, we have 10 b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji}, we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. The table reports the results of two hypothesis tests: (1) H_0: The summed difference across all 10 comparisons is equal to zero, (2) H_0: The proportion of positive and negative differences equals 50%. We test the first hypothesis with a Wald test and the second with a Binomial test. We present a "winning model" if the sum of coefficient differences is significantly different from zero. "–" indicates that there is no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table A.10: Model Horse Race with Quintile Sorts (continued)

A. Sharpe ratio

Winning model **Losing model**	**SR** **RR**	**SR** **Cb**	**SR** **Csb**	**SR** **B3**	**SR** **E4**	**SR** **F5**	**SR** **F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	3.894***	3.027***	2.115***	2.328***	1.311***	2.046***	2.070***
p-Value	0.000	0.000	0.000	0.000	0.003	0.000	0.000
% of coeff. Diff. >0	100.00***	100.00***	100.00***	100.00***	90.00**	100.00***	100.00***
Binomial p-Value	0.001	0.001	0.001	0.001	0.011	0.001	0.001

B. Raw return

Winning model **Losing model**	**Cb** **RR**	**Csb** **RR**	– –	**E4** **RR**	– –	– –
	RR_Cb	RR_Csb	RR_B3	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-1.450*	-1.021**	-0.456	-1.377***	-0.469	-0.359
p-Value	0.062	0.048	0.338	0.004	0.308	0.403
% of coeff. Diff. >0	10.000**	10.000**	30.00	0.000***	20.00*	20.00*
Binomial p-Value	0.011	0.011	0.172	0.001	0.055	0.055

C. CAPM bond

Winning model **Losing model**	– –	– –	**E4** **Cb**	– –	– –
	Cb_Csb	Cb_B3	Cb_E4	Cb_F5	Cb_F7
Sum of coeff. Diff.	-0.534	0.185	-0.988*	0.154	0.269
p-Value	0.333	0.719	0.062	0.760	0.573
% of coeff. Diff. >0	40.00	70.00	20.00*	70.00	60.00
Binomial p-Value	0.377	0.172	0.055	0.172	0.377

D. CAPM stock + bond

Winning model **Losing model**	– –	**E4** **Csb**	– –	– –
	Csb_B3	Csb_E4	Csb_F5	Csb_F7
Sum of coeff. Diff.	0.648	-1.166*	0.683	0.695
p-Value	0.349	0.061	0.233	0.201
% of coeff. Diff. >0	60.00	0.000***	80.00*	100.00***
Binomial p-Value	0.377	0.001	0.055	0.001

E. B3

Winning model **Losing model**	**E4** **B3**	– –	– –
	B3_E4	B3_F5	B3_F7
Sum of coeff. Diff.	-1.302**	0.132	0.199
p-Value	0.022	0.822	0.710
% of coeff. Diff. >0	10.000**	40.00	80.00*
Binomial p-Value	0.011	0.377	0.055

F. E4

Winning model **Losing model**	**E4** **F5**	**E4** **F7**
	E4_F5	E4_F7
Sum of coeff. Diff.	1.742***	1.707***
p-Value	0.007	0.003
% of coeff. Diff. >0	100.00***	100.00***
Binomial p-Value	0.001	0.001

G. F5

Winning model **Losing model**	– –
	F5_F7
Sum of coeff. Diff.	1.339
p-Value	0.147
% of coeff. Diff. >0	60.00
Binomial p-Value	0.377

Table A.11: Model Horse Race – Alternative Sharpe Ratio Calculations

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using the full bond fund sample. We estimate the relation between flow and the decile ranking of a fund share class based on different performance measures by running the regression:

$$F_{p,t} = a + \sum_{i}\sum_{j} b_{ij} D_{ij,p,t-1} + cX_{p,t-1} + \mu_t + e_{p,t}.$$

$F_{p,t}$ is the fund flow of mutual fund share class p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund share class p in month $t-1$ is in the decile i based on the first measure and decile j based on the second measure. To estimate the model, the dummy variable for $i = 5$ and $j = 5$ is excluded. $X_{p,t-1}$ contains the following control variables (observed at the end of month $t-1$): lagged fund flow, lagged expense ratio, a dummy for no-load share classes, return standard deviation estimated over the prior twelve months, log of fund share class size, log fund age, as well as Morningstar rating dummies. We also include time fixed effects (μ_t).

We consider two alternative ways to calculate the Sharpe ratio: (i) as the ratio of the monthly average twelve-month excess return to the one-year standard deviation in Panel A (SR) and (ii) employing a GARCH (1,1) model using 60-month past returns to estimate fund's variance to align with our method to estimate alphas in Panel B (SR_G). Other candidate performance measures are: the raw return (RR) and the alphas of the single-factor model with bond market factor (Cb), the two-factor model with both bond and stock market factors (Csb), the Bekaert & De Santis (2021) three-factor model (B3), the Elton et al. (1995) four-factor model (E4), the Fama & French (2015) five-factor model for bonds (F5), and an augmented F5 model with liquidity and momentum factors (F7).

For each pairwise comparison, we have 45 b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji}, we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. The table reports the results of two hypothesis tests: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of positive and negative differences equals 50%. We test the first hypothesis with a Wald test and the second with a Binomial test. We present a "winning model" if the sum of coefficient differences is significantly different from zero. "–" indicates that there is no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table A.11: Model Horse Race – Alternative Sharpe Ratio Calculations (continued)

A. Twelve-month return window

Winning model	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**
Losing model	**RR**	**Cb**	**Csb**	**B3**	**E4**	**F5**	**F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	17.71***	17.04***	15.80***	17.27***	14.46***	16.39***	16.68***
p-Value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
% of coeff. Diff. >0	100.00***	100.00***	95.56***	97.78***	91.11***	91.11***	93.33***
Binomial p-Value	0.000	0.000	0.000	0.000	0.000	0.000	0.000

B. Volatility based on a GARCH (1,1) model

Winning model	**SR_G**	**SR_G**	**SR_G**	**SR_G**	**SR_G**	**SR_G**	**SR_G**
Losing model	**RR**	**Cb**	**Csb**	**B3**	**E4**	**F5**	**F7**
	SR_G_RR	SR_G_Cb	SR_G_Csb	SR_G_B3	SR_G_E4	SR_G_F5	SR_G_F7
Sum of coeff. Diff.	13.73***	11.52***	6.726***	8.224***	4.469**	7.550***	7.280***
p-Value	0.000	0.000	0.004	0.000	0.022	0.000	0.000
% of coeff. Diff. >0	75.56***	77.78***	73.33***	75.56***	62.22*	77.78***	80.00***
Binomial p-Value	0.000	0.000	0.001	0.000	0.068	0.000	0.000

Table A.12: Model Horse Race (12-Month Window)

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using the full bond fund sample. We estimate the relation between flow and the decile ranking of a fund share class based on different performance measures by running the regression:

$$F_{p,t} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{p,t-1} + \mu_t + e_{p,t}.$$

$F_{p,t}$ is the fund flow of mutual fund share class p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund share class p in month $t-1$ is in the decile i based on the first measure and decile j based on the second measure. To estimate the model, the dummy variable for $i = 5$ and $j = 5$ is excluded. $X_{p,t-1}$ contains the following control variables (observed at the end of month $t-1$): lagged fund flow, lagged expense ratio, a dummy for no-load share classes, return standard deviation estimated over the prior twelve months, log of fund share class size, log fund age, as well as Morningstar rating dummies. We also include time fixed effects (μ_t).

Candidate performance measures are: the Sharpe ratio (SR), the raw return (RR), and the alphas of the single-factor model with bond market factor (Cb), the two-factor model with both bond and stock market factors (Csb), the Bekaert & De Santis (2021) three-factor model (B3), the Elton et al. (1995) four-factor model (E4), the Fama & French (2015) five-factor model for bonds (F5), and an augmented F5 model with liquidity and momentum factors (F7). Each of the measures is calculated as a weighted average of the prior twelve monthly alphas (or returns for the Sharpe ratio).

For each pairwise comparison, we have 45 b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji}, we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. The table reports the results of two hypothesis tests: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of positive and negative differences equals 50%. We test the first hypothesis with a Wald test and the second with a Binomial test. We present a "winning model" if the sum of coefficient differences is significantly different from zero. "–" indicates that there is no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table A.12: Model Horse Race (12-Month Window) (continued)

A. Sharpe ratio

Winning model	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**
Losing model	**RR**	**Cb**	**Csb**	**B3**	**E4**	**F5**	**F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	16.32***	6.280***	4.494**	5.725***	2.969*	6.377***	7.091***
p-Value	0.000	0.006	0.018	0.001	0.063	0.000	0.000
% of coeff. Diff. >0	95.56***	68.89***	57.78	73.33***	55.56	75.56***	75.56***
Binomial p-Value	0.000	0.008	0.186	0.001	0.276	0.000	0.000

B. Raw return

Winning model	**Cb**	**Csb**	**B3**	**E4**	–	–
Losing model	**RR**	**RR**	**RR**	**RR**	–	–
	RR_Cb	RR_Csb	RR_B3	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-9.834***	-6.461***	-3.741*	-6.672***	-2.114	-0.955
p-Value	0.003	0.003	0.063	0.001	0.263	0.601
% of coeff. Diff. >0	17.78***	22.22***	33.33**	13.33***	37.78*	46.67
Binomial p-Value	0.000	0.000	0.018	0.000	0.068	0.383

C. CAPM bond

Winning model	–	–	–	**Cb**	**Cb**
Losing model	–	–	–	**F5**	**F7**
	Cb_Csb	Cb_B3	Cb_E4	Cb_F5	Cb_F7
Sum of coeff. Diff.	-0.135	1.658	-2.223	3.547**	5.057***
p-Value	0.945	0.392	0.171	0.046	0.003
% of coeff. Diff. >0	51.11	53.33	48.89	64.44**	71.11***
Binomial p-Value	0.500	0.383	0.500	0.036	0.003

D. CAPM stock + bond

Winning model	**Csb**	–	**Csb**	**Csb**
Losing model	**B3**	–	**F5**	**F7**
	Csb_B3	Csb_E4	Csb_F5	Csb_F7
Sum of coeff. Diff.	4.665*	-2.361	5.194**	7.401***
p-Value	0.089	0.254	0.021	0.001
% of coeff. Diff. >0	62.22*	42.22	80.00***	84.44***
Binomial p-Value	0.068	0.186	0.000	0.000

E. B3

Winning model	**E4**	–	–
Losing model	**B3**	–	–
	B3_E4	B3_F5	B3_F7
Sum of coeff. Diff.	-5.335*	1.743	3.175
p-Value	0.056	0.470	0.151
% of coeff. Diff. >0	26.67***	62.22*	60.00
Binomial p-Value	0.001	0.068	0.116

F. E4

Winning model	**E4**	**E4**
Losing model	**F5**	**F7**
	E4_F5	E4_F7
Sum of coeff. Diff.	10.52***	12.36***
p-Value	0.000	0.000
% of coeff. Diff. >0	88.89***	84.44***
Binomial p-Value	0.000	0.000

G. F5

Winning model	–
Losing model	–
	F5_F7
Sum of coeff. Diff.	6.82
p-Value	0.211
% of coeff. Diff. >0	73.33***
Binomial p-Value	0.001

Table A.13: Model Horse Race – Alternative Factor Models

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using the full bond fund sample. We estimate the relation between flow and the decile ranking of a fund share class based on different performance measures by running the regression:

$$F_{p,t} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{p,t-1} + \mu_t + e_{p,t}.$$

$F_{p,t}$ is the fund flow of mutual fund share class p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund share class p in month $t-1$ is in the decile i based on the first model and decile j based on the second model (excluding the dummy variable for $i = 5$ and $j = 5$). $X_{p,t-1}$ contains the following control variables (observed at the end of month $t-1$): lagged fund flow, lagged expense ratio, a dummy for no-load share classes, return standard deviation estimated over the prior twelve months, log of fund share class size, log fund age, as well as Morningstar rating dummies. We also include time fixed effects (μ_t).

The main candidate performance measures are: the Sharpe ratio (SR), the raw return (RR), and the alphas of the single-factor model with bond market factor (Cb), the two-factor model with both bond and stock market factors (Csb), the Bekaert & De Santis (2021) three-factor model (B3), the Elton et al. (1995) four-factor model (E4), the Fama & French (2015) five-factor model for bonds (F5), and an augmented F5 model with liquidity and momentum factors (F7). In addition, we consider the factor models indicated in the respective panel headings.

For each pairwise comparison, we have 45 b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji}, we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. The table reports the results of two hypothesis tests: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of positive and negative differences equals 50%. We test the first hypothesis with a Wald test and the second with a Binomial test. We present a "winning model" if the sum of coefficient differences is significantly different from zero. "–" indicates that there is no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table A.13: Model Horse Race – Alternative Factor Models (continued)

Augumented Fama & French (1993) factor model with bond MOM (MOMb)

Winning model	**SR**	**MOMb**	–	–	–	**E4**	–	–
Losing model	**MOMb**	**RR**	–	–	–	**MOMb**	–	–
	SR_MOMb	RR_MOMb	Cb_MOMb	Csb_MOMb	B3_MOMb	E4_MOMb	F5_MOMb	F7_MOMb
Sum of coeff. Diff.	7.670***	-5.163**	0.094	0.687	-1.508	6.844*	-2.330	-2.631
p-Value	0.000	0.033	0.973	0.819	0.580	0.061	0.665	0.569
% of coeff. Diff. >0	71.11***	28.89***	44.44	53.33	42.22	75.56***	37.78*	42.22
Binomial p-Value	0.003	0.003	0.276	0.383	0.186	0.000	0.068	0.186

Bai et al. (2019) factor model (B4)

Winning model	**SR**	–	–	–	–	**E4**	–	–
Losing model	**B4**	–	–	–	–	**B4**	–	–
	SR_B4	RR_B4	Cb_B4	Csb_B4	B3_B4	E4_B4	F5_B4	F7_B4
Sum of coeff. Diff.	15.13***	1.707	0.233	3.957	3.407	5.155*	0.356	1.018
p-Value	0.000	0.682	0.950	0.141	0.192	0.051	0.889	0.670
% of coeff. Diff. >0	91.11***	55.56	60.00	57.78	62.22*	60.00	51.11	62.22*
Binomial p-Value	0.000	0.276	0.116	0.186	0.068	0.116	0.500	0.068

Chung et al. (2019) factor model (C7)

Winning model	**SR**	–	–	**Csb**	–	**E4**	–	–
Losing model	**C7**	–	–	**C7**	–	**C7**	–	–
	SR_C7	RR_C7	Cb_C7	Csb_C7	B3_C7	E4_C7	F5_C7	F7_C7
Sum of coeff. Diff.	10.26***	-0.190	1.974	3.956*	2.602	8.779***	2.990	0.916
p-Value	0.000	0.923	0.367	0.095	0.264	0.001	0.546	0.820
% of coeff. Diff. >0	91.11***	53.33	66.67**	71.11***	60.00	86.67***	62.22*	62.22*
Binomial p-Value	0.000	0.383	0.018	0.003	0.116	0.000	0.068	0.068

Ludvigson & Ng (2009) macro factors (Macro)

Winning model	**SR**	**RR**	**Cb**	**Csb**	**B3**	**E4**	**F5**	**F7**
Losing model	**Macro**	**Macro**	**Macro**	**Macro**	**Macro**	**Macro**	**Macro**	**Macro**
	SR_Macro	RR_Macro	Cb_Macro	Csb_Macro	B3_Macro	E4_Macro	F5_Macro	F7_Macro
Sum of coeff. Diff.	17.84***	10.22***	9.561***	9.556***	8.074***	10.69***	8.000***	7.414***
p-Value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
% of coeff. Diff. >0	95.56***	84.44***	84.44***	88.89***	86.67***	91.11***	88.89***	88.89***
Binomial p-Value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Augumented Fama & French (2015) factor model with TERM and DEF (FF7e)

Winning model	**SR**	–	–	–	–	**E4**	–	–
Losing model	**FF7e**	–	–	–	–	**FF7e**	–	–
	SR_FF7e	RR_FF7e	Cb_FF7e	Csb_FF7e	B3_FF7e	E4_FF7e	F5_FF7e	F7_FF7e
Sum of coeff. Diff.	9.321***	-1.283	0.993	3.617	1.516	7.238***	2.680	2.802
p-Value	0.000	0.518	0.639	0.1328	0.483	0.005	0.454	0.387
% of coeff. Diff. >0	86.67***	40.00	57.78	62.22*	57.78	84.44***	68.89***	55.56
Binomial p-Value	0.000	0.116	0.186	0.068	0.186	0.000	0.008	0.276

Augumented Hou et al. (2015) factor model with TERM and DEF (HXZ)

Winning model	**SR**	–	–	–	–	**E4**	–	–
Losing model	**HXZ**	–	–	–	–	**HXZ**	–	–
	SR_HXZ	RR_HXZ	Cb_HXZ	Csb_HXZ	B3_HXZ	E4_HXZ	F5_HXZ	F7_HXZ
Sum of coeff. Diff.	8.606***	-2.843	0.135	1.625	-0.200	6.774**	0.393	-1.215
p-Value	0.000	0.153	0.950	0.517	0.931	0.011	0.916	0.684
% of coeff. Diff. >0	82.22***	37.78*	57.78	64.44**	48.89	75.56***	42.22	40.00
Binomial p-Value	0.000	0.068	0.186	0.036	0.500	0.000	0.186	0.116

Augumented Stambaugh & Yuan (2017) factor model with TERM and DEF (M4)

Winning model	**SR**	–	–	–	–	**E4**	–	–
Losing model	**M4**	–	–	–	–	**M4**	–	–
	SR_M4	RR_M4	Cb_M4	Csb_M4	B3_M4	E4_M4	F5_M4	F7_M4
Sum of coeff. Diff.	9.676***	-1.185	0.552	2.957	1.088	7.747***	1.307	0.212
p-Value	0.000	0.549	0.798	0.230	0.634	0.006	0.761	0.951
% of coeff. Diff. >0	91.11***	40.00	55.56	68.89***	48.89	84.44***	48.89	51.11
Binomial p-Value	0.000	0.116	0.276	0.008	0.500	0.000	0.500	0.500

Table A.14: Model Horse Race – Fund-Level Sample

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using fund-level observations of full corporate bond mutual fund sample. We aggregate the fund flows and value-weight the share-class returns and other variables to obtain those on the fund level using total net asset values. We estimate the relation between flow and the decile ranking of a fund based on different performance measures by running the regression:

$$F_{p,t} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{p,t-1} + \mu_t + e_{p,t}.$$

$F_{p,t}$ is the fund flow of mutual fund p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund p in month $t-1$ is in the decile i based on the first model and decile j based on the second model (excluding the dummy variable for $i = 5$ and $j = 5$). $X_{p,t-1}$ contains the following control variables (observed at the end of month $t-1$): lagged fund flow, lagged expense ratio, a dummy for no-load funds, return standard deviation estimated over the prior twelve months, log of fund size, log fund age, as well as Morningstar rating dummies. We also include time fixed effects (μ_t).

Candidate performance measures are: the Sharpe ratio (SR), the raw return (RR), and the alphas of the single-factor model with bond market factor (Cb), the two-factor model with both bond and stock market factors (Csb), the Bekaert & De Santis (2021) three-factor model (B3), the Elton et al. (1995) four-factor model (E4), the Fama & French (2015) five-factor model for bonds (F5), and an augmented F5 model with liquidity and momentum factors (F7).

For each pairwise comparison, we have 45 b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji}, we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. The table reports the results of two hypothesis tests: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of positive and negative differences equals 50%. We test the first hypothesis with a Wald test and the second with a Binomial test. We present a "winning model" if the sum of coefficient differences is significantly different from zero. "–" indicates that there is no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table A.14: Model Horse Race – Fund-Level Sample (continued)

A. Sharpe ratio

Winning model **Losing model**	**SR** **RR**	**SR** **Cb**	**SR** **Csb**	**SR** **B3**	**SR** **E4**	**SR** **F5**	**SR** **F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	19.04***	16.44***	8.051***	8.220***	5.084**	7.678***	7.402***
p-Value	0.000	0.000	0.002	0.001	0.028	0.001	0.001
% of coeff. Diff. >0	86.67***	80.00***	77.78***	73.33***	62.22*	73.33***	73.33***
Binomial p-Value	0.000	0.000	0.000	0.001	0.068	0.001	0.001

B. Raw return

Winning model **Losing model**	– –	**Csb** **RR**	**B3** **RR**	**E4** **RR**	– –	– –
	RR_Cb	RR_Csb	RR_B3	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-4.340	-6.834***	-5.291**	-7.467***	-2.672	-3.167
p-Value	0.262	0.008	0.023	0.002	0.218	0.127
% of coeff. Diff. >0	37.78*	40.00	31.11***	24.44***	40.00	44.44
Binomial p-Value	0.068	0.116	0.008	0.000	0.116	0.276

C. CAPM bond

Winning model **Losing model**	**Csb** **Cb**	– –	**E4** **Cb**	– –	– –
	Cb_Csb	Cb_B3	Cb_E4	Cb_F5	Cb_F7
Sum of coeff. Diff.	-6.215**	-3.199	-6.711***	-1.381	-1.576
p-Value	0.040	0.227	0.009	0.585	0.510
% of coeff. Diff. >0	37.78*	42.22	22.22***	46.67	51.11
Binomial p-Value	0.068	0.186	0.000	0.383	0.500

D. CAPM stock + bond

Winning model **Losing model**	– –	– –	– –	– –
	Csb_B3	Csb_E4	Csb_F5	Csb_F7
Sum of coeff. Diff.	0.704	-5.100	3.618	2.288
p-Value	0.842	0.144	0.173	0.381
% of coeff. Diff. >0	48.89	35.56**	60.00	57.78
Binomial p-Value	0.500	0.036	0.116	0.186

E. B3

Winning model **Losing model**	– –	– –	– –
	B3_E4	B3_F5	B3_F7
Sum of coeff. Diff.	-4.240	4.006	1.902
p-Value	0.176	0.168	0.492
% of coeff. Diff. >0	42.22	62.22*	48.89
Binomial p-Value	0.186	0.068	0.500

F. E4

Winning model **Losing model**	**E4** **F5**	**E4** **F7**
	E4_F5	E4_F7
Sum of coeff. Diff.	6.528*	5.210*
p-Value	0.060	0.082
% of coeff. Diff. >0	68.89***	68.89***
Binomial p-Value	0.008	0.008

G. F5

Winning model **Losing model**	– –
	F5_F7
Sum of coeff. Diff.	-2.691
p-Value	0.692
% of coeff. Diff. >0	40.00
Binomial p-Value	0.180

Table A.15: Model Horse Race – Full Sample (monthFE x MS stars)

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using the full corporate bond fund sample. We estimate the relation between flow and the decile ranking of a fund share class based on different performance measures by running the regression:

$$F_{p,t} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{p,t-1} + e_{p,t}.$$

$F_{p,t}$ is the fund flow of mutual fund share class p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund share class p in month $t-1$ is in the decile i based on the first model and decile j based on the second model (excluding the dummy variable for $i = 5$ and $j = 5$). $X_{p,t-1}$ contains the following control variables (observed at the end of month $t-1$): lagged fund flow, lagged expense ratio, a dummy for no-load share classes, return standard deviation estimated over the prior twelve months, log of fund share class size, log fund age, as well as the month-times-Morningstar ratings fixed effects.

Candidate performance measures are: the Sharpe ratio (SR), the raw return (RR), and the alphas of the single-factor model with bond market factor (Cb), the two-factor model with both bond and stock market factors (Csb), the Bekaert & De Santis (2021) three-factor model (B3), the Elton et al. (1995) four-factor model (E4), the Fama & French (2015) five-factor model for bonds (F5), and an augmented F5 model with liquidity and momentum factors (F7).

For each pairwise comparison, we have 45 b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji}, we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. The table reports the results of two hypothesis tests: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of positive and negative differences equals 50%. We test the first hypothesis with a Wald test and the second with a Binomial test. We present a "winning model" if the sum of coefficient differences is significantly different from zero. "–" indicates that there is no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at

Table A.15: Model Horse Race – Full Sample (monthFE x MS stars) (continued)

A. Sharpe ratio

Winning model	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**
Losing model	**RR**	**Cb**	**Csb**	**B3**	**E4**	**F5**	**F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	18.94***	15.17***	10.09***	10.53***	7.087***	10.16***	10.14***
p-Value	0.000	0.000	0.000	0.000	0.001	0.000	0.000
% of coeff. Diff. >0	88.89***	95.56***	86.67***	86.67***	82.22***	86.67***	86.67***
Binomial p-Value	0.000	0.000	0.000	0.000	0.000	0.000	0.000

B. Raw return

Winning model	–	**Csb**	–	**E4**	–	–
Losing model	–	**RR**	–	**RR**	–	–
	RR_Cb	RR_Csb	RR_B3	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-6.210	-4.763*	-2.558	-5.632**	-1.537	-1.009
p-Value	0.102	0.059	0.269	0.018	0.486	0.640
% of coeff. Diff. >0	33.33**	26.67***	35.56**	13.33***	42.22	40.00
Binomial p-Value	0.018	0.001	0.036	0.000	0.186	0.116

C. CAPM bond

Winning model	–	–	**E4**	–	–
Losing model	–	–	**Cb**	–	–
	Cb_Csb	Cb_B3	Cb_E4	Cb_F5	Cb_F7
Sum of coeff. Diff.	-3.119	-0.581	-4.619*	0.387	1.220
p-Value	0.227	0.810	0.064	0.876	0.613
% of coeff. Diff. >0	37.78*	57.78	22.22***	55.56	60.00
Binomial p-Value	0.068	0.186	0.000	0.276	0.116

D. CAPM stock + bond

Winning model	–	–	–	**Csb**
Losing model	–	–	–	**F7**
	Csb_B3	Csb_E4	Csb_F5	Csb_F7
Sum of coeff. Diff.	2.348	-3.929	4.000	4.542*
p-Value	0.416	0.182	0.151	0.087
% of coeff. Diff. >0	64.44**	42.22	60.00	75.56***
Binomial p-Value	0.036	0.186	0.116	0.000

E. B3

Winning model	–	–	–
Losing model	–	–	–
	B3_E4	B3_F5	B3_F7
Sum of coeff. Diff.	-3.553	2.378	2.401
p-Value	0.212	0.383	0.350
% of coeff. Diff. >0	31.11***	57.78	62.22*
Binomial p-Value	0.008	0.186	0.068

F. E4

Winning model	**E4**	**E4**
Losing model	**F5**	**F7**
	E4_F5	E4_F7
Sum of coeff. Diff.	8.162***	8.023***
p-Value	0.005	0.002
% of coeff. Diff. >0	75.56***	84.44***
Binomial p-Value	0.000	0.000

G. F5

Winning model	–
Losing model	–
	F5_F7
Sum of coeff. Diff.	2.793
p-Value	0.565
% of coeff. Diff. >0	55.56
Binomial p-Value	0.276

Table A.16: Model Horse Race – Full Sample (monthFE x MS styles)

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using the full corporate bond fund sample. We estimate the relation between flow and the decile ranking of a fund share class based on different performance measures by running the regression:

$$F_{p,t} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{p,t-1} + e_{p,t}.$$

$F_{p,t}$ is the fund flow of mutual fund share class p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund share class p in month $t-1$ is in the decile i based on the first model and decile j based on the second model (excluding the dummy variable for $i = 5$ and $j = 5$). $X_{p,t-1}$ contains the following control variables (observed at the end of month $t-1$): lagged fund flow, lagged expense ratio, a dummy for no-load share classes, return standard deviation estimated over the prior twelve months, log of fund share class size, log fund age, as well as the month-times-Morningstar styles fixed effects.

Candidate performance measures are: the Sharpe ratio (SR), the raw return (RR), and the alphas of the single-factor model with bond market factor (Cb), the two-factor model with both bond and stock market factors (Csb), the Bekaert & De Santis (2021) three-factor model (B3), the Elton et al. (1995) four-factor model (E4), the Fama & French (2015) five-factor model for bonds (F5), and an augmented F5 model with liquidity and momentum factors (F7).

For each pairwise comparison, we have 45 b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji}, we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. The table reports the results of two hypothesis tests: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of positive and negative differences equals 50%. We test the first hypothesis with a Wald test and the second with a Binomial test. We present a "winning model" if the sum of coefficient differences is significantly different from zero. "–" indicates that there is no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table A.16: Model Horse Race – Full Sample (monthFE x MS styles) (continued)

A. Sharpe ratio

Winning model	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**
Losing model	**RR**	**Cb**	**Csb**	**B3**	**E4**	**F5**	**F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	19.68***	14.16***	10.70***	9.918***	6.934***	10.34***	10.35***
p-Value	0.000	0.000	0.000	0.000	0.001	0.000	0.000
% of coeff. Diff. >0	95.56***	95.56***	86.67***	84.44***	82.22***	91.11***	88.89***
Binomial p-Value	0.000	0.000	0.000	0.000	0.000	0.000	0.000

B. Raw return

Winning model	**Cb**	–	–	**E4**	–	–
Losing model	**RR**	–	–	**RR**	–	–
	RR_Cb	RR_Csb	RR_B3	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-7.484*	-3.539	-2.895	-5.476**	-1.337	-0.726
p-Value	0.054	0.212	0.253	0.029	0.586	0.758
% of coeff. Diff. >0	28.89***	31.11***	33.33**	15.56***	42.22	48.89
Binomial p-Value	0.003	0.008	0.018	0.000	0.186	0.500

C. CAPM bond

Winning model	–	–	–	–	–
Losing model	–	–	–	–	–
	Cb_Csb	Cb_B3	Cb_E4	Cb_F5	Cb_F7
Sum of coeff. Diff.	0.105	-0.473	-3.748	0.818	1.440
p-Value	0.972	0.857	0.135	0.758	0.577
% of coeff. Diff. >0	46.67	46.67	33.33**	57.78	62.22*
Binomial p-Value	0.383	0.383	0.018	0.186	0.068

D. CAPM stock + bond

Winning model	–	**E4**	–	–
Losing model	–	**Csb**	–	–
	Csb_B3	Csb_E4	Csb_F5	Csb_F7
Sum of coeff. Diff.	-0.093	-5.747*	2.607	3.258
p-Value	0.978	0.073	0.371	0.248
% of coeff. Diff. >0	48.89	28.89***	57.78	64.44**
Binomial p-Value	0.500	0.003	0.186	0.036

E. B3

Winning model	–	–	–
Losing model	–	–	–
	B3_E4	B3_F5	B3_F7
Sum of coeff. Diff.	-3.470	2.539	2.696
p-Value	0.268	0.364	0.308
% of coeff. Diff. >0	33.33**	62.22*	60.00
Binomial p-Value	0.018	0.068	0.116

F. E4

Winning model	**E4**	**E4**
Losing model	**F5**	**F7**
	E4_F5	E4_F7
Sum of coeff. Diff.	8.809***	8.690***
p-Value	0.004	0.002
% of coeff. Diff. >0	80.00***	80.00***
Binomial p-Value	0.000	0.000

G. F5

Winning model	–
Losing model	–
	F5_F7
Sum of coeff. Diff.	6.856
p-Value	0.259
% of coeff. Diff. >0	51.11
Binomial p-Value	0.5

Table A.17: Model Horse Race – Extended Corporate Bond Fund Sample

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using an extended corporate bond fund sample (no exclusion of observations with monthly TNA less than \$10 million). We estimate the relation between flow and the decile ranking of a fund share class based on different performance measures by running the regression:

$$F_{p,t} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{p,t-1} + \mu_t + e_{p,t}.$$

$F_{p,t}$ is the fund flow of mutual fund share class p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund share class p in month $t-1$ is in the decile i based on the first model and decile j based on the second model (excluding the dummy variable for $i = 5$ and $j = 5$). $X_{p,t-1}$ contains the following control variables (observed at the end of month $t-1$): lagged fund flow, lagged expense ratio, a dummy for no-load share classes, return standard deviation estimated over the prior twelve months, log of fund share class size, log fund age, as well as Morningstar rating dummies. We also include time fixed effects (μ_t).

Candidate performance measures are: the Sharpe ratio (SR), the raw return (RR), and the alphas of the single-factor model with bond market factor (Cb), the two-factor model with both bond and stock market factors (Csb), the Bekaert & De Santis (2021) three-factor model (B3), the Elton et al. (1995) four-factor model (E4), the Fama & French (2015) five-factor model for bonds (F5), and an augmented F5 model with liquidity and momentum factors (F7).

For each pairwise comparison, we have 45 b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji}, we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. The table reports the results of two hypothesis tests: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of positive and negative differences equals 50%. We test the first hypothesis with a Wald test and the second with a Binomial test. We present a "winning model" if the sum of coefficient differences is significantly different from zero. "–" indicates that there is no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table A.17: Model Horse Race – Extended Corporate Bond Fund Sample (continued)

A. Sharpe ratio

Winning model **Losing model**	**SR** **RR**	**SR** **Cb**	**SR** **Csb**	**SR** **B3**	**SR** **E4**	**SR** **F5**	**SR** **F7**
	SR_RR	SR_Cb	SR_Csb	SR_B3	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	19.84***	16.80***	12.99***	12.74***	10.84***	10.09***	10.89***
p-Value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
% of coeff. Diff. >0	88.89***	80.00***	84.44***	77.78***	71.11***	80.00***	84.44***
Binomial p-Value	0.000	0.000	0.000	0.000	0.003	0.000	0.000

B. Raw return

Winning model **Losing model**	– –	– –	– –	– –	– –	– –
	RR_Cb	RR_Csb	RR_B3	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-6.054	-4.009	-3.075	-4.400	-2.942	-1.222
p-Value	0.309	0.268	0.348	0.164	0.287	0.655
% of coeff. Diff. >0	31.11***	35.56**	40.00	35.56**	44.44	53.33
Binomial p-Value	0.008	0.036	0.116	0.036	0.276	0.383

C. CAPM bond

Winning model **Losing model**	– –	– –	– –	– –	– –
	Cb_Csb	Cb_B3	Cb_E4	Cb_F5	Cb_F7
Sum of coeff. Diff.	-2.034	-2.162	-3.607	-0.381	1.342
p-Value	0.614	0.527	0.249	0.907	0.675
% of coeff. Diff. >0	55.56	57.78	46.67	55.56	51.11
Binomial p-Value	0.276	0.186	0.383	0.276	0.500

D. CAPM stock + bond

Winning model **Losing model**	– –	– –	– –	– –
	Csb_B3	Csb_E4	Csb_F5	Csb_F7
Sum of coeff. Diff.	-1.804	-5.501	-1.818	0.586
p-Value	0.707	0.228	0.666	0.882
% of coeff. Diff. >0	35.56**	37.78*	48.89	51.11
Binomial p-Value	0.036	0.068	0.500	0.500

E. B3

Winning model **Losing model**	– –	– –	– –
	B3_E4	B3_F5	B3_F7
Sum of coeff. Diff.	-4.457	-1.401	0.760
p-Value	0.321	0.744	0.848
% of coeff. Diff. >0	37.78*	60.00	60.00
Binomial p-Value	0.068	0.116	0.116

F. E4

Winning model **Losing model**	– –	– –
	E4_F5	E4_F7
Sum of coeff. Diff.	0.682	4.447
p-Value	0.893	0.295
% of coeff. Diff. >0	60.00	66.67**
Binomial p-Value	0.116	0.018

G. F5

Winning model **Losing model**	– –
	F5_F7
Sum of coeff. Diff.	10.91
p-Value	0.136
% of coeff. Diff. >0	57.78
Binomial p-Value	0.186

Chapter 3

Which Factors for Corporate Bond Returns?*

3.1 Introduction

A pivotal issue in finance is understanding why certain types of assets, on average, earn vastly different returns than others do. Researchers and practitioners often attempt to explain these returns with factor models that consist of a sparse set of factors. In equity markets, hundreds of factors have been proposed, and equity managers have applied factor investing successfully for decades.[1] Factor investing in corporate bonds, on the other hand, is a relatively unexplored field. However, searching for bond factors has recently attracted growing interest, and, based on these discoveries, factor investing is likely to pick up substantially in the coming

*This chapter is based on the Article "Which Factors for Corporate Bond Returns?" authored by Thuy Duong Dang, Fabian Hollstein, and Marcel Prokopczuk, Review of Asset Pricing Studies, Volume 13(4), 615-652.

[1]Morgan Stanley reports that in 2017 $1.5 trillion was invested in smart beta, quant, and factor-based strategies and that assets under management have steadily grown at an average rate of 17% since 2010. By the end of 2018, exchange-traded funds (ETFs) had more than $900 billion in assets under management, and the top-two managers, Vanguard and BlackRock, each held more than $300 billion in assets in factor products. See at `https://www.robeco.com/hk/en/essentials/factor-investing/`.

years.

A plethora of factors lead to the necessity from both an academic and a practitioner's perspective to know which are genuine risk factors in corporate bond markets that provide incremental information about returns. In this chapter, we thus address the following questions: Do we really need all factors proposed in the corporate bond literature to explain the cross-section of returns? Which factors move corporate bond prices systematically? What set of factors overall best describes corporate bond returns? Are some factors redundant relative to others? To what extent does each needed factor play a role in explaining time-series and cross-sectional variation in corporate bond returns? Which economic forces drive the factors?

Our main contribution is a systematic analysis of the factors proposed in the corporate bond pricing literature. Our study helps academics and practitioners separate useful factors from redundant ones and search the growing list of bond factors for a set that spans the tangency portfolio and collectively best explains the differences in corporate bond returns. Based on this, we can build an "optimal" corporate bond factor model. To the best of our knowledge, we are the first to comprehensively compare a broad set of common and recently proposed factors and factor models for corporate bonds.

We start our empirical analysis by considering a collection of, from our point of view, the 23 most prominent risk factors in the corporate bond literature. We use the bond market (MKTb), term (TERM), and default risk (DEF) (Fama & French 1993), credit risk (CRF), downside risk (DRF), liquidity risk (LRF), and short-term reversal (STR) (Bai et al. 2019), momentum (MOMb) and long-term reversal (LTR) (Jostova et al. 2013; Bali, Subrahmanyam, & Wen 2017b, 2021a), bond volatility (BVL), carry (CRY), duration (DUR), stock momentum (MOMs), and value (VAL) (Israel, Palhares, & Richardson 2018; Kelly, Palhares, & Pruitt 2023), economic uncertainty (UNC) and (tax) policy uncertainty (EPU, EPUtax) (Bali, Brown, & Tang 2017a; Bali, Subrahmanyam, & Wen 2021b; Tao, Wang, Wang, & Wu 2022; Lee 2022), and volatility risk (VOL) (Chung et al. 2019), along with the five Fama & French (2015) stock market factors (MKT, SMB, HML, RMW, CMA) (Bektić, Wenzler, Wegener, Schiereck, & Spielmann 2019).

Given the apparent importance and need for replicability of factor premiums, as highlighted by a growing number of meta-studies, for instance, Welch & Goyal (2008), Harvey et al. (2016), McLean & Pontiff (2016), Green, Hand, & Zhang (2017), Linnainmaa & Roberts (2018), and Hou, Xue, & Zhang (2020), we examine these published factors on the same pedestal using the same period, data sources, and bond return definitions. While the choice of alternative specifications and procedures is not technically wrong, using factors that are constructed consistently helps us to avoid comparing apples with oranges.

In the first part of our empirical study, we use the necessary condition of the factor identification protocol popularized by Pukthuanthong et al. (2019). With this step, we basically separate factor candidates that systematically move corporate bond prices from those that do not. A factor candidate cannot be a viable risk factor if it does not move prices. Technically, we analyze whether the factor candidates can explain the canonical correlations between the entire set of factors and test asset principal components. We only retain those factors that pass the identification protocol for further analysis. We find that many prominent factors already fail this first test. For example, all the factors proposed by Bai et al. (2019) do not satisfy the necessary condition for being a risk factor in corporate bond markets. In addition, the Tao et al. (2022) and Lee (2022) policy uncertainty factors, the Israel et al. (2018) value factor, the short-term reversal factor, and all Fama & French (2015) stock market factors are eliminated.

As a second step, we employ the Bayesian marginal likelihood model comparison method recently developed by Barillas & Shanken (2018) and Chib et al. (2020) (BS-CZZ). The key advantages of this approach are that (a) it enables us to simultaneously compute the model probabilities for the collection of all possible models that are subsets of the given factors, while (b) it takes into account the matter of parsimony. The first main result is that a four-factor union of carry, duration, stock momentum, and term structure factors is revealed by the data as the best (no. 1) corporate bond risk factor model in terms of the Bayesian posterior probability. Based on a Bayes factor cutoff, only three further models remain as serious contenders. All four winning models contain the carry and stock momentum factors. The duration and term-structure factors have cumulative posterior probabilities around 50%.

To provide direct statistical evidence on the relative performance of different models, we use the Barillas, Kan, Robotti, & Shanken (2020) test of the equality of squared Sharpe ratios. We conduct pairwise comparisons among the winning models and various existing models. We find that the no. 1 winning factor model yields a substantially and significantly larger squared Sharpe ratio than all other contenders. Thus, the selected set of factors clearly dominates the existing models. We show that this is not only true in-sample. Also, out-of-sample, with two different sample splitting schemes, the winning models generate the highest Sharpe ratios.

We continue our analysis by running two sets of spanning tests. The purpose of the first set is to demonstrate why the various existing models fall short of explaining the winning factors. We find that the existing models largely fail to explain the average returns of the carry and stock momentum factors. However, even the Israel et al. (2018) and Kelly et al. (2023) models that contain one or both of these factors are rejected by the Gibbons, Ross, & Shanken (1989) (GRS) test. On the other hand, the no. 1 winning model can explain all other factors that pass the first-step identification protocol.

In a penultimate step, we use time-series and cross-sectional asset pricing tests with test assets to thoroughly analyze the performance of all models. We find that the winning models perform reasonably well for these tasks. In the time-series tests, the four winning models, along with those of Israel et al. (2018) and Kelly et al. (2023), which share many of the same factors, typically yield the smallest GRS test statistics, the lowest average absolute alphas, and the smallest squared Sharpe ratios of the alphas for different sets of test assets. In cross-sectional tests, the same set of models performs best. We find that the no. 1 winning model yields the largest cross-sectional R^2s.

Finally, we examine the economic drivers of the corporate bond risk factors contained in the winning models. We find that corporate bond illiquidity along with volatility are important drivers of the carry and duration factors. The duration factor, however, is also strongly driven by intermediary distress and inflation. The main determinant of the stock momentum factor is inflation, while the term factor is mainly driven by the change in industrial production.

The findings in this chapter have important practical implications. The winning set of factors can be used as a benchmark model for future research and in performance evaluation. Furthermore, investors in corporate bond markets can build on our findings to implement the most promising factor-investing strategies.

The remainder of this paper is organized as follows. In Section 3.2, we briefly review the literature. Section 3.3 describes our data and methodology. In Section 3.4, we present the factor and model selection results. We perform asset pricing tests to compare the winning models to existing ones in Section 3.5. In Section 3.6, we analyze the economic drivers of the set of winning factors. Section 3.7 provides concluding remarks.

3.2 Literature Review

Both stocks and corporate bonds are contingent claims on the value of the same underlying firm. However, several notable features distinguish bond from stock markets, suggesting potential market segmentation. Indeed, Chordia et al. (2017) and Choi & Kim (2018) find a discrepancy in risk premiums between corporate bond and equity markets. Therefore, it is important to investigate the cross-section of corporate bond returns by also using the factors constructed based on corporate bond characteristics, rather than only relying on the available commonly used factors from the equity market. This direction also helps to facilitate factor-based investing strategies in corporate bond markets. Hence, we focus our study mainly on corporate bond factors.

Earlier studies, notably Fama & French (1993), generally utilize aggregate bond indexes, such as the term and default spread factors, to explain the cross-sectional variation in corporate bond returns. Recently, inspired by the way characteristics have been used for constructing equity factors, substantial research efforts have been devoted to exploring new factors that drive corporate bond returns. Bali et al. (2017b) and Bali et al. (2021a) examine whether short-term reversal, momentum, and long-term reversal are priced in the corporate bond market. Bali et al. (2017b) introduce a return-based factor model including three factors constructed based on the bond market factor and these past return characteristics. Bai et al. (2019) propose a four-factor model, including the bond market as well as factors that build on the downside risk, credit risk, and liquidity risk characteristics, which appear to be prevalent in the corporate bond market. Israel et al. (2018) and Kelly et al. (2023) propose alternative factors and factor models based on the bond volatility, carry, duration, stock momentum, and value characteristics. Bektić et al. (2019) study the Fama & French (2015) five-factor model in corporate bond markets. Chung et al. (2019), Bali et al. (2021b), Tao et al. (2022), and Lee (2022) find that aggregate volatility and economic and policy uncertainty are priced in the cross-section of corporate bond returns. In this study, we comprehensively examine the properties of the factors introduced in these studies (in total 23) and form an optimal factor model.

Two competing approaches demonstrate the ability of factor models in explaining the cross-section of returns: left-hand-side (LHS) and right-hand-side (RHS) approaches, as classified by Fama & French (2018). LHS approaches introduce additional test assets and examine the models based on their abilities to price these test assets. For these, it often comes down to alphas, which are the estimated intercepts from time-series regressions of base asset returns on these factor models. Alphas capture the difference between the return an asset actually earns and what a factor model would predict, and hence gauges the model's error. A model with lower average pricing errors is deemed to perform better. Empirical implementations using characteristic/industry-sorted portfolios as the

LHS test assets and assessing model performance by alpha-based statistics from time-series regressions to explain the LHS assets are ubiquitous. In numerous studies in equity markets, such as Fama & French (2015, 2018), Hou et al. (2015), or Stambaugh & Yuan (2017), competing factor models are evaluated using several comparative alpha-based criteria, for example, the number of significant alphas based on t-statistics, the number of rejections by the GRS test, the point estimates of average absolute alpha, and the average absolute t-statistic. Among others, Bali et al. (2017b) and Bai et al. (2019) also aim to identify a superior model for corporate bond returns as the one that generates a smaller average absolute alpha and delivers a larger average time-series regression R^2 for certain test assets.

While the LHS alpha-based setting appears frequently in the empirical literature, some criticize it as being problematic. First, the selection of test assets is not innocuous. One important critique about standard asset pricing tests, brought forward by Lewellen, Nagel, & Shanken (2010), is that characteristic-sorted portfolios used as test assets do not have sufficient independent variation in the loadings of factors constructed with the same characteristics. Second, this framework ignores the pricing impact of factors from other models. Third, Barillas & Shanken (2017), Fama & French (2018), and Ahmed, Bu, & Tsvetanov (2019) show in numerous examples that informally comparing point estimates of alpha-based performance metrics may yield inconsistent model rankings, which can lead to incorrect judgments on the pricing performance of different models.

Barillas & Shanken (2017) straightforwardly emphasize that models should be judged in terms of their power to explain not only test assets but also the traded factors in other models (ideally, the entire universe of returns). They argue that LHS test portfolios do not provide any further information about model comparisons beyond what can be obtained by examining how well each model prices the factors of other models. Thus, following their argument, test assets are irrelevant for the purpose of model comparison.

This revealing insight leads to the development of the so-called “RHS approach.” Here, spanning regressions only involve other factors regressed on those of a model in order to decide whether candidate factors add explanatory power to a benchmark model. If the intercept is zero, the candidate factors contribute no additional information. Spanning tests are the main method adopted by Hou, Mo, Xue, & Zhang (2019) and Daniel, Hirshleifer, & Sun (2020) to compare factor models.

Barillas & Shanken (2018) develop a Bayesian RHS setting that permits us to compare a large set of models simultaneously and identify the best, parsimonious one. There are also alternative recent (LHS) approaches for model selection (e.g., Feng, Giglio, & Xiu 2020; Hwang & Rubesam 2020; Harvey & Liu 2021). We opt for the BS-CZZ approach because it is economically motivated for exactly the task we intend it for: to find an optimal factor model.[2] In addition, the approach turns out to be reasonably powerful, and the results of this RHS approach hold up well under an alternative LHS evaluation.[3]

Taking all the issues discussed above into consideration without being dogmatic on the LHS/RHS debate, the first part of this chapter is based on an RHS approach, in which we scan for the best corporate bond factor model using a Bayesian marginal-likelihood-based method. Afterward, we analyze the

[2]The approach of Feng et al. (2020) was mainly designed to accommodate very high-dimensional factor selection problems. This is much more relevant for equity markets than for corporate bond markets. After our first-step screening of whether the factors move corporate bond prices, only 11 factors remain. These can be well handled by standard statistical tools. Furthermore, the main goal of the Feng et al. (2020) approach is the evaluation of new factors rather than the selection of an optimal factor model.

[3]The Harvey & Liu (2021) approach is a very careful statistical approach to detect helpful factors in the presence of data mining and multiple testing. As such, it is very useful to gauge the significance of any new factor given the previous factors and possibly many others that have been tried. The multiple-testing adjustment, however, makes the approach quite conservative. For cross-sectional asset pricing in equity markets, the Harvey & Liu (2021) conclusion that the market factor is dominant is controversial. Indeed, when applying the Harvey & Liu (2021) approach to our corporate bond data, we obtain a similar result, that is, that only the corporate bond market factor is chosen. We back up the usefulness of the selected models by comparing them to existing models using state-of-the-art time-series and cross-sectional asset pricing tests. These clearly show that the corporate bond capital asset pricing model (CAPM) is inferior to the models selected by our application of the Barillas & Shanken (2018) approach for pricing the cross-section of corporate bonds. We address the (justified) criticism that the Sharpe-ratio-based methods, to which the Barillas & Shanken (2018) approach belongs, may choose well-performing factors that do not move prices by adding a first preselection step based on the factor identification protocol of Pukthuanthong et al. (2019).

winning models further by comparing their performance to existing models using two RHS approaches. Finally, we use multiple sets of LHS test portfolios to analyze the performance of the winning models and the existing ones for explaining cross-sectional and time-series variation in corporate bond returns.

3.3 Data and Methodology

3.3.1 Corporate bond data

We use the corporate bond data set of Kelly & Pruitt (2022). It is compiled from four sources: the Trade Reporting and Compliance Engine (TRACE) Enhanced, the Mergent Fixed Income Securities Database (FISD), Compustat, and the Center for Research in Security Prices (CRSP). Corporate bond transaction data (intraday clean price and volume) are from TRACE Enhanced and bond characteristics, such as bond ratings or coupons, are from FISD. Additional equity characteristics are from Compustat and CRSP. Special types of bonds, such as convertible bonds, bonds with floating coupon rates, and callable bonds are excluded from the data set.

The monthly return of corporate bond i in month t is calculated as

$$r_{i,t} = \frac{P_{i,t} + AI_{i,t} + C_{i,t}}{P_{i,t-1} + AI_{i,t-1}} - 1, \tag{3.1}$$

where $P_{i,t}$ is the clean transaction price, $AI_{i,t}$ is the accrued interest, and $C_{i,t}$ is the coupon payment, if any, of bond i in month t.[4]

3.3.2 Candidate factors and models

In Table 3.1, we briefly summarize the definitions of the bond variables (panel A) as well as the 23 candidate factors used in this study (panel B). More details on the exact construction of all variables and factors can be found in Appendix B.1.[5]

[4] To limit the effect of extreme outliers, Kelly & Pruitt (2022) winsorize the return data at the 0.05% and 99.95% quantiles.

[5] We only consider factors defined by one characteristic. Combinations of different characteristics suffer from overfitting bias (Novy-Marx 2016).

Table 3.1: Factor Definitions

This table provides brief descriptions of the main measures and factors used in this chapter. More details can be found in Appendix B.1.

Panel A		Measures	Definitions
1	β_{UNC}	uncertainty beta	the bond's sensitivity to the economic uncertainty, economic policy uncertainty, or tax policy uncertainty index
2	β_{VIX}	volatility beta	the bond's sensitivity to the change in the VIX index
3	cr	credit rating	a bond's rating as the average of the ratings provided by S&P and Moody's when both are available
4	cry	carry	option-adjusted spread (fixed difference of bond discount rate to Treasury curve)
5	dr	downside risk	5% VaR: the second lowest monthly return observation over the past 36 months, multiplied by -1
6	dur	duration	derivative of the bond value w.r.t. the credit spread, divided by current price
7	illiq	bond illiquidity	the negative of the bond return autocovariance
8	ltr	long-term reversal	the bond's past 36-month cumulative return from month $t-48$ to month $t-13$
9	momb	bond momentum	the bond's past 6-month cumulative return from month $t-7$ to month $t-2$
10	moms	stock momentum	the company's past 6-month cumulative stock return from month $t-7$ to month $t-2$
11	spr_d2d	spread to D2D	option-adjusted spread divided by distance to default
12	str	short-term reversal	the bond's previous $(t-1)$ month return
13	vol	bond volatility	the bond's volatility over the past 24 months
Panel B		Bond Factors	Definitions
1	MKTb	bond market factor	the value-weighted (by amount outstanding) average excess return of all corporate bonds in the sample
2	BVL	bond volatility factor	the average return difference between the highest- and lowest-bond-volatility portfolios across rating
3	CRF	credit risk factor	the average of the average return difference between the highest- and lowest-credit-risk portfolios across downside risk, illiquidity, and short-term reversal
4	CRY	carry factor	the average return difference between the highest- and lowest-carry portfolios across rating
5	DEF	default risk factor	the difference between the return on a market portfolio of long-term corporate and government bonds
6	DRF	downside risk factor	the average return difference between the highest- and lowest-VaR portfolios across rating
7	DUR	duration factor	the average return difference between the highest- and lowest-duration portfolios across rating
8	EPU, EPUtax, & UNC	uncertainty risk factors	the average return difference between the highest- and lowest-β_{UNC} portfolios across rating
9	LRF	liquidity risk factor	the average return difference between the highest- and lowest-illiquidity portfolios across rating
10	LTR	long-term reversal factor	the average return difference between the long-term loser and winner portfolios across rating
11	MOMb	bond momentum factor	the average return difference between the bond winner and loser portfolios across rating
12	MOMs	stock momentum factor	the average return difference between the stock winner and loser portfolios across rating
13	STR	short-term reversal factor	the average return difference between the short-term loser and winner portfolios across rating
14	TERM	term risk factor	the difference between the return on a market portfolio of long-term government bonds and the one-month Treasury bill
15	VAL	value factor	the average return difference between the highest- and lowest-spread-to-D2D portfolios across rating
16	VOL	volatility risk factor	the average return difference between the highest- and lowest-β_{VIX} portfolios across rating

In addition to the individual factors, for comparison, we also consider a set of existing factor models. We indicate the set of factors included in the models in braces. That is, we consider a corporate bond CAPM (CAPMbond: {MKTb}), the Fama & French (1993) three-factor model for corporate bonds (FF3: {MKTb, TERM, DEF}), the Fama & French (1993) three-factor model for corporate bonds augmented by a liquidity risk factor and a bond momentum factor (aug. FF3: {MKTb, TERM, DEF, LRF, MOMb}), the Fama & French (1993) five-factor model for corporate bonds (FF5stkb: {MKTs, SMB, HML, TERM, DEF}), the Bai et al. (2019) four-factor model (BBW: {MKTb, DRF, CRF, LRF}), the four-factor model in the spirit of Bali et al. (2017b) (BSW: {MKTb, STR, MOMb, LTR}), the five-factor model in the spirit of Israel et al. (2018) (IPR: {CRY, DUR, MOMb, MOMs, VAL}), and the observable five-factor model of Kelly et al. (2023) (KPP: {MKTb, CRY, DUR, BVL, VAL}).

3.3.3 First step: Factor identification protocol

For a first-step screening, we use the necessary condition of the factor identification protocol of Pukthuanthong et al. (2019). The goal of this step is to identify factors that systematically move corporate bond prices. That is, the factors should be related to the covariance matrix of corporate bond returns.

As representative test assets for this step, we use a set of 12 industry portfolios, 25 size-maturity portfolios, 25 rating-maturity portfolios, and further 5×5 double-sorted portfolios on the bond rating and 29 corporate bond characteristics provided by Kelly & Pruitt (2022). This set of portfolios clearly satisfies the requirements of Pukthuanthong et al. (2019) that the test assets should belong to different industries and have sufficient heterogeneity. From these portfolios, we extract the first 10 principal components using the method of Connor & Korajczyk (1988).[6] To account for possible nonstationarity, we cut our sample into two halves and do the analysis separately for each subperiod

[6] To be precise, we obtain the matrix $\Omega = (1/T)RR'$, where T is the number of time-series observations, and R is the $T \times N$ matrix of the N de-meaned test asset returns. The extracted principal components are the first 10 eigenvectors of Ω.

(Pukthuanthong et al. 2019). Next, we calculate the canonical correlations between the candidate factors and these 10 principal components. Finally, we regress each of the 10 canonical variates (which are all weighted averages of the 10 principal components) on a constant and the set of factor candidates. As in Pukthuanthong et al. (2019), for an eligible factor we require an average of the absolute t-statistics associated with the significant canonical correlations exceeding 1.96 and the average number of single absolute t-statistics exceeding 1.96 has to be higher than 2.5.

All factors that do not pass this first test apparently do not move bond prices and can be rejected as viable risk factors. Hence, we will only consider candidate factors that pass this factor identification protocol for the next steps.

3.3.4 Second step: BS-CZZ model comparison procedure

Among the factors that move corporate bond prices, we next aim to find those that best price the cross-section. We employ the Bayesian marginal-likelihood-based model comparison approach introduced by Barillas & Shanken (2018) and revisited by Chib et al. (2020).[7] This method allows us to simultaneously compare all possible models based on the subsets of the given factor space. To scan for the best model, we compute their log marginal likelihoods to perform the prior-posterior update and then rank them based on their posterior probabilities.

In more detail, starting with a set of K (traded) potential risk factors, in general $J = 2^K - 1$ factor combinations are possible. With the factor set resulting from the first-step factor identification and the restrictions on correlated factors described in Section 3.4.2, 1,024 candidate factor models remain. The model space is thus $\mathcal{M} = \{\mathbb{M}_1, \mathbb{M}_2, ..., \mathbb{M}_J\}$. $\mathbb{M}_j$ is one possible model defined by the vector of included factors $\tilde{f}_j$ and that of excluded factors f_j^*.

[7] Chib et al. (2020) show that the original prior definition by Barillas & Shanken (2018) is unsound for model comparisons. They propose an alternative approach with a modified prior. We follow exactly the general approach suggested by Chib et al. (2020). The authors show their approach performs the best in their simulations.

Each of the 1,024 factor models thus has a $L_j \times 1$ vector of included factors $\tilde{f}_j$ and a $(K - L_j) \times 1$ vector of excluded factors f_j^*. The data generating process of model j is thus given by

$$\tilde{f}_{j,t} = \tilde{\alpha}_j + \tilde{\epsilon}_{j,t}, \tag{3.2}$$

and

$$f_{j,t}^* = B_{j,f}^* \tilde{f}_{j,t} + \epsilon_{j,t}^*. \tag{3.3}$$

$\tilde{\alpha}_j$ is a $L_j \times 1$ parameter vector and $\tilde{\epsilon}_{j,t}$ is a multivariately normally distributed residual vector. $B_{j,f}^*$ is a $(K - L_j) \times L_j$ parameter matrix. $\epsilon_{j,t}^*$ is also a multivariately normally distributed residual vector. A special case applies when all factors are included in f_j.[8]

The log marginal likelihood of a model $\mathbb{M}_j$ $(j \neq J)$ with y given the sample data of the factors over T time periods in closed form is

$$\log \tilde{m}(y|\mathbb{M}_j) = \log \tilde{m}(\tilde{f}|\mathbb{M}_j) + \log \tilde{m}(f^*|\mathbb{M}_j). \tag{3.4}$$

We provide the details on the computation of the terms in Equation (3.4) in Appendix B.2.

The end product of the scanning procedure is a ranking of models

$$\{\mathbb{M}_1*, \mathbb{M}_2*, ..., \mathbb{M}_J*\} \tag{3.5}$$

by

$$\tilde{m}(y|\mathbb{M}_1*) > \tilde{m}(y|\mathbb{M}_2*) > ... > \tilde{m}(y|\mathbb{M}_J*), \tag{3.6}$$

where $\mathbb{M}_1*$ denotes the winning model, identified as the one that has the highest posterior model probability. Since the remaining terms in the posterior-probability

[8] $\tilde{\epsilon}_{j,t}$ and $\epsilon_{j,t}^*$ are also assumed to be not serially correlated. To thoroughly examine the potential issue of autocorrelations in the factor returns, we perform Ljung & Box (1978) tests of general dependency in the time series. For an overwhelming majority of factors, we cannot reject the null hypothesis of no significant dependency in the factor returns. Among the factors surviving the first-step screening, 10 of 11 factors have no significant dependency in their time series (only the VOL factor, which is not included in any of the top models, does). The White (1980) and Newey & West (1987) standard errors for the factor returns are also very similar (the average difference is 0.01 percentage points). The normality assumption is also not crucial. Chib & Zeng (2020) provide an alternative approach assuming student t-distributions of the factors. This approach yields a similar result with the same winning model.

calculation can be summarized by just a normalization constant, the ranking of posterior probabilities is equivalent to that of marginal livelihoods $\tilde{m}(y|\mathbb{M}_j)$.

3.3.5 Model comparison based on squared Sharpe ratios

After having determined the top model(s), the next step is a comparison to existing ones. For this purpose, among others, we use the Sharpe-ratio-based approach of Barillas et al. (2020) that requires a series of tests. First, we compute the differences between the bias-adjusted sample squared Sharpe ratios for various pairs of factor models.[9] Second, we calculate the p-values for the test of equality of the squared Sharpe ratios in two cases of nested models and non-nested models.

In the case of nested models (i.e., all of the factors in one model are included in the other model), in order to determine whether the model with more factors is superior, we check whether the squared Sharpe ratio of the larger model is higher than that of the model with fewer factors. This is a test of whether alphas of the noncommon factors in the larger model (i.e., that are not contained in the smaller model) regressed on the smaller one are significantly different from zero, which can be done simply with the GRS test.

In the case of non-nested models (i.e., each model contains factors not included in the other model), the statistical analysis is a sequential test. The preliminary step entails comparing the squared Sharpe ratios of the model composed of all the factors from both models and the smaller one that contains only the common factors. It becomes equivalent to testing the null hypothesis that the alphas of the nonoverlapping factors on the common ones are zero. If this test fails to reject, then the evidence is consistent with the notion that the common factors model is as good as the models that add the noncommon factors. If this test is rejected, some or all of the noncommon factors are not redundant and

[9] The squared Sharpe ratio for each model is modified to be unbiased for small samples under joint normality by multiplying it by $(T - K - 2)/T$ and subtracting K/T, where T is the number of return observations and K is the number of factors.

contribute to an increase in the squared Sharpe ratio compared to the common factors model. However, we still do not know which non-nested model has a higher squared Sharpe ratio. Therefore, we then proceed with a direct test of the equality of the squared Sharpe ratios from the two non-nested models by calculating the p-value based on the results in Proposition 1 of Barillas et al. (2020).[10]

3.4 Model Selection

3.4.1 Summary statistics

Our sample includes 8,759 U.S. corporate bonds issued by 1,220 unique firms with 443,485 bond-month return observations in total during the sample period from July 2002 to December 2019. Over the sample period, on average, 83.85% of our rated bond sample are investment grade and 16.15% are noninvestment grade.[11] On average, our sample includes approximately 5,361 bonds per month over the whole period.

Panel A of Table 3.2 reports the descriptive statistics of our bond sample. The average monthly bond return is 0.50%, with a standard deviation of 2.17%. The sample contains bonds with an average rating of 8.02 (i.e., BBB+), and an average amount outstanding of $913 million. The average corporate bond in our sample is 3.44 years old, has a time-to-maturity of 8.57 years, and a duration of 5.92 years. The average corporate bond return, its distribution, as well as the other summary statistics are very similar to those reported in other studies (e.g.,

[10] The p-value in this direct test is computed as the bias-adjusted squared Sharpe ratio difference divided by its standard error. The standard error of the squared Sharpe ratio difference is the square root of the asymptotic variance divided by the number of monthly observations. The asymptotic variance can be calculated as $d_t = 2(u_{A,t} - u_{B,t}) - (u_{A,t}^2 - u_{B,t}^2) + (\theta_A^2 - \theta_B^2)$, with $u_{A,t} = \mu_A' V_A^{-1}(f_{A,t} - \mu_A)$ and $u_{B,t} = \mu_B' V_B^{-1}(f_{B,t} - \mu_B)$. θ_A^2 and θ_B^2 are the bias-adjusted squared Sharpe ratios of models A and B, respectively. μ_A is a vector of the average returns of the factors in model A, V_A is the corresponding covariance matrix, and $f_{A,t}$ is the vector of factor returns at time t.

[11] The ratings are coded as numbers between 1 and 21. Higher numerical scores imply higher credit risk. Numerical ratings of 10 or below (i.e., BBB- or better) are labeled as investment grade and ratings of 11 or higher are considered as high yield.

Jostova et al. 2013; Bai et al. 2019; Bali et al. 2021a; Kelly et al. 2023).

Panel B of Table 3.2 presents the summary statistics of the monthly factor returns between August 2003 and December 2019. Since a certain amount of data is first necessary to obtain the measures, the time series of DRF, CRF, EPU, EPUtax, LTR, STR, UNC, and VOL returns start somewhat later. LTR is the last factor with data available and starts in August 2006. For all tests, including LTR, we use the common sample period from August 2006. To place all factors on equal footing, we follow the approach of Bai et al. (2019), who are inspired by the classical Fama & French (1993) approach to equity factors, and obtain the factors via double-sorts with credit ratings. This way, we ensure that the factors genuinely pick up the risk and return related to their underlying economic variables rather than just passive exposure to credit risk.

The average bond market excess return for our sample is 0.34% per month and highly statistically significant with a t-statistic of 3.35. The average return is very similar to the 0.39% per month reported by Bai et al. (2019) for a slightly shorter sample period. The corporate bond factors mostly yield significantly positive average returns that are consistent with the previous literature.[12]

[12] In a bit more detail for some of the historically most important factors: The TERM factor yields a mean return of 0.46% per month with a t-statistic of 2.18. The DEF factor, on the other hand, only has an insignificant average monthly return of 0.06%. The very small and insignificant return for the DEF factor is consistent with the 0.02% per month reported by Fama & French (1993) and the 0.04% per month reported by Gebhardt et al. (2005a). On the other hand, the TERM factor return is substantially larger for our sample period than that reported by Fama & French (1993) (0.06% per month). As in Bai et al. (2019), the credit risk, downside risk, liquidity risk, and short-term reversal factors yield large monthly average returns, which are all highly statistically significant. The downside risk factor has an average return of 0.66% per month. The credit risk factor yields 0.36%, the liquidity risk factor 0.43%, and the short-term reversal factor has a monthly average return of 0.39%. The only notable exception where our results differ is the bond momentum factor, which yields a significant negative return as opposed to a positive single-sorted excess return reported by Jostova et al. (2013) for the period 1973–2011. These results are consistent with the findings of Israel et al. (2018), who show that the lion's share of the positive combined bond and equity momentum profits accumulates in the pre-TRACE sample period. Thus, in their combined figure the positive equity momentum and the negative bond momentum approximately cancel out from 2002 on. Furthermore, Galvani & Li (2020) find that momentum returns in corporate bond markets crucially depend on outlier observations.

Table 3.2: Summary Statistics

Our sample contains 8,759 U.S. corporate bonds issued by 1,220 unique firms over the period from July 2002 to December 2019. Panel A reports the descriptive statistics including the mean, median, standard deviation, skewness, kurtosis, and percentiles of bond–month observations of returns (in %) and bond characteristics including the credit rating, the size (amount outstanding in $ million), the age (in years), the time-to-maturity (in years), the duration (in years), and the bond spread (in %). Ratings are numerical scores, where 1 refers to an AAA and 21 refers to a C rating. Panel B presents the summary statistics of the time series of the 18 corporate bond and 5 equity candidate factors. The t-statistics (in parentheses) are based on Newey & West (1987) standard errors with 4 lags. For each factor, we also report the exact sample period from the first to the last month the data are available. The definitions of the factors can be found in Appendix B.1.

Panel A	Mean	Median	SD	Skew	Kurt	Percentiles			
						$10th$	$25th$	$75th$	$90th$
Return (%)	0.50	0.32	2.17	0.45	18.6	−1.31	−0.23	1.20	2.56
Rating	8.02	8.00	2.97	0.31	3.22	5.00	6.00	10.0	12.0
Size ($ million)	913	700	717	3.39	28.2	350	500	1,000	1,750
Age (years)	3.44	2.67	2.96	1.72	7.88	0.50	1.25	4.75	7.42
Time to maturity (years)	8.57	5.76	8.45	1.75	6.87	1.50	3.12	9.17	25.5
Duration (years)	5.92	4.88	4.32	1.09	3.44	1.39	2.82	7.36	13.5
Spread (%)	1.78	1.31	1.50	1.98	8.31	0.47	0.77	2.28	3.79

Panel B	Mean	(t-statistic)	Median	SD	Skew	Kurt	First	Last
Bond Factors								
MKTb	0.34***	(3.35)	0.41	1.31	0.16	10.4	Aug-2003	Dec-2019
BVL	0.53***	(3.25)	0.64	2.16	0.38	8.28	Aug-2003	Dec-2019
CRF	0.36**	(2.02)	0.25	1.81	−0.32	8.55	Jul-2004	Dec-2019
CRY	0.95***	(5.60)	1.00	2.02	0.91	7.33	Aug-2003	Dec-2019
DEF	0.06	(0.45)	0.05	1.99	−0.49	7.95	Aug-2003	Dec-2019
DRF	0.66***	(3.16)	0.59	2.23	0.90	8.79	Jul-2004	Dec-2019
DUR	0.52***	(2.68)	0.64	2.52	0.01	7.95	Aug-2003	Dec-2019
EPU	0.11	(1.38)	0.14	0.83	−1.11	8.35	Aug-2005	Dec-2019
EPUtax	0.03	(0.47)	0.05	0.63	−1.63	11.4	Aug-2005	Dec-2019
LRF	0.43***	(3.10)	0.28	1.29	3.95	32.9	Aug-2003	Dec-2019
LTR	0.07	(0.46)	−0.09	1.73	1.80	12.6	Aug-2006	Dec-2019
MOMb	−0.38***	(−3.07)	−0.25	1.52	−3.01	23.0	Aug-2003	Dec-2019
MOMs	0.22***	(4.46)	0.20	0.78	−0.05	9.36	Aug-2003	Dec-2019
STR	0.39***	(3.60)	0.43	1.27	0.37	6.53	Sep-2003	Dec-2019
TERM	0.46**	(2.18)	0.39	3.16	0.37	5.29	Aug-2003	Dec-2019
UNC	0.00	(0.03)	0.07	1.18	−1.46	12.1	Aug-2005	Dec-2019
VAL	0.75***	(6.84)	0.81	1.35	−0.33	5.45	Aug-2003	Dec-2019
VOL	0.12*	(1.94)	0.09	0.65	2.25	18.1	Aug-2003	Dec-2019
Stock Factors								
MKTs	0.77**	(2.47)	1.29	4.00	−0.77	5.02	Aug-2003	Dec-2019
SMB	0.09	(0.57)	0.16	2.37	0.28	2.90	Aug-2003	Dec-2019
HML	−0.05	(−0.24)	−0.17	2.50	−0.03	5.27	Aug-2003	Dec-2019
RMW	0.26**	(2.16)	0.27	1.58	0.18	3.45	Aug-2003	Dec-2019
CMA	0.01	(0.05)	−0.02	1.43	0.32	2.75	Aug-2003	Dec-2019

3.4.2 Factor identification results

We begin the empirical analysis with the factor identification protocol step of Pukthuanthong et al. (2019). In panel A of Table 3.3, we show the results for the canonical correlations between the 10 principal components extracted from the large set of test assets and the 23 candidate factors. We do this analysis for the two equal halves of our sample period. We find that in both halves 9 of the 10 canonical correlations are statistically significant. Thus, several pairs of canonical variates between the test assets and factors, with each being orthogonal to the others, have substantial intercorrelations.

The main output of the factor identification protocol step is in panel B of Table 3.3. The bond market factor reaches an average t-statistic in the multiple regressions to explain the significant canonical variates of 4.9. In both subperiods, 7 of the 10 t-statistics for the bond market are statistically significant. Thus, MKTb clearly passes the hurdles set by Pukthuanthong et al. (2019) for the factor-identification-protocol step. Other prominent corporate bond factors, however, fail this step. For example, the Bai et al. (2019) CRF, DRF, and LRF factors are eliminated. They are not sufficiently strongly related to the significant canonical variates of the test asset principal components. Hence, they do not appear to sufficiently strongly and systematically move corporate bond prices. Similarly, the factor identification protocol step eliminates STR from consideration.

In total, 12 of the 23 candidate factors are eliminated by this step. Quite intuitively, this step eliminates all five Fama & French (2015) equity factors. The factors being kept for further consideration include {MKTb, BVL, CRY, DEF, DUR, LTR, MOMb, MOMs, TERM, UNC, VOL}. Next, we aim to form an optimal factor model from a subset of these factors.

Before doing so, we should have a look at the correlations of the factors surviving the first-step factor identification. We present these correlations in panel C of Table 3.3. Many factors are moderately correlated. Part of the factors have positive correlations in excess of 0.4 with the aggregate bond market: BVL, CRY, DEF, DUR, and TERM. On the other hand, LTR, MOMb, MOMs, UNC, and

Table 3.3: Factor Identification Protocol

This table presents the results of the first-step factor identification protocol. In Panel A are the canonical correlations of the 10 principal components extracted from our test portfolio set and the factor candidates. We do this separately for the two halves of our sample period. In Panel B, we present the results of the actual factor identification protocol. We regress the 10 canonical variates, which are all weighted averages of the 10 principal components, each on a constant and the set of factor candidates. For each factor, we report the average t-statistics from regressions with all canonical variates and the average t-statistics from regressions with those variates associated with significant (at 5%) canonical correlations. In addition, we report the number of significant t-statistics in the two halves of our sample period. The requirement for passing the first-step factor identification protocol is that the average of the absolute t-statistics associated with the significant canonical correlations exceeds 1.96 and the average number of significant t-statistics over the two periods is more than 2.5 Pukthuanthong et al. 2019. We indicate with a ✓ those factors that pass the first-stage screening implied by these two conditions. The factors failing this step of the factor identification protocol are indicated with a $-$. In Panel C, we present the correlations of factors surviving the first-stage screening.

Panel A: Canonical correlations

Canonical value	1	2	3	4	5	6	7	8	9	10
First half										
Canonical correlation	1.00***	0.99***	0.98***	0.96***	0.94***	0.91***	0.89***	0.78***	0.62**	0.51
z-statistic	(37.0)	(32.6)	(28.6)	(24.6)	(21.1)	(17.8)	(14.3)	(10.5)	(7.09)	(4.42)
Second half										
Canonical correlation	1.00***	0.99***	0.98***	0.95***	0.92***	0.86***	0.85***	0.73***	0.63**	0.46
z-statistic	(37.9)	(31.9)	(27.2)	(23.0)	(19.4)	(16.2)	(13.3)	(9.85)	(6.93)	(3.97)

continued on the next page

Table 3.3: Factor Identification Protocol (continued)

Panel B: Factor identification protocol

	Bond Factors											
	MKTb	BVL	CRF	CRY	DEF	DRF	DUR	EPU	EPUtax	LTR	LRF	MOMb
Avg. t-stat	4.45	2.01	1.00	2.76	1.87	1.74	3.57	1.06	1.27	1.99	1.63	2.26
Avg. t-stat sig.	4.90	2.16	1.07	2.94	2.01	1.91	3.94	1.08	1.37	2.12	1.71	2.37
Sig. t-stats first half	7.0	3.0	0.0	3.0	5.0	3.0	4.0	2.0	2.0	3.0	6.0	3.0
Sig. t-stats second half	7.0	4.0	2.0	6.0	3.0	3.0	5.0	2.0	2.0	5.0	1.0	5.0
Avg.	7.0	3.5	1.0	4.5	4.0	3.0	4.5	2.0	2.0	4.0	3.5	4.0
Pass first step?	✓	✓	–	✓	✓	–	✓	–	–	✓	–	✓

	Bond Factors						Equity Factors				
	MOMs	STR	TERM	UNC	VAL	VOL	MKTs	SMB	HML	RMW	CMA
Avg. t-stat	2.38	1.19	2.29	1.92	1.57	2.25	1.39	1.06	1.13	1.17	1.01
Avg. t-stat sig.	2.59	1.22	2.51	2.02	1.71	2.39	1.51	1.12	1.22	1.27	1.11
Sig. t-stats first half	5.0	2.0	6.0	2.0	3.0	2.0	4.0	1.0	2.0	2.0	1.0
Sig. t-stats second half	3.0	1.0	4.0	6.0	2.0	6.0	2.0	1.0	0.0	0.0	1.0
Avg.	4.0	1.5	5.0	4.0	2.5	4.0	3.0	1.0	1.0	1.0	1.0
Pass first step?	✓	–	✓	✓	–	✓	–	–	–	–	–

Panel C: Correlations of the surviving factors

	MKTb	BVL	CRY	DEF	DUR	LTR	MOMb	MOMs	TERM	UNC	VOL
MKTb	1.00	0.87	0.65	0.45	0.89	0.09	−0.33	−0.12	0.48	0.22	0.08
BVL		1.00	0.71	0.56	0.97	0.08	−0.35	−0.18	0.34	0.24	0.10
CRY			1.00	0.59	0.65	0.11	−0.40	−0.37	−0.04	−0.03	0.27
DEF				1.00	0.44	0.26	−0.35	−0.38	−0.45	−0.00	0.09
DUR					1.00	0.01	−0.28	−0.08	0.47	0.29	0.09
LTR						1.00	−0.49	−0.25	−0.21	−0.37	0.09
MOMb							1.00	0.37	0.04	0.21	−0.28
MOMs								1.00	0.26	0.13	0.07
TERM									1.00	0.24	−0.21
UNC										1.00	0.01
VOL											1.00

VOL have rather small correlations with the aggregate bond market. Among the other factors, we find that TERM and DEF are negatively correlated, consistent with Fama & French (1993). BVL, CRY, DEF, and DUR are also positively correlated among one another. MOMb and MOMs have negative correlations with most other factors. Finally, LTR, UNC, and VOL are not strongly correlated to most other factors.

The highest correlations are between the set of factors {MKTb, BVL, DUR}, each of which exceeds 0.8. It is thus likely that these factors capture similar economic risk sources (Gospodinov & Robotti 2021). Hence, going forward we will not consider models that include more than one of these factors.

The fact that the surviving factors apparently drive systematic movements in corporate bond returns and that most of them yield a statistically significant average return indicates that they could all be useful for pricing corporate bonds. The substantial correlations among different factors, on the other hand, suggest that the factors are not all that different. Thus, some factors are likely redundant and a parsimonious optimal factor model does not need all of them.

3.4.3 Model selection results

In a next step, we thus use the model selection approach to form an optimal model out of the 11 candidate factors. That is, we perform the second model selection step using the approach of Barillas & Shanken (2018) and Chib et al. (2020).

Panel A of Table 3.4 reports the log marginal likelihoods, the posterior probabilities, and the ratios of the posterior probability to the prior probability of the top models. We further illustrate the posterior probabilities for all models in Figure 3.1. The best combination of factors includes CRY, DUR, MOMs, and TERM. The model made up of these four factors yields the highest log marginal likelihood and, hence, the largest posterior probability. Thus, carry risk, duration risk, stock momentum, and term risk appear to be the most important factors in corporate bond markets in our sample. We call this set of factors the "no. 1 winning model," or just "winning model."

Table 3.4: Model Scan Result

This table summarizes the main results of the model scanning algorithm. We examine all 1,024 asset pricing models that, subject to the restrictions on factor correlations, can be formed with the eleven candidate factors that survive the first-step factor identification from August 2006 to December 2019 using the BS–CZZ approach. Panel A reports the log marginal likelihoods ($\log \tilde{m}(y|M_j)$), posterior probabilities ($Pr(M_j|y)$), and the ratios of the posterior over the prior probabilities ($\frac{Pr(M_j|y)}{Pr(M_j)}$) of the top four models. We pick the reported top models based on the Bayes factor. All others have a posterior probability more than 3.2 times lower than the no. 1 winning model. Panel B reports the same statistics for the existing models that are spanned by the surviving factors. These include CAPMbond: {MKTb} and FF3: {MKTb, TERM, DEF}. In addition, we present the results for the model with all eleven factors. Panel C provides the cumulative posterior probabilities of the individual factors. For this, we sum up the posterior probabilities of all models that contain a factor.

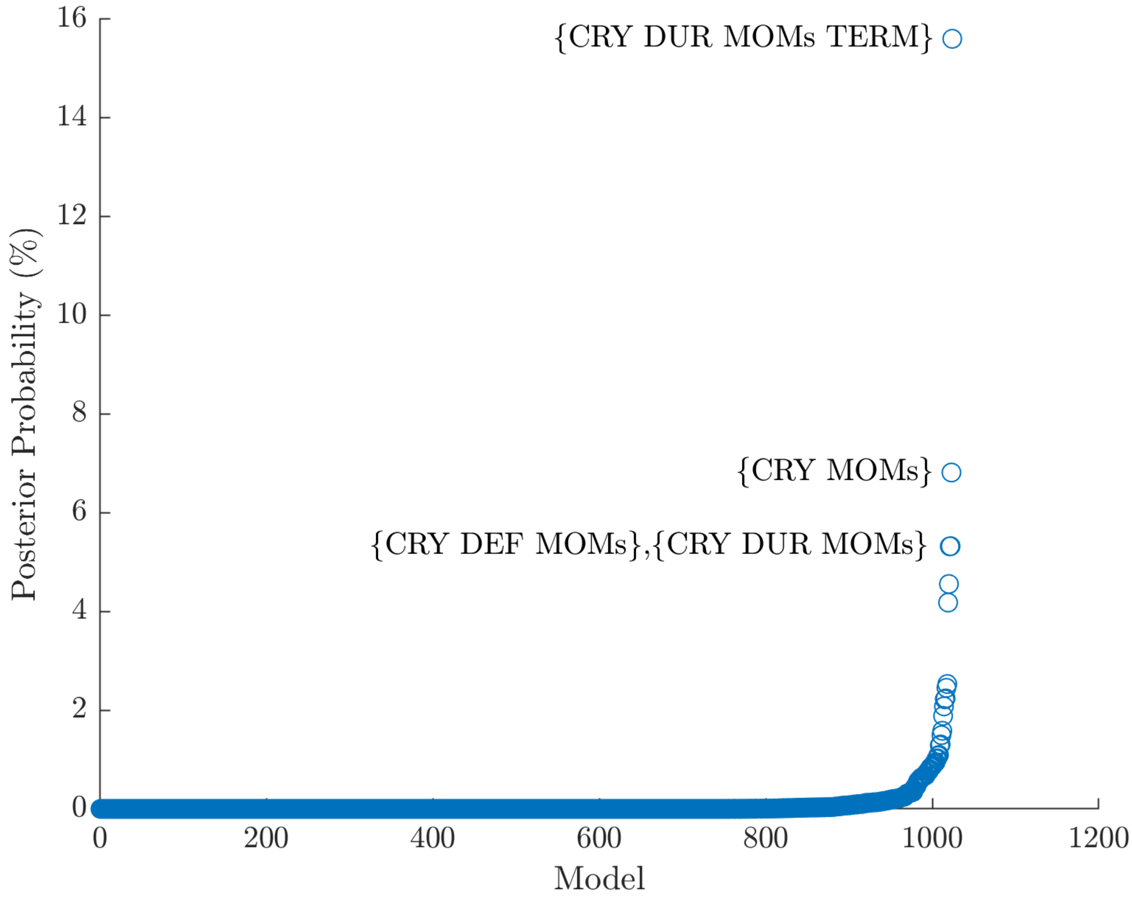
Risk Factors	$\log \tilde{m}(y\|M_j)$	$Pr(M_j\|y)$	$\frac{Pr(M_j\|y)}{Pr(M_j)}$
Panel A: Top models			
CRY DUR MOMs TERM	4,833.81	15.6	160
CRY MOMs	4,832.98	6.82	69.8
CRY DEF MOMs	4,832.73	5.32	54.5
CRY DUR MOMs	4,832.73	5.32	54.5
Panel B: CAPMbond, FF3, and the model with all factors			
MKTb	4,817.33	0.00	0.00
MKTb DEF TERM	4,814.41	0.00	0.00
MKTb BVL CRY DEF DUR LTR MOMb MOMs TERM UNC VOL	4,824.74	0.00	0.02

Panel C: Cumulative posterior probabilities of the factors

	MKTb	BVL	CRY	DEF	DUR	LTR
Cumulative posterior probability (%)	4.765	12.37	100.0	29.20	55.79	19.72
	MOMb	MOMs	TERM	UNC	VOL	
Cumulative posterior probability (%)	21.16	99.96	47.78	12.66	15.74	

Figure 3.1: The Model Scan Result

This figure illustrates the results of the model scanning algorithm. We plot the posterior probabilities $Pr(M_j|y)$ for all 1,024 models. The models are ranked in ascending order by their posterior probabilities. Beside the dots for the four winning models, separated by the Bayes factor cutoff from the remaining ones, we indicate the respective factors.



We view the model selection approach not only as a tool to determine the optimal model but also as one that helps us find the most important factors. Thus, it is also worth having a look at the next-best factor sets. In panel A of Table 3.4, we report the top-four models based on a Bayes factor cutoff.[13] The second-best factor model contains only 2 of the 4 factors of the winning model: CRY and MOMs. The third- and fourth-best models also include these two factors. They only differ in the additional factor included (DEF in case of the third-best and DUR in case of the fourth-best model).[14]

Thus, while quite naturally, based on a large set of candidate factors and a rather short sample period, the posterior probability of the winning model does not approach unity, a clear pattern emerges around the set of winning factors. This information is also reflected by the cumulative posterior probabilities of the factors presented in panel C of Table 3.4. These are 100.0% for CRY, 99.96% for MOMs, 55.79% for DUR, and 47.78% for TERM. All other factors have cumulative posterior probabilities lower than 30%. Interestingly, the bond market factor yields the lowest cumulative posterior probability of 4.77%.

The DEF factor is only included in one of the top-four models and has a cumulative posterior probability of 29.20%. Thus, its explanatory power for the cross-section of corporate bond returns in our sample appears to be limited. This is surprising in light of the results of Gebhardt et al. (2005a), who show that DEF betas perform well in explaining cross-sectional variation in beta-sorted portfolios. Thus, while performing well for these, the DEF factor appears to be much less able to explain the returns of other characteristics-sorted portfolios.[15]

Another metric to judge the performance of the selected models is the ratio of the posterior model probability to the prior model probability of any model

[13] The Bayes factor postulates substantial/significant differences between two models if the marginal likelihood is different by more than $\log(10^{0.5}) = 1.15$ or, equivalently, the posterior probability is lower by a factor of more than 3.2 (Kass & Raftery 1995).

[14] Note that as further factors are added to the winning set, the posterior probabilities may deteriorate markedly. For example, when adding all factors, the posterior is 0.00%. This is because the model selection algorithm is designed to encourage parsimony. Models that include redundant factors receive lower posterior probabilities.

[15] One reason for the difference is likely that Gebhardt et al. (2005a) study investment-grade bonds only, while our sample contains both high-yield and investment-grade bonds.

$\mathbb{M}_j$, denoted by $\frac{Pr(M_j|y)}{Pr(M_j)}$. This ratio reflects the information improvement of the posterior over the prior, which is the same for all models, given the data observed. In the case of the winning model, improvement is very clear. Its posterior is more than 160 times as high as its prior.

In panel B of Table 3.4, we also examine the performance of the existing factor models spanned by the set of factors that survive the first-step screening. We find that both the corporate bond CAPM and the FF3 model perform poorly. The posterior probabilities are 0.00% and the ratios of the posterior probability to the prior probability are 0.00. A model with all factors (which we consider as a benchmark by way of exception, despite the high correlations of MKTb, BVL, and DUR) also performs rather poorly with a posterior-to-prior-probability ratio of 0.02. This implies that all 11 candidate factors together contain information redundancies. Thus, in the trade-off between slightly enhanced in-sample performance and the parsimony encouraged by the Barillas & Shanken (2018) and Chib et al. (2020) model selection approach, adding all these additional factors hurts the model performance.

Thus, the most important set of factors in corporate bond markets appears to consist of CRY, DUR, MOMs, and TERM. Carry reflects the return of an asset if the market conditions stay the same (Koijen, Moskowitz, Pedersen, & Vrugt 2018). As such, it is not unique to corporate bond markets. However, it is an important measure of risk and expected return. Duration and TERM are important due to the interest rate risk, which is a unique feature of bond markets that strongly differs from equity markets. Corporate bonds with higher interest rate risk earn systematically larger returns. Finally, high stock momentum increases the equity cushion available and reduces firm leverage, hence making the more senior claims of corporate bonds less risky. The factors associated with these corporate bond characteristics seem to systematically drive and explain corporate bond returns.

3.5 Asset Pricing Tests

3.5.1 Model Sharpe ratios

Having selected an optimal set of factors, we next turn to analyzing whether the selected winning models outperform other factor models based on more traditional model comparison approaches. That is, instead of Bayesian statistics, in this section, we use classical statistics and conduct pairwise tests of the equality of squared Sharpe ratios following Barillas et al. (2020). Their method enables us to provide reliable inference regarding relative model performance gauged by the squared Sharpe ratio improvement.

Table 3.5 reports the differences between the sample squared Sharpe ratios (column model minus row model) of different pairs of models. The estimated model squared Sharpe ratios are modified to be unbiased in small samples. The associated p-values are shown in brackets.

The final column of Table 3.5 clearly indicates that the top factor model of the model selection approach dominates all other existing models by producing a higher Sharpe ratio. The bias-adjusted squared Sharpe ratio of the winning model is higher by 0.45 compared to the CAPMbond model. For the other factor models, the improvements are generally only somewhat smaller. For example, compared to the FF3 model, the improvement is 0.43, compared to the augmented FF3 model 0.38, and compared to the BBW model 0.40. In terms of squared Sharpe ratios, the IPR and KPP models perform best among the existing models. However, the squared Sharpe ratio improvement of the no. 1 winning model is still 0.04 and 0.17, respectively. All these Sharpe ratio differences are highly statistically significant, as specified by the corresponding p-values that are virtually zero in all instances.

The no. 1 winning model also outperforms the second- to fourth-best models of the model selection. Its squared Sharpe ratios are significantly larger. The no. 2 to 4 winning models also outperform all other models except for the IPR

Table 3.5: Tests of the Equality of Squared Sharpe Ratios

This table presents pairwise tests of the equality of the squared Sharpe ratios of the existing corporate bond pricing models and the four winning models from the model scan (sorted in ascending order such that winning 1 is the top model from Table 3.4). The existing corporate bond factor models include: (i) CAPMbond: {MKTb}, (ii) FF3: {MKTb, TERM, DEF}, (iii) aug. FF3: {MKTb, TERM, DEF, LRF, MOMb}, (iv) FF5stkb: {MKTs, SMB, HML, TERM, DEF}, (v) BBW: {MKTb, DRF, CRF, LRF}, (vi) BSW: {MKTb, STR, MOMb, LTR}, (vii) IPR: {CRY, DUR, MOMb, MOMs, VAL}, and (viii) KPP: {MKTb, CRY, DUR, BVL, VAL}. The main body of the table reports the differences between the monthly (bias-adjusted) sample squared Sharpe ratios of the models in column i and row j, $\hat{\theta}_i^2 - \hat{\theta}_j^2$. In brackets, we report the associated p-values for the test of the null hypothesis $H_0 : \hat{\theta}_i^2 = \hat{\theta}_j^2$. *,**, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	FF3	aug. FF3	FF5stkb	BBW	BSW	IPR	KPP	winning 4	winning 3	winning 2	winning 1
CAPMbond	0.020*	0.074***	−0.015***	0.054***	0.082***	0.414***	0.278***	0.390***	0.389***	0.349***	0.450***
	[0.083]	[0.001]	[0.002]	[0.003]	[0.002]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
FF3		0.054**	−0.035**	0.034***	0.061***	0.394***	0.258***	0.370***	0.369***	0.329***	0.429***
		[0.011]	[0.013]	[0.002]	[0.005]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
aug. FF3			−0.089***	−0.020*	0.008***	0.340***	0.204***	0.316***	0.315***	0.275***	0.376***
			[0.000]	[0.062]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
FF5stkb				0.069***	0.096***	0.429***	0.293***	0.405***	0.404***	0.364***	0.464***
				[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
BBW					0.027***	0.360***	0.223***	0.336***	0.335***	0.295***	0.395***
					[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
BSW						0.333***	0.196***	0.309***	0.308***	0.268***	0.368***
						[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
IPR							−0.136***	−0.024	−0.025*	−0.065	0.036***
							[0.000]	[0.202]	[0.072]	[0.183]	[0.001]
KPP								0.113***	0.111***	0.072***	0.172***
								[0.000]	[0.000]	[0.000]	[0.000]
winning 4									−0.001**	−0.041	0.059***
									[0.044]	[0.132]	[0.010]
winning 3										−0.040**	0.061**
										[0.027]	[0.014]
winning 2											0.100***
											[0.006]

model. Among the existing ones, the IPR model clearly performs best. This is not surprising as the model overlaps with the winning model in three of its factors.

Relying on comparisons of in-sample Sharpe ratios is not enough, though. Kan, Wang, & Zheng (2022) show that in the presence of estimation risk, the multifactor in-sample Sharpe ratios are typically unattainable for investors in real time. Therefore, we also analyze out-of-sample Sharpe ratios.

We present the results in Table 3.6. As in Barillas & Shanken (2018), we show the full-sample Sharpe ratios of the models as well as the in- and out-of-sample Sharpe ratios for two different sample splitting schemes. Consistent with the results of Table 3.5, the no. 1 winning model has the highest in-sample Sharpe ratio of 0.76, followed by the IPR model with 0.74 and the other winning models (0.67 up to 0.71).[16]

Next, we have a look at the out-of-sample Sharpe ratios with the different sample splitting schemes. When using the first half of the sample to determine the weights in the tangency portfolio, the out-of-sample Sharpe ratios are smaller than those in-sample and also those that could be achieved with an optimal ex post weighting of the factors. However, the winning models also provide the highest out-of-sample Sharpe ratios, with the best performance being achieved by the no. 4 winning model (0.55), followed by the no. 2 and no. 1 winning models (0.51 and 0.50, respectively). It is not surprising that the no. 4 and no. 2 winning models perform somewhat better than the no. 1 winning model for this exercise as the estimation risk is smaller in these models that have fewer factors than the no. 1 winning model. However, the four winning models have clearly higher out-of-sample Sharpe ratios than all existing models.

Finally, we also examine the out-of-sample Sharpe ratios for a different sample splitting scheme, using two-thirds of the sample period for estimation of the optimal weights in the tangency portfolio. We find, again, that the no. 1, no. 2, and no. 4 winning models achieve the best out-of-sample performance with

[16] Note that the IPR model is not obtainable with the model selection approach because the VAL factor is knocked out by the first-step factor identification. This factor performs extremely well with a mean return of 0.75% per month and a t-statistic of 6.84 (see Table 3.2).

Table 3.6: Out-of-Sample Sharpe Ratios

This table presents the in- and out-of-sample performance of the existing corporate bond pricing models and the four winning models from the model scan (sorted in ascending order such that winning 1 is the top model from Table 3.4). The existing corporate bond factor models include: (i) CAPMbond: {MKTb}, (ii) FF3: {MKTb, TERM, DEF}, (iii) aug. FF3: {MKTb, TERM, DEF, LRF, MOMb}, (iv) FF5stkb: {MKTs, SMB, HML, TERM, DEF}, (v) BBW: {MKTb, DRF, CRF, LRF}, (vi) BSW: {MKTb, STR, MOMb, LTR}, (vii) IPR: {CRY, DUR, MOMb, MOMs, VAL}, and (viii) KPP: {MKTb, CRY, DUR, BVL, VAL}. The first column shows the full-sample monthly Sharpe ratio of the model tangency portfolios. The remainder of the table shows the results for out-of-sample tests where the initial estimation period for the factor weights in the tangency portfolio is half of the sample period ($T/2$) or two thirds of the sample period ($2T/3$). In each case, EST shows the in-sample Sharpe ratio of the estimation period, PERF the in-sample Sharpe ratio of the remaining period, and PERFw the actual out-of-sample Sharpe ratio when using the weights from the first in-sample estimation period.

	T	$T/2$			$2T/3$		
	Sample SR	EST	PERF	PERFw	EST	PERF	PERFw
winning 1	0.756	0.977	0.819	0.503	0.811	0.881	0.615
winning 2	0.670	0.826	0.547	0.511	0.703	0.655	0.589
winning 3	0.706	0.980	0.547	0.461	0.795	0.655	0.523
winning 4	0.707	0.834	0.776	0.545	0.726	0.855	0.648
CAPMbond	0.288	0.312	0.271	0.271	0.273	0.356	0.356
FF3	0.343	0.422	0.303	0.237	0.360	0.371	0.305
aug. FF3	0.434	0.542	0.377	0.300	0.488	0.399	0.310
FF5stkb	0.309	0.320	0.396	0.223	0.324	0.412	0.243
BBW	0.401	0.469	0.351	0.325	0.428	0.444	0.347
BSW	0.434	0.676	0.347	0.153	0.541	0.388	0.157
IPR	0.738	0.922	0.814	0.448	0.782	0.932	0.568
KPP	0.634	0.927	0.693	0.360	0.767	0.767	0.379

Sharpe ratios between 0.59 and 0.65.

3.5.2 Spanning tests

In this section, we conduct two sets of factor spanning tests. The main questions are: Which of the factors are most important? Which factors explain time-series variation in others? For what factors do the existing models fail most strongly? Thus, while these exercises do not provide new insights into which model is the best, they help us to better understand the winning model(s) superior performance compared to the existing ones.

First, we run the spanning regressions of the nonoverlapping factors of the no. 1 winning model (which largely also overlap with those in models two to four) on the alternative existing models to see how those factors not included in the existing models add information to the existing model benchmarks. Sizable and significant alphas indicate that the noncommon factors of the best model can add more power to explain average returns, which is missed by the benchmark models.

We present the results in Table 3.7. We find that the bond CAPM fails for the CRY and MOMs factors. It can explain the DUR factor (with which it is highly correlated) and the TERM factor. All versions of the Fama-French factor models and the BBW and BSW models also fail for the CRY and MOMs factors. The IPR model, which contains both CRY and MOMs, in turn fails to explain the TERM factor. Finally, for the KPP model, both MOMs and TERM have significant positive alphas. For each model, the GRS test rejects the null hypothesis that all alphas for a given factor model are jointly zero.

Next, we turn the table and try to explain the factors not in the winning set with the selected factor model. We present the results in Table 3.8. Splitting the analysis into two parts, we first present the spanning tests for the factors that pass the first-step identification before also turning to those that do not.

Starting with the factors that pass the identification protocol, we present the results in panel A of Table 3.8. We find that the bond market factor is well explained by the winning factor model. It has significant exposures to the DUR,

Table 3.7: Spanning Tests: Regressions of the Winning Factors on Various Existing Models

This table summarizes the results of spanning regressions of the non-overlapping factors of the no. 1 winning model {CRY, DUR, MOMs, TERM} on the existing corporate bond factor models including: (i) CAPMbond: {MKTb}, (ii) FF3: {MKTb, TERM, DEF}, (iii) aug. FF3: {MKTb, TERM, DEF, LRF, MOMb}, (iv) FF5stkb: {MKTs, SMB, HML, TERM, DEF}, (v) BBW: {MKTb, DRF, CRF, LRF}, (vi) BSW: {MKTb, STR, MOMb, LTR}, (vii) IPR: {CRY, DUR, MOMb, MOMs, VAL}, and (viii) KPP: {MKTb, CRY, DUR, BVL, VAL}. α is the intercept from a spanning regression. The t-statistics in parentheses are based on robust Newey & West (1987) standard errors with 4 lags. R^2 presents the coefficient of determination of the single spanning regressions. *GRS* indicates the results for the Gibbons et al. (1989) test of the null hypothesis that all alphas are jointly zero for a model. Beside the GRS test statistics in brackets we present the corresponding p-values. *,**, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	CAPMbond			FF3			aug. FF3			FF5stkb		
	α	(t-statistic)	R^2	α	(t-statistic)	R^2	α	(t-statistic)	R^2	α	(t-statistic)	R^2
CRY	0.56***	(3.03)	43.8	0.58***	(3.15)	60.9	0.43***	(3.18)	67.7	0.77***	(3.90)	44.8
DUR	−0.13	(−1.38)	79.0	−0.06	(−0.54)	81.5	−0.00	(−0.04)	82.2	0.00	(0.01)	80.2
MOMs	0.26***	(3.77)	2.11	0.23***	(3.55)	16.3	0.31***	(5.28)	28.4	0.23***	(2.66)	19.8
TERM	0.09	(0.39)	18.5									
GRS	17.7***	[0.00]		21.1***	[0.00]		18.0***	[0.00]		29.5***	[0.00]	

	BBW			BSW			IPR			KPP		
	α	(t-statistic)	R^2	α	(t-statistic)	R^2	α	(t-statistic)	R^2	α	(t-statistic)	R^2
CRY	0.38***	(2.97)	56.4	0.46***	(2.73)	48.8						
DUR	−0.12	(−1.34)	81.7	−0.04	(−0.36)	80.8						
MOMs	0.32***	(4.41)	11.7	0.26***	(3.81)	16.3				0.26***	(3.20)	43.6
TERM	0.21	(1.10)	33.3	0.20	(0.84)	26.0	0.62**	(2.26)	42.5	0.50**	(2.10)	55.3
GRS	15.6***	[0.00]		13.7***	[0.00]		6.30**	[0.01]		12.9***	[0.00]	

Table 3.8: Spanning Tests: Regressions of the Other Factors on the Winning Model

This table reports the results of spanning regressions of factors not selected on the no. 1 winning model {CRY, DUR, MOMs, TERM} from the model scanning procedure. We categorize the alternative factors into those that pass the first-step identification (Panel A) and those that do not (Panel B). We present the intercept from the spanning regressions (α) as well as the excluded factors' loadings on the winning model factors. The t-statistics in parentheses are based on robust Newey & West (1987) standard errors with 4 lags. R^2 presents the coefficient of determination of the single spanning regressions. *GRS* indicates the results for the Gibbons et al. (1989) test of the null hypothesis that all alphas are jointly zero. Below the GRS test statistic in brackets we present the corresponding p-value. We separately test the joint significance of the factors that pass the first-step identification (in Panel A) and that of all factors including those that pass the first-step identification and those that do not. *,**, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	α	CRY	DUR	MOMs	TERM	R^2
Panel A: Factors that pass the first-step identification						
MKTb	0.03	0.16***	0.33***	−0.06	0.08***	82.6
	(0.44)	(2.67)	(8.55)	(−0.77)	(4.85)	
BVL	0.10*	0.06	0.84***	−0.20**	−0.05**	96.6
	(1.89)	(1.61)	(34.1)	(−2.36)	(−2.32)	
DEF	0.03	0.01	0.64***	−0.29*	−0.51***	76.2
	(0.28)	(0.09)	(4.84)	(−1.74)	(−6.74)	
LTR	0.23	−0.05	0.08	−0.43**	−0.11	9.17
	(1.61)	(−0.51)	(0.48)	(−2.35)	(−1.58)	
MOMb	−0.24*	−0.20*	−0.07	0.53*	−0.01	24.7
	(−1.70)	(−1.69)	(−0.66)	(1.85)	(−0.24)	
UNC	0.03	−0.20	0.24**	0.08	0.00	17.9
	(0.34)	(−1.36)	(2.21)	(0.44)	(0.01)	
VOL	−0.02	0.12	−0.00	0.24	−0.06	18.5
	(−0.28)	(1.27)	(−0.01)	(1.42)	(−1.42)	
GRS	1.44					
	[0.19]					
Panel B: Remaining corporate bond factors						
CRF	0.41***	−0.22	0.51***	0.06	−0.23***	23.9
	(3.19)	(−1.50)	(3.06)	(0.33)	(−3.08)	
DRF	0.26	0.36*	0.05	−0.19	0.19**	21.5
	(1.26)	(1.89)	(0.48)	(−0.46)	(2.41)	
EPU	0.15**	−0.14**	0.25***	−0.12	−0.07**	26.9
	(2.47)	(−2.04)	(3.16)	(−1.21)	(−2.10)	
EPUtax	0.06	−0.04	0.07**	−0.03	−0.07**	9.43
	(0.95)	(−0.64)	(2.27)	(−0.35)	(−2.06)	
LRF	0.08	0.46***	−0.10	−0.19	0.07**	41.2
	(0.99)	(3.43)	(−1.31)	(−1.31)	(2.11)	
STR	0.15	0.27**	−0.28***	0.30	0.03	15.2
	(1.10)	(2.56)	(−4.58)	(1.37)	(0.60)	
VAL	0.05	0.48***	0.07	0.60***	0.04	62.8
	(0.55)	(6.86)	(1.47)	(3.68)	(1.27)	
GRS	1.86**					
	[0.04]					

CRY, and TERM factors (sorted by the size of the factor sensitivities). The alpha is 0.03% and clearly not statistically significant. All other factors in this set also can be explained reasonably well by the set of winning factors. The individual factor alphas are all close to zero and generally much smaller than the factor average returns. None of the alphas is statistically significant at 5%. The BVL and MOMb factors, though, have alphas that are significant at 10%. The GRS test does not reject the hypothesis that all these alphas are jointly zero. Thus, the winning factor model does a very good job in summarizing the information contained in those factors that systematically move corporate bond prices.

Next, we cast the net wider and test if the winning model can also explain the factors that have been rejected by the factor identification protocol. We present the results in panel B of Table 3.8.[17] The DRF, EPUtax, LRF, STR, and VAL long-short returns are well explained by the winning factors. However, those for CRF and EPU are not. The GRS test rejects the null hypothesis that the alphas of all factor candidates from panels A and B are jointly zero. Thus, credit risk and economic policy uncertainty still appear to be anomalies with respect to the winning factor model. These either reflect mispricing or suggest that the optimal corporate bond factor model should also include further, yet undiscovered factors.[18] Finally, all four factors help to explain time variation in the returns of other factors. Seven other factor candidates are significantly exposed to CRY, eight to DUR, five to MOMs, and eight to TERM.

3.5.3 Time-series tests with test assets

In this section, we investigate the empirical performance of the winning model for various test assets in the time-series domain. While the RHS approach is

[17]We skip the five equity factors for this analysis since corporate bond and equity markets are potentially segmented (Chordia et al. 2017; Choi & Kim 2018).

[18]The model, though, should not be expanded with CRF and EPU factors as these do not significantly move corporate bond prices. They both fail the factor identification clearly, not narrowly. It is possible that the true factor(s) behind these anomalies are only weakly correlated with the CRF and EPU portfolio returns and noise in these returns masks the price-moving signals.

elegant and useful, in practice many factor model users may remain interested in understanding how factor models explain LHS returns. Furthermore, it is interesting if there are any (and if yes, which) sets of test portfolios that still produce significant alphas with respect to the best factor models.

A factor model that can explain a variety of unrelated anomalies appears more useful than one that is only able to explain its own factors. More importantly, to improve the power of asset pricing tests, Lewellen et al. (2010) suggest testing risk factors based on additional test portfolios that are not related to the risk characteristics used to construct those factors. Thus, we use a comprehensive set of test assets.

Table 3.9 summarizes the results. We start with long-short portfolios generated from the 23 corporate bond characteristics of Kelly & Pruitt (2022) that are not used to construct factors.[19] As for the factors, we use 25 double-sorted portfolios (generally with rating) to calculate the value-weighted long-short returns. We present the results of time-series tests for these test assets in panel A of Table 3.9. We find that none of the models jointly explains all characteristic long-short returns. The GRS test rejects in every instance. However, since it is well known that the GRS test tends to overreject its null hypothesis in finite samples (Bekaert & De Santis 2021), its results should not be taken at face value. We can see that the four winning models perform quite well compared to the existing models. They yield the lowest GRS statistics and also the lowest squared Sharpe ratios achievable from the alphas of the characteristic long-short test portfolios. The no. 1 winning model yields one of the lowest average absolute alphas and one of the largest time-series R^2s. Again, the IPR and KPP models, which have some overlap with the winning model(s) in their factors, perform quite well, too.

In panels B–D of Table 3.9, we also examine alternative test portfolios (as,

[19] These characteristics include bond face value, maturity, bond age, coupon, face value, book-to-price, debt-to-EBITDA, earnings-to-price, equity market cap, equity volatility, firm total debt, industry momentum, momentum times ratings, book leverage, market leverage, turnover volatility, operating leverage, profitability, profitability change, rating, distance-to-default, bond skewness, and momentum spread. For more information on these characteristics, see Table A.I of Kelly et al. (2023).

Table 3.9: Time-Series Asset Pricing Tests with Test Assets

This table reports the results for test-asset-based time-series asset pricing tests of the existing corporate bond pricing models and the four winning models from the model scan (sorted in ascending order such that winning 1 is the top model from Table 3.4). The existing corporate bond factor models include: (i) CAPMbond: {MKTb}, (ii) FF3: {MKTb, TERM, DEF}, (iii) aug. FF3: {MKTb, TERM, DEF, LRF, MOMb}, (iv) FF5stkb: {MKTs, SMB, HML, TERM, DEF}, (v) BBW: {MKTb, DRF, CRF, LRF}, (vi) BSW: {MKTb, STR, MOMb, LTR}, (vii) IPR: {CRY, DUR, MOMb, MOMs, VAL}, and (viii) KPP: {MKTb, CRY, DUR, BVL, VAL}. In Panel A, we examine the performance for 23 long–short portfolios based on the Kelly & Pruitt (2022) dataset. In Panels B and C, we use 25 double-sorted size–maturity and maturity–rating portfolios as test assets, respectively. Finally, in Panel D, we report the results for 12 Fama & French (1997) industry portfolios. In the different panels, GRS indicates the results for the Gibbons et al. (1989) test of the null hypothesis that all alphas are jointly zero for a model, with the corresponding p-value in brackets. $A|\alpha_i|$ is the average absolute alpha of the test portfolios. $\#sig\ \alpha_i$ reports how many test portfolios have significant alphas at the 10% level. We use Newey & West (1987) standard errors with 4 lags. $\frac{A|\alpha_i|}{A|r_i|}$ is the ratio of the average absolute alpha to the average absolute portfolio return. $\frac{A\alpha_i^2}{Ar_i^2}$ is the ratio of the respective squares. $\frac{As^2(\alpha_i)}{A\alpha_i^2}$ is the ratio of the average squared standard error of the alphas to the average squared alpha. $A(R^2)$ is the average adjusted R^2 of the regressions (in percentage points). $SH^2(f)$ is the squared Sharpe ratio of the optimal portfolio from the model factors and $SH^2(\alpha)$ is the squared Sharpe ratio attainable with the alphas of the test assets. *,**, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	GRS	[p-value]	$A\|\alpha_i\|$	$\#sig\ \alpha_i$	$\frac{A\|\alpha_i\|}{A\|r_i\|}$	$\frac{A\alpha_i^2}{Ar_i^2}$	$\frac{As^2(\alpha_i)}{A\alpha_i^2}$	$A(R^2)$	$SH^2(f)$	$SH^2(\alpha)$
Panel A: Long–short anomaly portfolios										
winning 1	2.91***	[0.00]	0.09	8	0.41	0.17	0.29	55.5	0.57	0.76
winning 2	3.13***	[0.00]	0.10	5	0.48	0.29	0.51	35.4	0.45	0.75
winning 3	3.00***	[0.00]	0.11	6	0.53	0.25	0.56	40.5	0.50	0.75
winning 4	2.81***	[0.00]	0.05	6	0.26	0.07	0.93	50.4	0.50	0.70
CAPMbond	5.21***	[0.00]	0.12	7	0.56	0.35	0.24	20.9	0.08	0.94
FF3	5.61***	[0.00]	0.10	6	0.48	0.32	0.24	34.3	0.12	1.04
aug. FF3	5.45***	[0.00]	0.11	11	0.53	0.29	0.19	40.0	0.19	1.08
FF5stkb	5.99***	[0.00]	0.11	5	0.53	0.42	0.22	35.4	0.10	1.09
BBW	4.54***	[0.00]	0.13	11	0.59	0.29	0.24	31.5	0.16	0.88
BSW	5.11***	[0.00]	0.10	7	0.47	0.26	0.31	27.0	0.19	1.01
IPR	3.42***	[0.00]	0.07	7	0.32	0.09	0.57	56.9	0.54	0.88
KPP	3.63***	[0.00]	0.08	6	0.36	0.11	0.46	55.8	0.40	0.85

continued on the next page

Table 3.9: Time-Series Asset Pricing Tests with Test Assets (continued)

	GRS	[p-value]	$A\|\alpha_i\|$	$\#sig\ \alpha_i$	$\frac{A\|\alpha_i\|}{A\|r_i\|}$	$\frac{A\alpha_i^2}{Ar_i^2}$	$\frac{As^2(\alpha_i)}{A\alpha_i^2}$	$A(R^2)$	$SH^2(f)$	$SH^2(\alpha)$
Panel B: Size–maturity portfolios										
winning 1	2.70***	[0.00]	0.09	13	0.22	0.05	0.41	66.9	0.57	0.78
winning 2	2.69***	[0.00]	0.12	6	0.29	0.15	0.74	38.8	0.45	0.71
winning 3	2.57***	[0.00]	0.11	6	0.27	0.12	0.90	39.5	0.50	0.71
winning 4	2.91***	[0.00]	0.10	17	0.25	0.07	0.36	65.8	0.50	0.80
CAPMbond	4.03***	[0.00]	0.10	20	0.25	0.06	0.21	73.3	0.08	0.80
FF3	3.82***	[0.00]	0.08	17	0.19	0.04	0.31	77.3	0.12	0.78
aug. FF3	3.50***	[0.00]	0.08	19	0.21	0.05	0.20	78.4	0.19	0.76
FF5stkb	4.25***	[0.00]	0.17	20	0.41	0.16	0.23	55.9	0.10	0.86
BBW	4.05***	[0.00]	0.10	20	0.26	0.07	0.18	76.1	0.16	0.86
BSW	3.88***	[0.00]	0.08	19	0.20	0.04	0.32	75.9	0.19	0.85
IPR	2.66***	[0.00]	0.10	16	0.24	0.06	0.43	68.7	0.54	0.75
KPP	3.48***	[0.00]	0.07	12	0.16	0.03	0.47	79.3	0.40	0.90
Panel C: Maturity–rating portfolios										
winning 1	2.64***	[0.00]	0.11	12	0.27	0.09	0.27	64.0	0.57	0.76
winning 2	2.95***	[0.00]	0.13	4	0.32	0.18	0.68	34.7	0.45	0.78
winning 3	2.84***	[0.00]	0.11	7	0.27	0.12	0.95	36.1	0.50	0.78
winning 4	2.71***	[0.00]	0.11	17	0.27	0.08	0.40	59.5	0.50	0.75
CAPMbond	4.65***	[0.00]	0.13	16	0.30	0.11	0.29	64.0	0.08	0.92
FF3	4.78***	[0.00]	0.11	14	0.26	0.09	0.32	68.9	0.12	0.98
aug. FF3	4.13***	[0.00]	0.11	13	0.28	0.12	0.23	70.5	0.19	0.90
FF5stkb	5.22***	[0.00]	0.17	17	0.41	0.18	0.30	53.6	0.10	1.05
BBW	4.06***	[0.00]	0.12	17	0.29	0.10	0.22	69.1	0.16	0.87
BSW	4.05***	[0.00]	0.11	12	0.26	0.10	0.37	66.6	0.19	0.88
IPR	2.54***	[0.00]	0.10	12	0.24	0.07	0.53	62.0	0.54	0.72
KPP	2.89***	[0.00]	0.09	8	0.21	0.08	0.37	72.2	0.40	0.74
Panel D: Industry portfolios										
winning 1	1.38	[0.19]	0.07	3	0.17	0.04	0.60	81.1	0.57	0.16
winning 2	1.29	[0.23]	0.11	1	0.25	0.10	0.96	38.7	0.45	0.14
winning 3	1.38	[0.19]	0.10	1	0.24	0.08	1.09	39.4	0.50	0.15
winning 4	1.56	[0.12]	0.10	6	0.24	0.07	0.38	78.6	0.50	0.17
CAPMbond	1.31	[0.22]	0.07	4	0.16	0.03	0.58	77.6	0.08	0.10
FF3	1.18	[0.31]	0.06	5	0.14	0.02	0.73	80.1	0.12	0.10
aug. FF3	1.50	[0.14]	0.08	7	0.19	0.04	0.37	81.0	0.19	0.13
FF5stkb	1.75*	[0.07]	0.16	8	0.37	0.14	0.31	67.2	0.10	0.14
BBW	1.87**	[0.05]	0.08	6	0.19	0.04	0.39	81.2	0.16	0.16
BSW	1.52	[0.13]	0.07	4	0.16	0.04	0.55	78.4	0.19	0.13
IPR	1.97**	[0.04]	0.09	3	0.22	0.07	0.42	80.0	0.54	0.22
KPP	1.53	[0.13]	0.07	5	0.17	0.04	0.53	86.6	0.40	0.16

e.g., in Bai et al. 2019). In panel B, we show the results for 25 size-maturity portfolios and in panel C those for 25 maturity-rating portfolios. The results are overall very similar to those for the characteristic long-short portfolios. The four winning models along with the IPR model perform best. For the maturity-rating portfolios, the KPP model also performs well. These models yield the lowest (although still significant) GRS statistics, small average absolute alphas, and the lowest squared Sharpe ratios of the portfolio alphas.

Panel D of Table 3.9 shows the results for 12 Fama & French (1997) corporate bond industry portfolios. Most of the factor models can price these. The GRS statistics are generally insignificant. The four winning models are among those not rejected and perform well. The IPR model, on the other hand, fails for industry portfolios with a significant GRS test and the largest squared Sharpe ratio from the industry portfolio alphas.

3.5.4 Cross-sectional asset pricing tests

To complement our time-series asset pricing tests, we next perform cross-sectional tests of the factor models. With these, we can test which factors and factor models perform best for explaining cross-sectional differences in corporate bond returns. To do so, we first regress the time series of each test asset return on a constant and the model factors to determine the full-sample betas. Then, we run a cross-sectional regression of the average test-asset excess returns on a constant and the betas estimated in the first step. We account for model-misspecification and errors-in-variables by using the robust standard errors of Kan, Robotti, & Shanken (2013). We also use the standard errors and hypothesis tests for the ordinary least squares (OLS) and generalized least squares (GLS) R^2s provided by Kan et al. (2013) and report the result of the Shanken (1992) T^2 test, for which the null hypothesis is that all cross-sectional pricing errors are jointly zero.

As test assets, we use all the portfolios examined in the previous subsection: 23 characteristic long-short portfolios, 25 size-maturity portfolios, 25 maturity-rating portfolios, and 12 industry portfolios. This wide range and heterogeneity of test assets is designed to obtain robust results (Lewellen et al. 2010).

The results are in Table 3.10. In the winning models, we find that mainly the CRY and DUR factors can explain cross-sectional differences in corporate bond returns. The risk premiums associated with these factors are economically large and highly statistically significant. The MOMs factor has a marginally insignificant positive risk premium. Finally, TERM does not seem to add much in terms of explanatory power for the cross-section of corporate bond returns. Thus, MOMs and TERM in the no. 1 winning model appear to primarily explain time-series variation in corporate bond prices.[20] Looking at the existing models, we find that MKTb often yields a significant positive risk premium estimate. Also, LRF, STR, MOMb, VAL, and BVL appear to be priced in (part of) the models that they are included in.

Most important, however, when comparing models are the cross-sectional R^2s. The OLS R^2s are all significantly greater than zero. For the no. 1 winning model and the IPR model, the cross-sectional R^2s are highest with 91.3%. Even more importantly from an investment perspective is the GLS cross-sectional R^2, which gives a direct indication of the relative mean-variance efficiency of a factor model (Kandel & Stambaugh 1995). The GLS R^2 is clearly largest for the no. 1 winning model, with 10.1%.[21] Thus, the test-asset-based cross-sectional regression test further underlines the very good performance of the selected winning model.

[20] This is akin to the market factor for equity pricing. While it is essential for explaining time-series variation and the level in equity prices, it has little power to explain cross-sectional differences in average returns. Since the former is also very important, the equity market remains an undisputed risk factor.

[21] Given the comparably short sample period and large overlaps in the factors of the models, the differences in cross-sectional R^2s are often not statistically significant. The OLS R^2 of the no. 1 winning model is significantly larger than those of the CAPMbond and FF5stkb models. The GLS R^2 is significantly larger than those of the CAPMbond, FF3, FF5stkb, and BBW models (the GLS R^2 is also significantly larger than that of the no. 4 winning model).

Table 3.10: Cross-Sectional Asset Pricing Tests

This table reports the results for test-asset-based cross-sectional asset pricing tests of the existing corporate bond pricing models and the four winning models from the model scan (sorted in ascending order such that winning 1 is the top model from Table 3.4). The existing corporate bond factor models include: (i) CAPMbond: {MKTb}, (ii) FF3: {MKTb, TERM, DEF}, (iii) aug. FF3: {MKTb, TERM, DEF, LRF, MOMb}, (iv) FF5stkb: {MKTs, SMB, HML, TERM, DEF}, (v) BBW: {MKTb, DRF, CRF, LRF}, (vi) BSW: {MKTb, STR, MOMb, LTR}, (vii) IPR: {CRY, DUR, MOMb, MOMs, VAL}, and (viii) KPP: {MKTb, CRY, DUR, BVL, VAL}. As test assets, we use the 23 long–short portfolios based on the Kelly & Pruitt (2022) dataset along with the 25 double-sorted size–maturity and maturity–rating portfolios and the 12 Fama & French (1997) industry portfolios. We present the results of cross-sectional tests, where we first estimate full-sample betas for each factor model and each test asset. Then we regress the average test asset returns on these betas, the results of which are presented in this table. In the main part of the table (below the heading Variables), we present the intercept (Const) as well as the cross-sectional risk premia of the factors. For the t-statitics below in parentheses we use the errors-in-variables and model-misspecification consistent standard errors of Kan et al. (2013). In the next two columns we present the OLS R^2 and the GLS R^2 (both in percentage points). For both, the standard errors in braces are based on Kan et al. (2013) and the stars indicate the outcome of the test of the null hypothesis $H_0 : R^2 = 0$. The final column presents the result of the Shanken (1992) T^2 test, for which the null hypothesis is that all cross-sectional pricing errors are jointly zero. The corresponding p-values are in brackets. *,**, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Model	Variables					OLS R^2	GLS R^2	T^2
	Const	CRY	DUR	MOMs	TERM			
winning 1	0.08***	1.07***	0.54***	0.15*	−0.03	91.3***	10.1***	365***
	(3.67)	(6.30)	(2.65)	(1.46)	(−0.11)	{4.92}	{3.51}	[0.00]
	Const	CRY	MOMs					
winning 2	0.08***	0.66***	0.06			84.3**	7.96***	491***
	(3.11)	(2.68)	(0.54)			{11.6}	{2.81}	[0.00]
	Const	CRY	DEF	MOMs				
winning 3	0.09***	0.72***	0.11	0.03		85.4***	8.87***	472***
	(3.44)	(2.78)	(0.40)	(0.28)		{11.1}	{3.18}	[0.00]
	Const	CRY	DUR	MOMs				
winning 4	0.08***	0.99***	0.57***	0.13		89.7***	8.82***	398***
	(3.60)	(4.78)	(2.64)	(1.22)		{5.76}	{3.45}	[0.00]

continued on the next page

Table 3.10: Cross-Sectional Asset Pricing Tests (continued)

Model	Variables						OLS R^2	GLS R^2	T^2
	Const	MKTb							
CAPMbond	0.10***	0.33***					69.6***	1.98***	567***
	(3.90)	(2.96)					{15.9}	{1.24}	[0.00]
	Const	MKTb	TERM	DEF					
FF3	0.04**	0.41***	0.02	0.20			83.5***	2.74***	483***
	(1.87)	(3.18)	(0.07)	(0.79)			{9.44}	{1.56}	[0.00]
	Const	MKTb	TERM	DEF	LRF	MOMb			
aug. FF3	0.04**	0.41***	0.07	0.16	0.38**	−0.31	84.0***	4.79***	463***
	(1.91)	(3.35)	(0.20)	(0.60)	(2.20)	(−0.58)	{8.14}	{1.93}	[0.00]
	Const	MKTs	SMB	HML	TERM	DEF			
FF5stkb	0.07*	0.81	0.88	−2.26***	−0.21	0.53	81.0**	1.65	268***
	(1.61)	(0.62)	(0.54)	(−3.50)	(−0.37)	(1.16)	{12.0}	{1.71}	[0.00]
	Const	MKTb	DRF	CRF	LRF				
BBW	0.09***	0.33***	0.58	0.59**	0.43**		76.9**	4.34***	506***
	(3.71)	(2.72)	(0.55)	(1.69)	(2.04)		{12.1}	{1.73}	[0.00]
	Const	MKTb	STR	MOMb	LTR				
BSW	0.05**	0.42***	0.91***	−0.49	0.37		82.0***	3.87**	349***
	(1.96)	(3.51)	(2.73)	(−1.07)	(0.86)		{10.6}	{2.28}	[0.00]
	Const	CRY	DUR	MOMb	MOMs	VAL			
IPR	0.06***	0.99***	0.56***	−0.59**	0.16	0.53***	91.3***	8.98***	365***
	(3.66)	(4.55)	(2.64)	(−2.22)	(1.01)	(3.68)	{5.24}	{3.59}	[0.00]
	Const	MKTb	CRY	DUR	BVL	VAL			
KPP	0.09***	0.32***	1.00***	0.53***	0.48***	0.75***	89.1***	7.97***	400***
	(2.36)	(2.49)	(5.45)	(2.60)	(2.45)	(4.96)	{7.48}	{2.82}	[0.00]

3.6 Explaining Corporate Bond Factors

Having established the very good performance of the winning models both using the RHS and the test-asset-based LHS approach, we finally turn to a fundamental question: What are the fundamental economic drivers behind the winning set of factors?

The traditional view would be that the factors are likely related to changes in perceptions about macroeconomic variables (e.g., Cochrane 2005; Pukthuanthong et al. 2019). A recent alternative strand in the literature also suggests intermediary frictions as an important driver of variation in asset prices (e.g., He & Krishnamurthy 2013; Adrian, Etula, & Muir 2014; He, Kelly, & Manela 2017; Friewald & Nagler 2019; He, Khorrami, & Song 2022). In particular the corporate bond market, which largely operates with over-the-counter transactions, relies strongly on the services of broker-dealer intermediaries. On top of that, it is also possible that the factors are driven by (il)liquidity, a very important feature in corporate bond markets, or total market risk and risk aversion.

As explanatory variables, we thus follow He et al. (2022) and consider intermediary distress and intermediary inventory. For intermediary distress, we obtain data on the squared intermediary leverage ratio from He et al. (2017) and data on the noise variable from Hu, Pan, & Wang (2013). Intermediary distress is the first principal component of the changes in the two variables. For intermediary inventory, we aggregate the inventories of dealers using data from TRACE. Furthermore, we consider the TED spread as a proxy for intermediary funding costs (Friewald & Nagler 2019). We obtain the data from the Federal Reserve Bank of St. Louis (FRED). As macroeconomic variables, we consider the change in the seasonally adjusted monthly industrial production and the monthly inflation rate. For both, we use the Archival FRED (ALFRED) database, which contains the vintage data available at each point in time. We also consider the corporate bond market illiquidity of Dick-Nielsen, Feldhütter, & Lando (2012). Finally, we include the VIX as a measure of equity risk and investor risk aversion.

The data are from the Chicago Board Options Exchange (CBOE).

For each factor contained in at least 1 of the 4 winning models, we then perform a regression of the monthly returns on a constant and the contemporaneous changes in these variables. We present the results in Table 3.11. We standardize all explanatory variables to have a mean of zero and a standard deviation of one.

Starting with CRY, we find that the factor is significantly negatively related to the change in industrial production, the change in bond illiquidity, and the change in the VIX. Thus, the factor returns are particularly low in times of increasing illiquidity and stock market volatility or risk aversion. This result is also consistent with what one would intuitively expect. Carry returns are high if market conditions stay the same, but if they do not, as indicated by an increase in illiquidity or volatility, the factor performs poorly. On the other hand, CRY returns tend to be high if industrial production decreases. Thus, from a macroeconomic perspective it partially behaves like a hedge.

DEF is also negatively exposed to changes in bond illiquidity and the VIX. DUR, on the other hand, has significant negative exposures to intermediary distress, inflation, illiquidity, and the VIX. Thus, the duration factor is indeed related to intermediary frictions. An increase in intermediary distress clearly reduces the DUR return. This result is intuitively consistent, as corporate bonds with long duration are likely subject to particularly high demand for intermediation since there are likely few counterparties willing to trade in them. Similarly, an increase in consumer prices has a negative impact on the factor. Inflation tends to be followed by interest rate rises, to which long-duration corporate bonds are particularly sensitive. Thus, both the traditional view and the intermediary asset pricing view have some merit in explaining the returns of the duration factor. The exposures to illiquidity and the VIX are similar to those of the CRY and DEF factors.

Next, we analyze the MOMs factor. It has only a weakly significant exposure to one of the explanatory variables: inflation. Thus, when consumer prices

Table 3.11: Explaining Corporate Bond Factors

This table reports the results of regressions of the excess returns of the factors in any of the four winning models on different economic variables. We run contemporaneous multiple time-series regressions of the monthly factor long–short returns on a constant, the change in intermediary distress, the change in inventories held by intermediaries, the change in the TED spread, the change in industrial production, the inflation rate, the change in bond illiquidity, and the change in the VIX. The factor returns are in percentage points and all explanatory variables are standardized to have a mean of zero and a standard deviation of one. The t-statistics (in parentheses) are based on Newey & West (1987) standard errors with 4 lags. *,**, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Adj. R^2 presents the adjusted R^2s (in percentage points).

	CRY	DEF	DUR	MOMs	TERM
Const	0.95***	0.06	0.52***	0.22***	0.46**
	(7.00)	(0.71)	(3.10)	(4.93)	(2.30)
Δintermediary distress	−0.26	−0.21	−0.36**	0.13	0.34
	(−1.30)	(−0.84)	(−2.05)	(1.26)	(0.88)
Δinventory	−0.08	−0.08	−0.19	−0.03	−0.25
	(−0.56)	(−0.93)	(−0.97)	(−0.59)	(−1.44)
ΔTED spread	−0.05	−0.08	0.22	0.13	−0.13
	(−0.30)	(−0.41)	(1.00)	(1.01)	(−0.43)
ΔINDPRO	−0.43**	−0.21	0.20	−0.01	0.65**
	(−2.00)	(−1.38)	(1.21)	(−0.07)	(1.98)
INFL	−0.22	−0.07	−0.38**	−0.11*	−0.32
	(−1.53)	(−0.83)	(−2.03)	(−1.87)	(−1.32)
Δbond illiquidity	−0.57***	−0.80***	−0.72***	−0.03	0.21
	(−2.60)	(−2.90)	(−2.76)	(−0.19)	(0.42)
ΔVIX	−0.42**	−0.55***	−0.69**	0.08	0.23
	(−2.48)	(−3.56)	(−2.46)	(0.95)	(0.53)
Adj. R^2	24.6	38.6	27.7	6.94	5.26

increase, the MOMs factor returns decrease. Finally, TERM is positively exposed to industrial production. When industrial production falls, TERM returns are negative. Thus, this factor appears to be a proxy for macroeconomic risk.

Thus, the main drivers of three of the five most important factors are illiquidity and volatility. However, changes in macroeconomic conditions and intermediary frictions also play a key role for part of the factors.

3.7 Conclusion

To the best of our knowledge, we are the first to comprehensively examine a large set of the most prominent corporate bond factors. We pool factors that originate from different previous studies. First, we establish whether the factors systematically move corporate bond prices. For those that do, we adopt a Bayesian marginal likelihood-based approach proposed by Barillas & Shanken (2018) and Chib et al. (2020). In this second step, we simultaneously compare all 1,024 possible models that can be formed as subsets of these factors.

The main finding that emerges from our analysis is that the best factor model for corporate bond returns is based on the combination of carry, duration, stock momentum, and term structure factors. The result indicates that only a small subset of the 23 considered factors really matters for corporate bond pricing. For example, we find that the prominent recent factors of Bai et al. (2019) among many others do not systematically move prices. Among those that do, the bond market, bond volatility, long-term reversal, bond momentum, uncertainty, and volatility risk seem to be redundant factors.

The prominent existing factor models suggested in the corporate bond literature deliver significantly smaller squared Sharpe ratios than the winning model and fail to explain its noncommon factors. Further analysis shows that the winning model from the Bayesian model scan overall explains reasonably well the time-series and cross-sectional variation of corporate bond returns (represented by various test assets). Among the best-performing existing models are the Israel et al. (2018) and Kelly et al. (2023) models, which share many factors with the winning model.

Our study can help academics and practitioners separate useful factors from redundant ones. Based on our search from the expanding list of bond factors, we build an "optimal" corporate bond factor model. The findings in this chapter thus have important practical implications. The winning factor model can be used as a benchmark model for future research, for investors in corporate bond markets to implement factor-investing strategies, and to evaluate performance.

B Appendix

B.1 Variable Definitions and Factor Construction

Variable Definitions

- **Bond illiquidity (illiq)** (Bao, Pan, & Wang 2011) is constructed to extract the transitory component from the bond price. Specifically, let $\Delta p_{i,t,d} = p_{i,t,d} - p_{i,t,d-1}$ be the log price change for bond i on day d of month t. Then, the final illiquidity measure uses the daily returns of bond i during month t to calculate $illiq_{i,t} = -Cov_t(\Delta p_{i,t,d}, \Delta p_{i,t,d+1})$. Under the assumption that the fundamental value of a bond follows a random walk, this measure only depends on the transitory component of the price. The higher the value of $illiq_{i,t}$ the more illiquid is a bond.

- **Bond volatility (vol)** (Bai et al. 2019) is the bond's volatility over the past 24 months.

- **Credit rating (cr)** (Bai et al. 2019) is measured via the credit ratings provided by rating agencies. Bond-level rating information is from the Mergent FISD historical ratings. All ratings are assigned a number to facilitate the analysis. A larger number indicates higher credit risk, or lower credit rating. Investment-grade bonds have ratings from 1 (refers to AAA) to 10 (BBB-). Non-investment-grade bonds have ratings starting from 11 (BB+).

- **Carry (cry)** (Israel et al. 2018) is measured using the option-adjusted spread (OAS). It is the fixed difference between a bond's (option-adjusted) yield for which the discounted expected payments match the market price and the corresponding Treasury yield.

- **Downside risk (dr)** (Bai et al. 2019) is proxied by the 5% VaR, which is the second-lowest monthly return observation over the past 36 months, then multiplied by -1 for the ease of interpretation.

- **Duration (dur)** (Israel et al. 2018) is the derivative of the value of the bond with respect to the credit spread, divided by the current bond price.

- **Economic uncertainty beta ($\beta^{\mathbf{JLN}}_{\mathbf{UNC}}$)** (Bali et al. 2017a, 2021b) is estimated from monthly rolling regressions of excess bond returns on the economic uncertainty index over a 36-month window, while controlling for the bond market portfolio return (MKTb) for each bond and each month of our sample. We use the Jurado, Ludvigson, & Ng (2015) 1-month-ahead economic uncertainty index from Sydney Ludvigson's website.

- **Policy uncertainty beta ($\beta^{\mathbf{EPU}}_{\mathbf{UNC}}$, $\beta^{\mathbf{EPUtax}}_{\mathbf{UNC}}$)** (Tao et al. 2022) is estimated from monthly rolling regressions of excess bond returns on a policy uncertainty index over a 36-month window, while controlling for MKTs, SMB, HML, DEF, and TERM. We use the economic policy uncertainty index (β^{EPU}_{UNC}) of Baker, Bloom, & Davis (2016) as well as the tax policy uncertainty subindex, as proposed by Lee (2022) (β^{EPUtax}_{UNC}). We download both from https://www.policyuncertainty.com.

- **Short-term reversal, bond momentum, and long-term reversal (str, momb, ltr)** (Jostova et al. 2013; Bali et al. 2017b, 2021a; Bai et al. 2019) are measures based on the bonds' past returns. The short-term reversal of a bond i for month t is its return during the previous month. Bond momentum is the past 6-month cumulative return, while skipping the most recent month. Long-term reversal is the past 36-month cumulative return.

- **Spread to D2D (spr_d2d)** (Correia, Richardson, & Tuna 2012; Kelly et al. 2023) is the option-adjusted spread (see *Carry*) divided by one minus the cumulative density function of the Shumway (2001) distance-to-default measure.

- **Stock momentum (moms)** (Gebhardt et al. 2005a) is the past 6-month cumulative stock return, while skipping the most recent month.
- **Value-at-risk (VaR)** (Bai et al. 2019) is the second-lowest corporate bond excess return during the previous 36 months (minimum 24 months).
- **Volatility beta ($\beta_{\mathbf{VIX}}$)** (Chung et al. 2019) is estimated from the monthly rolling regressions of excess bond returns on the change in the volatility index (ΔVIX) and its first lag over a 60-month window, while controlling for MKTs, SMB, HML, DEF, and TERM. β_{VIX} is the sum of of the sensitivities toward the ΔVIX and its first lag, which captures the response and lagged response, respectively, to aggregate volatility shocks. The VIX data are from the Chicago Board Options Exchange (CBOE).

Factor Construction

- **Bond market factor (MKTb)** is computed as the value-weighted (using the bonds' amount outstanding) average return of all corporate bonds in the sample minus the 1-month Treasury-bill rate.
- **Bond momentum factor (MOMb)** (Bali et al. 2017b; Jostova et al. 2013) is the difference between the average returns of the high-bond-momentum portfolios and the low-bond-momentum portfolios across the rating quintile portfolios. We form the value-weighted bivariate portfolios by independently sorting bonds into five portfolios based on their credit ratings, and five portfolios based on their bond momentum.
- **Bond volatility factor (BVL)** (Kelly et al. 2023) is the difference between the average returns of the high-bond-volatility portfolios and the low-bond-volatility portfolios across the rating quintile portfolios. We form the value-weighted bivariate portfolios by independently sorting bonds into five portfolios based on their credit ratings, and five portfolios based on their bond volatility.

- **Carry factor (CRY)** (Israel et al. 2018; Kelly et al. 2023) is the difference between the average returns of the high-carry portfolios and the low-carry portfolios across the rating quintile portfolios. We form the value-weighted bivariate portfolios by independently sorting bonds into five portfolios based on their credit ratings, and five portfolios based on their carry.
- **Credit risk factor (CRF)** (Bai et al. 2019) is the average of the credit risk factors based on the bivariate sorts with downside risk, illiquidity, and short-term reversal (CRF_{dr}, CRF_{illiq}, and CRF_{str}). In each case, the CRF factor is the difference between the average returns of the low-rating portfolios and the high-rating portfolios across the quintile portfolios based on the respective other characteristics. We take the MKTb, CRF, DRF, and LRF factors directly from Bai et al. (2019).
- **Duration factor (DUR)** (Israel et al. 2018; Kelly et al. 2023) is the difference between the average returns of the high-duration portfolios and the low-duration portfolios across the rating quintile portfolios. We form the value-weighted bivariate portfolios by independently sorting bonds into five portfolios based on their credit ratings, and five portfolios based on their duration.
- **Default factor (DEF)** (Fama & French 1993) is the difference between the return on a market portfolio of long-term corporate bonds (the composite portfolio on the corporate bond module of Ibbotson Associates) and the long-term government bond return. The data for DEF and TERM are from Amit Goyal's webpage.
- **Downside risk factor (DRF)** (Bai et al. 2019) is the difference between the average returns of the high-VaR portfolios and the low-VaR portfolios across the rating quintile portfolios.
- **Liquidity risk factor (LRF)** (Bai et al. 2019) is the difference between the average returns of the high-illiquidity portfolios and the low-illiquidity portfolios across the rating quintile portfolios.

- **Long-term reversal factor (LTR)** (Bali et al. 2017b, 2021a) is the difference between the average returns of the low-long-term-reversal portfolios and the high-long-term-reversal portfolios across the rating quintile portfolios. We form the value-weighted bivariate portfolios by independently sorting bonds into five portfolios based on their credit ratings, and five portfolios based on their long-term reversal.

- **Short-term reversal factor (STR)** (Bai et al. 2019) is the difference between the average returns of the short-term-loser portfolios and the short-term-winner portfolios across the rating quintile portfolios. We form the value-weighted bivariate portfolios by independently sorting bonds into five portfolios based on their credit ratings, and five portfolios based on their short-term reversal.

- **Stock momentum factor (MOMs)** (Israel et al. 2018) is the difference between the average returns of the high-stock-momentum portfolios and the low-stock-momentum portfolios across the rating quintile portfolios. We form the value-weighted bivariate portfolios by independently sorting bonds into five portfolios based on their credit ratings, and five portfolios based on their stock momentum.

- **Term factor (TERM)** (Fama & French 1993) is the difference between the monthly long-term government bond return (from Ibbotson Associates) and the 1-month Treasury-bill rate.

- **Uncertainty risk factors (UNC, EPU, & EPUtax)** (Bali et al. 2017a, 2021b; Tao et al. 2022; Lee 2022) is the difference between the average returns of the high-β_{UNC} portfolios and the low-β_{UNC} portfolios across the rating quintile portfolios. We form the value-weighted bivariate portfolios by independently sorting bonds into five portfolios based on their credit ratings, and five portfolios based on their uncertainty beta (β_{UNC}) estimates. For UNC, we use β^{JLN}_{UNC}, for EPU β^{EPU}_{UNC}, and for EPUtax β^{EPUtax}_{UNC}.

- **Value factor (VAL)** (Kelly et al. 2023) is the difference between the average returns of the high spread-to-D2D portfolios and the low spread-to-D2D portfolios across the rating quintile portfolios. We form the value-weighted bivariate portfolios by independently sorting bonds into five portfolios based on their credit ratings, and five portfolios based on their spread to D2D.

- **Volatility risk factor (VOL)** (Chung et al. 2019) is the difference between the average returns of the high-β_{VIX} portfolios and the low-β_{VIX} portfolios across the rating quintile portfolios. We form the value-weighted bivariate portfolios by independently sorting bonds into five portfolios based on their credit ratings, and five portfolios based on their uncertainty beta (β_{VIX}) estimates.

- **Equity factors** (Fama & French 2015; Bektić et al. 2019). In addition to the corporate bond factors above, we also consider the five factors of Fama & French (2015). These include the stock market (**MKTs**), size (**SMB**), value (**HML**), profitability (**RMW**), and investment (**CMA**) factors. We take the factors from Kenneth French's data library.[22]

B.2 Model Selection Method Implementation Details

The first term on the RHS of Equation (3.4) is

$$
\begin{aligned}
&-\frac{(K-L_j)L_j}{2}\log 2-\frac{\tilde{T}L_j}{2}\log\pi-\frac{L_j}{2}\log(\tilde{T}k_j+1)\\
&-\frac{(\tilde{T}+L_j-K)}{2}\log|\psi_j|+\log\Gamma_{L_j}\left(\frac{\tilde{T}+L_j-K}{2}\right).
\end{aligned}
$$

[22] Bektić et al. (2019) show that investment and profitability factors based on corporate bond data have some explanatory power for corporate bond returns. When using these instead of the Fama & French (2015) equity factors, the results are similar. Both factors are eliminated by the first-step identification protocol.

The second term on the RHS of Equation (3.4) is

$$\frac{(K-L_j)L_j}{2}\log 2-\frac{(K-L_j)(\tilde{T}-L_j)}{2}\log\pi$$
$$-\frac{(K-L_j)}{2}\log|W_j^*|-\frac{\tilde{T}}{2}\log|\psi_j^*|+\log\Gamma_{K-L_j}\left(\frac{\tilde{T}}{2}\right),$$

where $\tilde{T}=T-n_t$ and

$$W_j^*=\sum_{t=n_t+1}^{T}\tilde{f}_{j,t}\tilde{f}_{j,t}',$$
$$\psi_j=\sum_{t=n_t+1}^{T}(\tilde{f}_{j,t}-\hat{\tilde{\alpha}}_j)(\tilde{f}_{j,t}-\hat{\tilde{\alpha}}_j)'+\frac{\tilde{T}}{\tilde{T}k_j+1}\left(\hat{\tilde{\alpha}}_j-\tilde{\alpha}_{j0}\right)\left(\hat{\tilde{\alpha}}_j-\tilde{\alpha}_{j0}\right)'$$
$$\psi_j^*=\sum_{t=n_t+1}^{T}(f_{j,t}^*-\hat{B}_{j,f}^*\tilde{f}_{j,t})(f_{j,t}^*-\hat{B}_{j,f}^*\tilde{f}_{j,t})'.$$

$\Gamma_d(.)$ denotes the d-dimensional multivariate gamma function. All other variables are as previously defined. Hats on the parameters indicate that they are the estimates obtained by linear regressions of Equations (3.2) and (3.3).

Following the recommendation of Chib et al. (2020), we use this model along with the model-specific prior $\tilde{\alpha}_j|M_j\sim\mathcal{N}(\tilde{\alpha}_{j0},k_j\Sigma_j)$ with

$$\tilde{\alpha}_{j0}=n_t^{-1}\sum_{t=1}^{n_t}\tilde{f}_{j,t},$$

where $n_t=tr\times T$ is the size of the training sample, which we set to $tr=10\%$ of the data, as in Chib et al. (2020). The model-specific multiplier k_j can be computed as

$$k_j=\frac{1-tr}{tr}\times L_j^{-1}sum(diag(V_{j0})/diag(\hat{\Sigma}_{j0})),$$

where V_{j0} is the negative inverse Hessian over $\tilde{\alpha}_j$ and $\hat{\Sigma}_{j0}$ the estimate of the covariance matrix Σ_j in the training sample.

Chapter 4

Factor Pricing Across Asset Classes*

4.1 Introduction

Well-diversified investors should construct their portfolios from as broad a range of assets as possible. However, most academic studies typically examine different markets in isolation, developing asset-class-specific factor models for stocks, bonds, commodities, and the like. Only a few recent studies examine anomalies and other phenomena jointly across asset classes (e.g., Asness, Moskowitz, & Pedersen, 2013; Koijen et al., 2018). Under the law of one price and free portfolio formation, however, Cochrane (2009) shows that theoretically there should be a single stochastic discount factor that prices assets of all classes. Therefore, from both a theoretical and a practical perspective, it is worthwhile to search for a handful of factors that span the Markowitz (1952) mean-variance-efficient frontier and capture the returns of all assets.

However, to the best of our knowledge, important questions related to this endeavor have not yet been fully resolved, for example: To what extent can

*This chapter is based on the Working Paper "Factor Pricing Across Asset Classes" authored by Thuy Duong Dang, Fabian Hollstein and Marcel Prokopczuk, 2023.

the factors of different asset classes also price the assets of others? What is the degree of integration between the different asset classes? What does an optimal (empirical) stochastic discount factor across all asset classes look like? What is an appropriate benchmark model for portfolios of global securities across asset classes? The objectives of this chapter are therefore twofold. First, we comprehensively examine the prominent traded factors proposed in the asset pricing literature for various individual asset classes, and investigate the extent of market integration based on their explanatory power across other asset classes. Second, we attempt to identify an integrated empirical model based on a sparse number of risk factors that spans and explains returns across multiple asset classes.

To capture the widest possible range of investment opportunities, we consider a broad set of factors across seven major asset classes: U.S. equities, international equities, corporate bonds, commodities, currencies, equity indices, and government bonds. In total, we use 77 well-known empirical factors that typically enjoy consensus support from both academics and practitioners.

The first major objective of this chapter is to examine the degree of market integration. We do this through the lens of the pricing power of factor models from one asset class for others. As a first simple step, we show that among the market factors of the different asset classes, at least two are necessary to capture the risk premia of the others.

More importantly, we examine the pricing power of the best models of each individual asset class for other asset classes. To do this, we first have to identify the best models in each asset class. Rather than relying solely on existing models, we also combine the existing factors together to create new optimal models. Thus, for each individual asset class, we first identify viable risk factors among the candidates by subjecting them to the necessary condition of the factor identification protocol proposed by Pukthuanthong et al. (2019) (PRS). Then, we scan for the best model among the surviving factors by applying the Bayesian marginal-likelihood factor model selection algorithm developed by Barillas & Shanken (2018) and Chib et al. (2020) (BS–CZZ).

We find that factor models that specialize in one asset class typically have difficulty pricing the factors from other asset classes. Thus, we reject perfect integration. There appear to be multiple underlying systematic risk drivers across asset classes and markets. However, we also detect some cross-market linkages.

These findings further motivate us to pursue the second main objective of this study, which is to find an optimal integrated factor model across all asset classes. To avoid creating high-dimensional factor models, we focus on the best factors for each asset class when building the combined model by again using the BS–CZZ method. The optimal model consists of a total of eight factors, including the U.S. equity market, the size, management, and quality-minus-junk factors for international equities, the carry and equity momentum factors for corporate bonds, the currency momentum factor, and the equity index carry factor. Factors from the major asset classes (equities and corporate bonds) prove to be the most important. Furthermore, not all asset classes need to be included in the optimal model: commodity and government bond factors are completely absent from the top model and a currency factor is only included in one of the top three models. Thus, the fact that not all asset classes are needed suggests the presence of some degree of cross-market linkages.

Having constructed an optimal integrated model, the natural next step is to investigate its performance. Specifically, we analyze its superiority over existing models and its explanatory power for a comprehensive list of prominent factors and a large battery of test portfolios across the seven major asset classes. To begin, we conduct pairwise comparisons of the relative performance of the optimal unified factor model and several prominent existing models using the Barillas et al. (2020) test for equality of squared Sharpe ratios. We find that the top integrated factor model achieves a substantially higher squared Sharpe ratio than all the existing single- and multi-asset-class models we consider. Its Sharpe ratio also far exceeds those of all the optimal single-asset-class models we identify. All of these differences are highly statistically and economically significant. Furthermore, we show that the performance differences also persist in an out-of-sample Sharpe ratio analysis. Thus, the optimal integrated model clearly dominates all single-asset-class models. The findings indicate that a multi-asset, multi-factor investment approach provides substantial diversification benefits and significantly enhances the investment opportunity set available to investors.

We continue the examination of model performance by performing spanning tests of all remaining factors across asset classes on the top integrated model to identify explicit cross-market linkages. Overall, this model is able to explain an extensive list of prominent existing factors from multiple asset classes. While we reject perfect integration overall, we clearly detect some cross-market-linkages, in that the factors of one market are typically also exposed to some extent to the factors in the top integrated model from other asset classes.

Finally, we examine the explanatory power of the top integrated model on a broad set of test portfolios across asset classes. We compare this model to the existing models for each asset class, based on several ubiquitous performance measures. The optimal integrated model generally produces low Gibbons et al. (1989) GRS statistics and small numbers of significant alphas. At the same time, it can deliver time-series R^2s that are comparatively high, and average absolute alphas that are comparatively low, on a par with many existing asset-class-specific models. These results suggest that the integrated model has competitive explanatory power for the returns of each individual asset class.

Our main contribution is to provide an integrated view of asset pricing. Moving beyond the convention of studying individual asset classes separately, we uncover an optimal factor model that can span the multi-asset return space. Our findings have important implications for both academics and practitioners. Our proposed model can be a benchmark for future research in pricing securities across different asset classes, and a useful guide for investors to exploit factor-based investing through a multi-asset, multi-factor lens.

A few previous studies link risk factors in one asset class to returns in other asset classes. These typically focus on the relationships between only two asset classes. Even for stocks and bonds, which are contingent claims on the value of the same underlying assets, the exact degree of integration is still a matter of debate (e.g., Choi & Kim, 2018; Kelly et al., 2023). Other papers also examine further empirical linkages between markets, usually U.S. equities and others, such as currencies (e.g., Karolyi & Wu, 2021; Fan, Londono, & Xiao, 2022), sovereign bonds (e.g., Borri & Verdelhan, 2011), or commodities (e.g., Bakshi, Gao, & Rossi, 2019). Links between currency and fixed income markets may result from hedging activities of intermediaries (e.g., Korsaye, Trojani, & Vedolin, 2023).

However, most of the asset pricing literature has been devoted to discovering new factors to explain different asset classes in isolation. Only a few previous studies suggest that there is some common structure among asset pricing factors in different markets. For example, factors such as value and momentum (Asness et al., 2013), carry (Koijen et al., 2018), betting-against-beta (Frazzini & Pedersen, 2014), quality-minus-junk (Asness, Frazzini, & Pedersen, 2019), and time-series momentum (Moskowitz, Ooi, & Pedersen, 2012) have been shown to work across multiple asset classes.

Our work contributes to the small but growing literature on asset pricing across asset classes. Two important studies in this area are Asness et al. (2013) and Cooper, Mitrache, & Priestley (2022). The former introduce a simple global three-factor model consisting of a global market factor plus value and momentum to explain the returns of their 48 value-and-momentum-everywhere (VME) portfolios across different markets and asset classes. Cooper et al. (2022) study the pricing performance of a global five-macroeconomic-factor model for, among others, the 48 VME portfolios.[1]

Our research approach allows us to analyze market integration across asset classes much more directly than in previous studies. In particular, a broad consideration of a wide range of factors is important to comprehensively analyze the extent and limits of cross-market linkages. Moreover, previous studies that focus on a few factors or fixed models are unlikely to have come close to identifying

[1] He et al. (2017) also study an intermediary asset pricing factor across different asset classes. However, Gospodinov & Robotti (2021) argue that the intermediary factor does not provide incremental information to the market factor.

optimal factor models for multi-asset, multi-factor investors.

The remainder of this chapter is organized as follows. Section 4.2 describes the data. Sections 4.3 and 4.4 analyze market integration across asset classes. In Section 4.5, we generate and analyze an optimal integrated factor model across asset classes. Finally, Section 4.6 provides concluding remarks. The appendix to this chapter contains a robustness check and additional descriptions of the data and methods used in this chapter.

4.2 Data

4.2.1 Candidate Factors

We consider a total of 77 factor candidates from seven major asset classes, including U.S. equities, international equities (global ex-U.S.), corporate bonds, currencies, commodities, government bonds, and global equity indices.[2] Their common sample period available for analysis is from August 2006 to December 2019.[3] Note that our sample period is largely out-of-sample or post-publication for many of the factors. Thus, our results are likely to provide a realistic assessment of the risk and return that investors can achieve with these factors going forward.

In Table 4.1, we summarize the list of the candidate factors proposed in the specialized literature for each of the asset classes we consider in our analysis. We include a total of 21 U.S. equity factors, including those of Fama & French (1993, 2018), Pástor & Stambaugh (2003), Asness & Frazzini (2013), Frazzini & Pedersen (2014), Hou et al. (2015); Hou, Mo, Xue, & Zhang (2021), Stambaugh & Yuan (2017), Asness et al. (2019), and Daniel et al. (2020). We also include a

[2]Factor and portfolio data are taken directly from the authors' websites. Detailed links to the web sources are in Table C.1 of the appendix to this chapter . We calculate a small number of factors ourselves that are not directly available. Information on these is also given in the same table.

[3]These starting and ending dates of the sample period are necessary to have as broad coverage as possible across asset classes. In the appendix to this chapter, we also consider a longer sample period, while excluding some important asset classes with shorter time periods. The results are qualitatively similar.

total of 13 international equity factors from Hanauer (2020) and Jensen, Kelly, & Pedersen (2022). Our sample also includes 18 corporate bond factors from Fama & French (1993), Bali et al. (2017b, 2021b), Bai et al. (2019), Chung et al. (2019), Lee (2022), Tao et al. (2022), and Kelly et al. (2023). In addition, we have five commodity factors from Moskowitz et al. (2012) and Ilmanen, Israel, Moskowitz, Thapar, & Lee (2021), eight foreign exchange factors from Lustig, Roussanov, & Verdelhan (2011), Moskowitz et al. (2012), Verdelhan (2018), and Ilmanen et al. (2021), six equity index factors and six government bond factors, both also from Moskowitz et al. (2012) and Ilmanen et al. (2021). Thus, there are at least five candidate factors for each asset type.[4]

Table C.2 of the appendix to this chapter provides the summary statistics of the monthly returns of all candidate factors. Consistent with McLean & Pontiff (2016), we find that many of the U.S. equity factors do not yield a statistically significant average return over our sample period. Performance is stronger for other asset classes, particularly international equities, corporate bonds, and government bonds. Across asset classes, commodity factors appear to be the most volatile. Among asset classes, time-series momentum factors tend to have the highest volatilities.

Figure C.2 of the appendix to this chapter shows that there are significant correlations between different factors within and across asset classes. In particular, the highest correlations, exceeding 0.8, are among the following factor sets within asset classes: {SMB_useq, ME_useq, SMB_useqsy}, and {CMA_useq, IA_useq} for U.S. equities, {RMW_inteq, ROE_inteq} for international equities, and {MKT_cb, VOL_cb, DUR_cb} for corporate bonds. However, there are also high correlations between the factors across asset classes for the sets {MKT_useq, MKT_inteq, MKT_eqi}, {MOM_useq, MOM_inteq, MOM_EW}, and {TERM_cb, MKT_govtb}. It is therefore likely that these factors are exposed to similar sources of economic risk and are not all that different

[4]We do not include the intermediary factor of He et al. (2017) because Gospodinov & Robotti (2021) argue that it carries essentially the same information as the market factor. For our sample period, we also find that the intermediary factor can be spanned by the U.S. equity market factor, leaving an alpha that is insignificant at the 5% level.

Table 4.1: List of Candidate Factors across Asset Classes

This table provides the list of prominent factors from seven major asset classes that we use in this chapter. The list includes the abbreviations and names of the factors. The abbreviations are typically a combination of the original name of the factor and an indicator of the asset class to which it belongs after the underscore. The last column lists the authors and/or sources of each factor. The data on factors are, if available, directly from the authors' websites. If not, we construct them following the methodologies described in the corresponding papers.

#	Factor Abbr.	Factor Names	Authors/ Sources
		U.S. Equities	
1	MKT_useq	market	
2	SMB_useq	size	
3	HML_useq	value	
4	RMW_useq	operating profitability	Fama & French (1993, 2018) (FF)
5	CMA_useq	investment	
6	MOM_useq	momentum	
7	STR_useq	short-term reversal	
8	LTR_useq	long-term reversal	
9	ME_useq	size	
10	IA_useq	investment	Hou et al. (2015, 2021) (HXZ)
11	ROE_useq	return on equity	
12	EG_useq	expected growth	
13	SMB_useqsy	size	
14	MGMT_useq	mispricing - management	Stambaugh & Yuan (2017) (SY)
15	PERF_useq	mispricing - performance	
16	PEAD_useq	inattention	Daniel et al. (2020) (DHS)
17	FIN_useq	financing	
18	HMLm_useq	monthly value	Asness & Frazzini (2013)
19	BAB_useq	betting against beta	Frazzini & Pedersen (2014)
20	QMJ_useq	quality minus junk	Asness et al. (2019)
21	LIQ_useq	liquidity	Pástor & Stambaugh (2003)
		International Equities (excluding the U.S.)	
1	MKT_inteq	global market	
2	SMB_inteq	size	
3	HML_inteq	value	
4	HMLm_inteq	value devil	
5	RMW_inteq	operating profitability	
6	ROE_inteq	return on equity	Hanauer (2020)
7	CMA_inteq	investment	
8	MOM_inteq	momentum	
9	iMOM_inteq	idosyncratic momentum	
10	MGMT_inteq	mispricing - management	
11	PERF_inteq	mispricing - performance	
12	BAB_inteq	betting against beta	Jensen et al. (2022)
13	QMJ_inteq	quality minus junk	
		Corporate Bonds	
1	MKT_cb	corporate bond marker	
2	DRF_cb	downside risk	Bai et al. (2019) (BBW)
3	LRF_cb	liquidity risk	
4	CRF_cb	credit risk	
5	TERM_cb	term risk	Fama & French (1993)
6	DEF_cb	default risk	
7	Carry_cb	carry	
8	Value_cb	value	
9	DUR_cb	duration	Kelly et al. (2023) (KPP)
10	VOL_cb	bond volatility	
11	MOM_cbeq	stock momentum	
12	MOM_cb	bond momentum	
13	STR_cb	short-term reversal	Bali et al. (2017b) (BSW)
14	LTR_cb	long-term reversal	
15	VIX_cb	volatility risk	Chung et al. (2019)
16	UNC_cb	macroeconomic uncertainty risk	Bali et al. (2021b)
17	EPU_cb	economic policy uncertainty risk	Tao et al. (2022)
18	EPUtax_cb	tax policy uncertainty risk	Lee (2022)
		Commodities	
1	MKT_cm	commodity market	
2	Value_cm	commodity value	Ilmanen et al. (2021) (AQR)
3	MOM_cm	commodity momentum	
4	Carry_cm	commodity carry	
5	TSMOM_cm	commodity time-series momentum	Moskowitz et al. (2012)
		Currencies	
1	MKT_fx	Dollar risk factor	Lustig et al. (2011) (LRV)
2	HML_fx	currency carry trade risk	
3	Carry_fx	currency carry factor	Verdelhan (2018)
4	Dollar_fx	Dollar factor	
5	Value_fxaqr	currency value	
6	MOM_fxaqr	currency momentum	Ilmanen et al. (2021) (AQR)
7	Carry_fxaqr	currency carry	
8	TSMOM_fx	currency time-series momentum	Moskowitz et al. (2012)
		Equity Indices	
1	MKT_eqi	equity index market	
2	Value_eqi	equity index value	
3	MOM_eqi	equity index Momentum	Ilmanen et al. (2021) (AQR)
4	Carry_eqi	equity index carry	
5	Defensive_eqi	equity index defensive	
6	TSMOM_eqi	equity index time-series momentum	Moskowitz et al. (2012)
		Government Bonds	
1	MKT_govtb	fixed income market	
2	Value_govtb	fixed income value	
3	MOM_govtb	fixed income Momentum	Ilmanen et al. (2021) (AQR)
4	Carry_govtb	fixed income carry	
5	Defensive_govtb	fixed income defensive	
6	TSMOM_govtb	fixed income time-series momentum	Moskowitz et al. (2012)

(Gospodinov & Robotti, 2021). Therefore, when constructing optimal models, we consider only those that include at most one of the factors in these sets.

4.2.2 Existing Models

In addition to the individual factors, we also consider a list of prominent existing factor models of all asset classes for comparison, as detailed in Table 4.2. In the spirit of the Capital Asset Pricing Model (CAPM), we consider a single-factor model of the corresponding market factor for each asset class. All models for international equities are in the spirit of the models proposed for the U.S. equities, but their factors are constructed from stocks in markets other than the U.S. Some models are augmented versions of the corresponding models. The number of factors included in a model specialized for one asset class typically ranges from one to six. Most of the factor models include a market factor.

4.2.3 Test Assets

Part of the model selection and the asset pricing tests are also based on test assets. Thus, for each asset class, we also obtain portfolios that capture the cross-sectional heterogeneity in these markets. For U.S. equities, we use 207 characteristic long-short portfolios from Chen & Zimmermann (2021) and 30 industry portfolios from Kenneth French's library. For global equities (excluding the U.S.), we use 153 characteristic long-short portfolios of Jensen et al. (2022) and 125 double-sorted portfolios based on different pairs of characteristics, derived from developed and emerging markets from Kenneth French's library.

For corporate bonds, as in Chapter 3, we use the double-sorted portfolios on size and the other 23 bond characteristics from the Kelly & Pruitt (2022) dataset, 25 size-maturity and 25 rating-maturity portfolios, and 12 industry portfolios. For commodities, we consider a list of 23 commodities, using data from the Commodity Research Bureau (CRB). For each commodity, we use the nearest-to-maturity futures contracts. We roll over the contracts at the end of the

Table 4.2: List of Existing Models for Different Asset Classes

This table lists the prominent existing factor models specialized for each of the seven major asset classes, as well as a global factor model. For each asset class, there is a single-factor model of the corresponding market factor in the spirit of the CAPM. All models for international equities are in the spirit of the models proposed for U.S. equities. Some models are augmented versions of the corresponding models.

Authors	**Model**	**Factors**
	U.S. Equities	
	CAPM_useq	MKT_useq
Fama & French (1993)	FF3_useq	MKT_useq, SMB_useq, HML_useq
Carhart (1997)	C4_useq	MKT_useq, SMB_useq, HML_useq, MOM_useq
Fama & French (2015)	FF5_useq	MKT_useq, SMB_useq, HML_useq, RMW_useq, CMA_useq
Fama & French (2018)	FF6_useq	MKT_useq, SMB_useq, HML_useq, RMW_useq, CMA_useq, MOM_useq
Hou et al. (2015)	HXZ4_useq	MKT_useq, ME_useq, IA_useq, ROE_useq
Hou et al. (2021)	HXZ5_useq	MKT_useq, ME_useq, IA_useq, ROE_useq, EG_useq
Daniel et al. (2020)	DHS_useq	MKT_useq, PEAD_useq, FIN_useq
Stambaugh & Yuan (2017)	SY_useq	MKT_useq, SMB_useqsy, MGMT_useq, PERF_useq
	International Equities	
	CAPM_inteq	MKT_inteq
	FF3_inteq	MKT_inteq, SMB_inteq, HML_inteq
	C4_inteq	MKT_inteq, SMB_inteq, HML_inteq, MOM_inteq
	FF5_inteq	MKT_inteq, SMB_inteq, HML_inteq, RMW_inteq, CMA_inteq
	FF6_inteq	MKT_inteq, SMB_inteq, HML_inteq, RMW_inteq, CMA_inteq, MOM_inteq
	HXZ4_inteq	MKT_inteq, SMB_inteq, CMA_inteq, ROE_inteq
	SY_inteq	MKT_inteq, SMB_inteq, MGMT_inteq, PERF_inteq
	Corporate Bonds	
	CAPM_cb	MKT_cb
Fama & French (1993)	FF3_cb	MKT_cb, TERM_cb, DEF_cb
Bai et al. (2019)	BBW	MKT_cb, DRF_cb, CRF_cb, LRF_cb
Bali et al. (2017b)	BSW	MKT_cb, STR_cb, MOM_cb, LTR_cb
Israel et al. (2018)	IRP	Carry_cb, Value_cb, DUR_cb, MOM_cb, MOM_cbeq
Kelly et al. (2023)	KPP	MKT_cb, Carry_cb, VAL_cb, DUR_cb, VOL_cb
	aug. FF3_cb	MKT_cb, TERM_cb, DEF_cb, LRF_cb, MOM_cb
Fama & French (1993)	FF5_cb	MKT_useq, SMB_useq, HML_useq, TERM_cb, DEF_cb
	Commodities	
	CAPM_cm	MKT_cm
Bakshi et al. (2019)	BGR_cm	MKT_cm, MOM_cm, Carry_cm
Ilmanen et al. (2021)	AQR_cm	MKT_cm, Value_cm, MOM_cm, Carry_cm
	Currencies	
	CAPM_fx	MKT_fx
Ilmanen et al. (2021)	AQR_fx	Value_fxaqr, MOM_fxaqr, Carry_fxaqr
Lustig et al. (2011)	LRV2011	MKT_fx, HML_fx
Verdelhan (2018)	Verdelhan2018	Carry_fx, Dollar_fx
	Equity Indices	
	CAPM_eqi	MKT_eqi
Ilmanen et al. (2021)	AQR_eqi	MKT_eqi, Value_eqi, MOM_eqi, Carry_eqi, Defensive_eqi
	Government Bonds	
	CAPM_govtb	MKT_govtb
Ilmanen et al. (2021)	AQR_govtb	MKT_govtb, Value_govtb, MOM_govtb, Carry_govtb, Defensive_govtb
	Across Asset Classes	
Asness et al. (2013)	AMP_across	MKT_global, VAL_EW, MOM_EW

month that is two months before expiration. We also add six value and momentum commodity portfolios from Asness et al. (2013) to the pool of commodity test assets. For currencies, we combine two sets of currency portfolios to arrive at a total of twelve portfolios. The first set of six currency portfolios is sorted on the interest rate from Lustig et al. (2011). The second set contains three sorted portfolios (low, medium, and high) for each value and momentum from Asness et al. (2013).

For global government bonds, we use returns on 23 developed and emerging market government bond indices from Refinitiv Datastream, as in Zaremba & Czapkiewicz (2017). Similarly, we use returns on 43 global equity indices from Refinitiv Datastream. We also include 6 value and momentum portfolios for each of these two asset classes from Asness et al. (2013).

4.3 Market Integration at the Aggregate Level

We begin with a simple analysis of asset class integration, based only on the respective market factors. The empirical asset pricing literature centers largely on the U.S. stock market. Thus, we examine the extent to which U.S. stock market excess returns can explain market excess returns in other asset classes. We present the results of spanning regressions of these market factors on the U.S. stock market factor in Panel A of Table 4.3. If the market factors of other asset classes have significant exposures to the U.S. stock market, it suggests that the stock market can explain (time-series) variation in these returns, and that both are driven, to some extent, by similar economic forces. However, the more important part of the spanning regressions is the alpha. If the market factors of other asset classes have significant positive alphas relative to the U.S. stock market, this would suggest that an additional factor is needed to explain their average return. On the other hand, if a significant average excess return turns into an insignificant alpha, it suggests that the U.S. stock market spans the market factor of the other asset class.

First, we find that all asset class market factors have significant exposures to the U.S. stock market. Quite naturally, the exposures are largest for international equities and equity indices, where the U.S. stock market alone can explain 71.2% and 78.6% of the variation, respectively. For other markets, however, both the slope coefficients and the R^2s are lower. The U.S. stock market can explain 36.1%, and 26.1% of the variation in the returns of the foreign exchange, and commodity market returns, respectively. Finally, the stock market explains only 10.9% and 8.01% of the variation in corporate and government bond market returns, respectively.[5]

The corporate and government bond market factors are the most interesting for this analysis because, unlike most others, they generate significant positive average excess returns over our sample period. Indeed, we find that these two also generate significant positive alphas relative to the U.S. stock market factor. Thus, the equity market alone is not sufficient to span the full set of market factors, suggesting that there are multiple underlying systematic risk drivers across asset classes and markets.

Next, we add another market factor: that of government bonds.[6] We present the results in Panel B of Table 4.3. Indeed, the combination of the two market factors can span the monthly market returns of the five remaining asset classes. So there is some preliminary evidence of integration: while one factor alone is not enough, just two different market factors can explain the market factors across asset classes quite well.

[5] These results are consistent with the correlations between the market factors, as shown in Figure C.2 of the appendix to this chapter. There are also relatively high and positive correlations between the foreign exchange market and international equities (almost 80%), as well as between currencies and commodities (more than 70%). The government bond market, on the other hand, has quite low and in most cases negative correlations with the remaining asset classes, except for a moderate correlation with the corporate bond market. The corporate bond market is moderately correlated with the other six markets (correlations between 26% and 46%).

[6] We do not do this entirely randomly. The two-factor model with equity and government bond market factors is also the one selected from the candidate set of different market factors by the BS–CZZ model scan approach. We leave the details of this approach for later in the chapter and simply use this two-factor model to get some initial insights.

Table 4.3: Spanning Regressions with Market Factors of Different Asset Classes

This table reports the results of spanning regressions of the market factors of the other asset classes on the U.S. equity market factor (Panel A) and a combination of the U.S. equity and government bond market factors (Panel B). We present the intercept from the spanning regressions (α) as well as the loadings of the excluded factors on the respective model factors from August 2003 to December 2019. The t-statistics in parentheses are based on robust Newey & West (1987) standard errors with four lags. GRS reports the results of the Gibbons et al. (1989) test of the null hypothesis that all alphas are jointly zero. Below the GRS test statistics in brackets are the corresponding p-values. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A:

	α	MKT_useq	R^2
MKT_inteq	−0.35	0.98***	71.2
	(−1.36)	(15.8)	
MKT_cb	0.31***	0.11***	10.9
	(2.67)	(3.34)	
MKT_govtb	0.39***	−0.09***	8.01
	(3.91)	(−3.36)	
MKT_eqi	−0.19	0.85***	78.6
	(−1.12)	(16.2)	
MKT_cm	−0.51	0.53***	26.1
	(−1.38)	(4.68)	
MKT_fx	−0.17	0.26***	36.1
	(−1.31)	(7.41)	
GRS	3.62***		
	[0.00]		

Panel B:

	α	MKT_useq	MKT_govtb	R^2
MKT_inteq	−0.29	0.97***	−0.13	71.3
	(−1.04)	(14.9)	(−0.64)	
MKT_cb	0.08	0.16***	0.58***	40.2
	(0.82)	(5.46)	(6.42)	
MKT_eqi	−0.17	0.84***	−0.05	78.6
	(−0.88)	(15.2)	(−0.35)	
MKT_cm	−0.25	0.48***	−0.65***	29.5
	(−0.67)	(3.87)	(−2.79)	
MKT_fx	−0.18	0.27***	0.02	36.1
	(−1.30)	(6.85)	(0.18)	
GRS	1.42			
	[0.22]			

However, this simple broad market analysis is unlikely to provide the full picture. It has been widely documented that the market factors are not sufficient to price even their own asset classes. Therefore, in the next section, we conduct a more comprehensive analysis using a large collection of other empirical factors across major asset classes.

4.4 A Second Look at Market Integration with Asset-Class-Specific Optimal Models

4.4.1 Factor Identification Results

Before we go any further, though, we need to isolate the viable risk factors of the different asset classes. To do this, we first subject all factors to the factor identification protocol of Pukthuanthong et al. (2019). The key point of this analysis is that for a factor candidate to be considered a genuine risk factor, it must be significantly associated with the return covariance matrix of its asset class. After eliminating non-viable factors, we set out to select optimal factor models that best represent each asset class. Chapter 3 use a similar two-step approach for corporate bond factors.

For the factor identification step, for each asset class, we extract the first ten principal components from the universe of test portfolios using the method of Connor & Korajczyk (1988). The set of test assets should be big enough to accurately capture the cross-sectional heterogeneity within each asset class. Therefore, we use the largest possible sets of test assets for each asset class, as described in Section 4.2.3. To account for possible non-stationarity, we split our sample in two and run the analysis separately for each sub-period (Pukthuanthong et al. 2019). Next, we compute the canonical correlations between the candidate factors and these ten principal components. Finally, we regress each of the canonical variates on a constant and the set of candidate factors. As in Pukthuanthong et al. (2019), for an eligible factor we require that the average of the absolute t-statistics associated with the significant canonical correlations exceeds 1.96 and the average number of significant t-statistics over the two periods is more than 25% of the number of canonical variates (Pukthuanthong et al., 2019).

In Panel A of Table 4.4, we summarize the results of factor candidates that pass the necessary condition of the factor identification for each asset class. Table 3.3 of the appendix to this chapter shows the detailed results. As one might expect, almost all market factors pass the necessary condition to be considered as a risk factor. Only for the equity indices does the market factor fail. Furthermore, value, betting-against-beta, quality-minus-junk, and carry factors are often selected as viable risk factors in multiple asset classes.

Any factors that do not pass this initial test appear not to move prices in their respective asset classes, and can be discarded as viable risk factors. Therefore, in the following steps, we will only consider candidate factors that pass this factor identification protocol.

4.4.2 Asset-Class-Specific Model Selection Results

Among the factors that pass the necessary condition of the first-step factor identification protocol, for each asset class, we next aim to find the factor combination that can best explain the returns in its own asset class. That is, we perform the second model selection step using the approach of Barillas & Shanken (2018) and Chib et al. (2020) to identify the best set of factors for each asset class. We describe the approach in detail in Section B.2 of the appendix to this chapter. The best set of factors or top model is identified as the one with the highest posterior model probability.

Panel B of Table 4.4 summarizes the results for each asset class. Detailed results are in Table C.4 of the appendix to this chapter. Although market factors are considered viable risk factors for almost all asset classes, they do not seem to

Table 4.4: Summary Results of Factor Identification and Model Scan for each Asset Class

This table summarizes the results of the first-step factor identification protocol and the second-step model selection separately for each asset class. In Panel A, we document the factors that satisfy the necessary condition of the Pukthuanthong et al. (2019) factor protocol for each of the seven asset classes we study. Panel B reports the best model for each asset class in terms of its posterior probability according to the BS–CZZ model scan approach.

Panel A	**Summary list of factors passing the necessary condition of the PRS protocol**
Asset classes	**Factors passing the protocol**
U.S. EQ	MKT_useq, MOM_useq, HMLm_useq, BAB_useq, QMJ_useq
Int. EQ	MKT_inteq, SMB_inteq, CMA_inteq, MGMT_inteq, BAB_inteq, QMJ_inteq
Corporate Bonds	MKT_cb, TERM_cb, DEF_cb, Carry_cb, DUR_cb, VOL_cb, MOM_cb, MOM_cbeq, LTR_cb, VIX_cb, UNC_cb
Commodities	MKT_cm, Value_cm, Carry_cm, TSMOM_cm
FX	MKT_fx, MOM_fxaqr, Carry_fxaqr
EQ indices	Carry_eqi, Defensive_eqi
Govt. Bond indices	MKT_govtb, Carry_govtb

Panel B	**Summary of BS–CZZ model scan results**
Asset classes	**Top models**
U.S. EQ	MKT_useq, QMJ_useq
Int. EQ	MKT_inteq, SMB_inteq, MGMT_inteq, QMJ_inteq
Corporate Bonds	TERM_cb, Carry_cb, DUR_cb, MOM_cbeq
Commodities	Value_cm, Carry_cm
FX	MOM_fxaqr
EQ indices	Carry_eqi
Govt. Bond indices	MKT_govtb

be crucial for explaining the returns of corporate bonds, commodities, currencies, and equity indices. For U.S. equities, the overall top model selected consists of the market and the quality-minus-junk factor.[7] For international equities, a four-factor model is selected with the market and size, management, and quality-minus-junk factors. For corporate bonds, consistent with Chapter 3, the optimal model includes the term-structure, carry, duration, and equity momentum factors. For commodities, the optimal factor model includes a value factor and a carry factor. The best currency factor model includes only one momentum factor, the best equity index model includes only one carry factor, and the best government bond model includes only one market factor. Thus, both carry and quality-minus-junk factors appear to be important across asset classes, consistent with Koijen et al. (2018) and Asness et al. (2019).

4.4.3 Implications for Market Integration

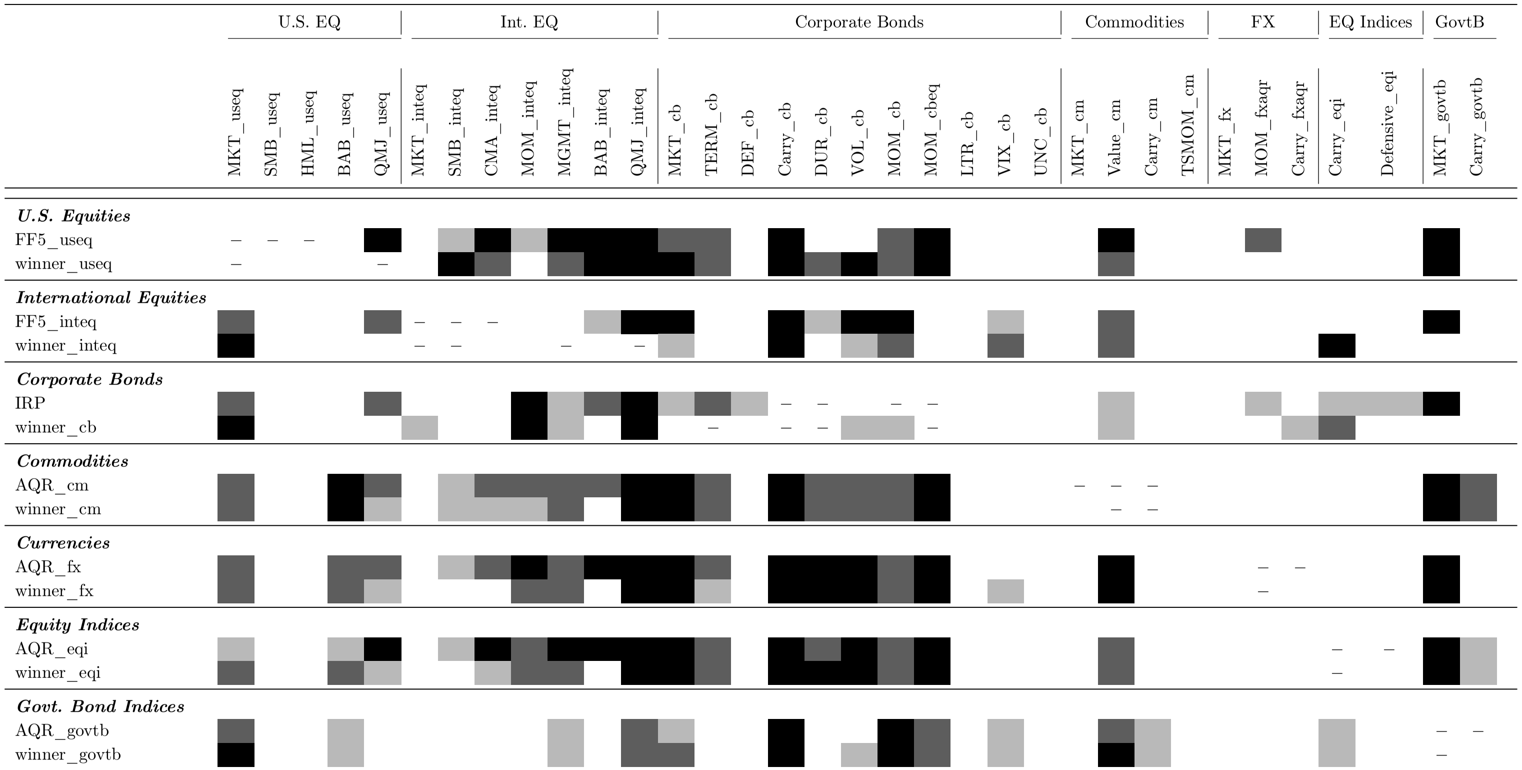
If markets are fully integrated, factor models for one asset class should be able to explain the returns of factors from other asset classes. Having selected the optimal factor model for each asset class, we next examine the ability of factor models specialized for one asset class to price the other asset classes. For this analysis, we use all factors from the different asset classes that can be considered as viable risk factors based on the first-step screening. We perform spanning regressions of these factors on the leading existing models and the best set of factors selected for each asset class. Table 4.5 shows the significance levels of the alphas from these spanning regressions.

Typically, the asset-class-specific models do a relatively good job of capturing other factors of their own asset class. For example, the optimal selected models leave no significant alphas for U.S. and international equities, commodities, currencies, equity indices, and government bonds. Only for corporate bonds do

[7] These results differ from those of Barillas & Shanken (2018). There are three main reasons for the difference. First, our sample period is very different from theirs (1972–2015). Second, they do not include the quality-minus-junk factor. Third, their results are difficult to interpret because Chib et al. (2020) show that the prior used in the original Barillas & Shanken (2018) method is unsound for model comparisons.

Table 4.5: Explaining Viable Factors in One Asset Class with Specialized Models of Other Asset Classes

This heatmap table summarizes the significance levels of the alphas from the spanning regressions of viable factors in different asset classes (i.e., those factors that pass the necessary condition of the PKR factor identification protocol for each asset class) on the representative existing models as well as models selected by the BS–CZZ scan approach specialized for each asset class. ■, ■, ■ indicate significance at the 1%, 5%, and 10% levels, respectively. White space indicates insignificance. Significance tests use robust Newey & West (1987) standard errors with four lags. – indicates that a spanning regression is not possible because the factor is already included in the model.



two factors have weakly significant alphas. Thus, reassuringly, the model selection within asset classes seems to work quite well.

More interesting, however, is how the models perform for factors in other asset classes. Again, the story is different. Typically, the models fail to explain most of the factors in other asset classes that have significant average returns. For example, the equity factor models (both the optimal selected model and the Fama & French, 2015 model) already fail to price almost all international equity factors. This finding is consistent with the results of Hollstein (2022), who shows that local factors explain anomaly returns far better than global factors. The situation is not much better for other asset classes. For corporate bonds, commodities, and government bonds, both U.S. equity factor models also leave significant alphas.

The results for the models of other asset classes are similar. It is interesting to note that the best international equity factor models do a little better at explaining U.S. equity factors than the other way around. However, they also fail for many corporate bond factors, one commodity factor, and equity index or government bond factors. The corporate bond, commodity, currency, equity index, and government bond factor models also perform rather poorly overall in explaining each other's factors.

However, there are also cases where factors with significant excess returns can be explained by models from other asset classes. For example, international equity and corporate bond factor models can explain the U.S. equity betting-against-beta factor. International equity and government bond factor models can explain the term-structure factor. Many models can explain the government bond carry factor, and government bond factor models, in turn, do comparatively well in explaining international equity and corporate bond factors.

Therefore, we clearly reject perfect integration between asset classes. There is strong evidence of multiple underlying systematic drivers across asset classes and markets. On the other hand, there appear to be some interdependencies across asset classes. Taken together, these results call for an integrated factor model across asset classes.

4.5 A Unified Model Across Asset Classes

4.5.1 Optimal Model Selection

To build an optimal integrated model that can characterize returns across asset classes, we pool all the factors in the best models from the seven asset classes together and perform the Bayesian model selection among them. We report the results for the top three models in Table 4.6.

Panel A reports the model selection results. Consistent with the results of the previous section, we find that the top models contain factors from different asset classes. Only commodity factors are completely absent from the top three models. Furthermore, a currency or government bond factor only makes it into one of the top three models. All other asset classes are represented in each of the top models. The most influential asset classes appear to be international equities and corporate bonds, with two to three factors in each of the top three models.

The most important single factors seem to be MKT_useq, SMB_inteq, MGMT_inteq, QMJ_inteq, Carry_cb, MOM_cbeq, and Carry_eqi. These seven factors are included in each of the three top models. The overall best model is an eight-factor model that also includes the MOM_fxaqr factor. The second-best model according to the posterior model probability contains the seven most important single factors plus DUR_cb and MKT_govtb, while the third-best model consists only of these.

Panel B shows the cumulative posterior probabilities of each of the factors selected to best represent their respective asset classes. We find that the factors MKT_useq, SMB_inteq, MGMT_inteq, QMJ_inteq, and Carry_cb have cumulative posterior probabilities close to 100%. Thus, these factors should be included in any decent integrated pricing model across asset classes. Among the remaining top factors, MOM_cbeq has a cumulative posterior probability of 87.50% and Carry_eqi has one of 63.89%. The additional factor in the top model, MOM_fxaqr, has a posterior probability of 49.48%. The factors that

Table 4.6: Summary Results for the Top Integrated Models

This table summarizes the model selection across asset classes, the cumulative posterior probabilities of the top factors, and the spanning regressions of viable factors in different asset classes on the integrated models. In Panel A, we report the top three integrated factor models in terms of their posterior probabilities according to the BS–CZZ model scan approach. Eligible factors are those from the top models of their asset classes as reported in the Table 4.4. Panel B shows the cumulative posterior probabilities of each of these factors. In Panel C, a heatmap summarizes the significance levels of the alphas from the spanning regressions of viable factors in different asset classes on the integrated models across asset classes. ■, ■, ■ indicate significance at the 1%, 5%, and 10% levels, respectively. White space indicates insignificance. Significance tests use robust Newey & West (1987) standard errors with four lags. − indicates that a spanning regression is not possible because the factor is already included in the model.

Panel A: Model selection across asset classes

winner1_across	MKT_useq, SMB_inteq, MGMT_inteq, QMJ_inteq, Carry_cb, MOM_cbeq, MOM_fxaqr, Carry_eqi
winner2_across	MKT_useq, SMB_inteq, MGMT_inteq, QMJ_inteq, Carry_cb, DUR_cb, MOM_cbeq, Carry_eqi, MKT_govtb
winner3_across	MKT_useq, SMB_inteq, MGMT_inteq, QMJ_inteq, Carry_cb, MOM_cbeq, Carry_eqi

Panel B: Cumulative posterior probabilities of the factors

	MKT_useq	QMJ_useq	MKT_inteq	SMB_inteq	MGMT_inteq	QMJ_inteq	TERM_cb	Carry_cb
Cumulative posterior probability (%)	99.97	43.40	0.025	99.40	99.95	99.98	13.97	100.0
	DUR_cb	MOM_cbeq	Value_cm	Carry_cm	MOM_fxaqr	Carry_eqi	MKT_govtb	
Cumulative posterior probability (%)	51.69	87.50	36.49	19.97	49.48	63.89	33.46	

Panel C: Explaining viable factors with integrated asset pricing models

	U.S. EQ					Int. EQ							Corporate Bonds											Commodities				FX			EQ Indices		GovtB	
	MKT_useq	SMB_useq	HML_useq	BAB_useq	QMJ_useq	MKT_inteq	SMB_inteq	CMA_inteq	MOM_inteq	MGMT_inteq	BAB_inteq	QMJ_inteq	MKT_cb	TERM_cb	DEF_cb	Carry_cb	DUR_cb	VOL_cb	MOM_cb	MOM_cbeq	LTR_cb	VIX_cb	UNC_cb	MKT_cm	Value_cm	Carry_cm	TSMOM_cm	MKT_fx	MOM_fxaqr	Carry_fxaqr	Carry_eqi	Defensive_eqi	MKT_govtb	Carry_govtb
AMP_across																																		
winner1_across	−						−			−		−				−				−									−		−			
winner2_across	−						−			−		−				−	−			−											−		−	
winner3_across	−						−			−		−				−				−											−			

appear in the second-best model have cumulative posterior probabilities of 51.69% (DUR_cb) and 33.46% (MKT_govtb). The other factors that do not make it into any of the top models have cumulative posterior probabilities ranging from 0.03% (MKT_inteq) to 43.40% (QMJ_useq). Thus, while the posterior probabilities of the individual top models may not be overwhelmingly large, the cumulative posterior probabilities of the factors clearly separate the wheat from the chaff.

Having established which are the best factor models across asset classes, it is natural to ask how well they perform. In what follows, we thus continue our empirical analysis by examining the pricing performance of the top models. First, we compare their Sharpe ratios with those of several existing models. Then, in order to gain insights from different angles, we examine the explanatory power of the optimal integrated model using spanning tests and a large battery of test portfolios across asset classes. These tests also provide further insight into the extent of cross-market linkages.

4.5.2 Model Sharpe Ratios

First, we analyze the mean-variance frontiers achievable by the models. Then, we perform pairwise tests of equality of squared Sharpe ratios following Barillas et al. (2020) to analyze whether the selected top models outperform other factor models. The latter method allows us to make reliable inferences when comparing relative model performance as measured by squared Sharpe ratio improvements.

Figure 4.1 visualizes the mean-variance frontiers of different models. We show existing representatives of each asset class as well as the top integrated model across all asset classes. The efficient frontier of the top integrated model is the furthest to the northwest, suggesting that investors can improve their optimal portfolios by implementing multi-asset, multi-style (factor) strategies. In other words, factors from other asset classes can expand the investment opportunity set even for multi-style, single-asset-class investors by adding variance hedges and diversification. This finding is consistent with and supports the model selection results.

Figure 4.1: Efficient Frontiers

This figure plots the efficient frontiers and the tangency portfolios (red dots) for the representative models from each asset class: FF5_useq (U.S. equities), FF5_inteq (international equities, excluding the U.S.), IRP_cb (corporate bonds), AQR_cm (commodities), AQR_fx (currencies), AQR_eqi (equity indices), and AQR_govtb (government bond indices). In addition, we also plot the optimal integrated model based on eight selected factors across asset classes though the two-step approach {MKT_useq SMB_inteq MGMT_inteq QMJ_inteq Carry_eqi Carry_cb MOM_cbeq MOM_fxaqr} as a multi-asset, multi-factor investment strategy. To construct the efficient frontier, we require that all weights are positive and sum to one. The analysis is performed over the sample period from August 2006 to December 2019.

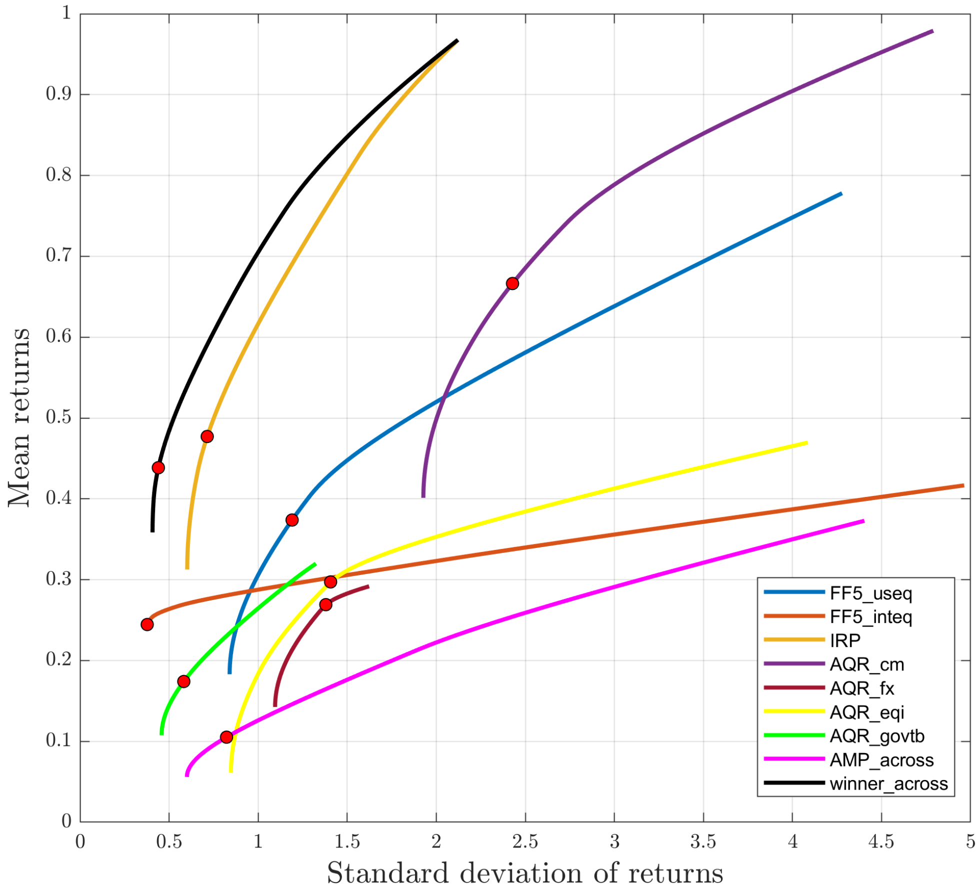


Table 4.7 reports the differences between the sample squared Sharpe ratios (column model minus row model) of different model pairs, including the best existing models (in terms of their full-sample Sharpe ratio, see Table 4.8) and the top generated models from each asset class, the Asness et al. (2013) (AMP) global model, and the top three integrated models selected from the model scan. The estimated model squared Sharpe ratios are modified to be unbiased in small samples. The associated p-values are shown in brackets.

There are some differences in the squared Sharpe ratios across asset classes. For example, the squared Sharpe ratio differences are significantly negative when comparing rows (1) to (6) and columns (7) to (14). Thus, investors in U.S. equities, international equities (ex-U.S.), and corporate bonds obtain a better unconditional risk–return tradeoff than investors limited to commodities, currencies, equity indices, and government bonds.

The most important results, however, relate to the comparison of the integrated models across asset classes with those that operate only in the individual asset classes. The three last columns of Table 4.7 clearly show that the top three models of the model selection approach dominate all other existing models across asset classes, producing higher Sharpe ratios. All of these Sharpe ratio differences are highly statistically significant, as indicated by the corresponding p-values, which are virtually zero in all cases. However, by including an additional factor, there is a slight improvement in the squared Sharpe ratios of the second top model over the first.

We also show the full-sample Sharpe ratios of the models as well as the in- and out-of-sample Sharpe ratios for two different sample splitting schemes using the first half or two-thirds of the sample to determine the weights in the tangency portfolio, as in Barillas & Shanken (2018). We report the results of the top model and existing models of each asset class, as well as the top models across asset classes and the AMP global three-factor model in Table 4.8.

We find that the two-step selection approach generally does a pretty good job of selecting the best factors and models for each individual asset class. The

Table 4.7: Tests of the Equality of Squared Sharpe Ratios

This table presents pairwise tests of the equality of the squared Sharpe ratios of the top existing models that have the highest sample squared Sharpe ratio among those within their own asset class (see Table 4.8), the top models selected by the BS–CZZ scanning approach that have the highest posterior probability within each asset class (see Panel B of Table 4.4), and the three winning models across asset classes (see Panel A of Table 4.6). The body of the table reports the differences between the monthly (bias-adjusted) sample squared Sharpe ratios of the models in column i and row j, $\hat{\theta}_i^2 - \hat{\theta}_j^2$. The column models are indicated by numbers, as defined in the first column with the model names. In brackets, we report the corresponding p-values for testing the null hypothesis $H_0 : \hat{\theta}_i^2 = \hat{\theta}_j^2$. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	winner1_across
(1) HXZ5_useq	0.013***	0.281***	0.270***	0.345***	0.380***	−0.095***	−0.085***	−0.124***	−0.147***	−0.119***	−0.136***	−0.090***	−0.093***	−0.145***	0.813***	0.917***	0.845***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.002]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
(2) winner_useq		0.268***	0.256***	0.331***	0.367***	−0.109***	−0.098***	−0.138***	−0.161***	−0.133***	−0.149***	−0.103***	−0.107***	−0.158***	0.799***	0.904***	0.832***
		[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
(3) HXZ4_inteq			−0.012***	0.064***	0.099***	−0.377***	−0.366***	−0.406***	−0.428***	−0.401***	−0.417***	−0.371***	−0.374***	−0.426***	0.532***	0.636***	0.564***
			[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
(4) winner_inteq				0.075***	0.111***	−0.365***	−0.354***	−0.394***	−0.417***	−0.389***	−0.405***	−0.359***	−0.363***	−0.414***	0.543***	0.648***	0.576***
				[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
(5) IRP					0.036***	−0.440***	−0.429***	−0.469***	−0.492***	−0.464***	−0.481***	−0.434***	−0.438***	−0.489***	0.468***	0.573***	0.501***
					[0.001]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
(6) winner_cb						−0.476***	−0.465***	−0.505***	−0.528***	−0.500***	−0.516***	−0.470***	−0.473***	−0.525***	0.432***	0.537***	0.465***
						[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
(7) AQR_cm							0.011	−0.029**	−0.052**	−0.024**	−0.040**	0.006***	0.002***	−0.049**	0.908***	1.013***	0.941***
							[0.843]	[0.018]	[0.025]	[0.032]	[0.021]	[0.003]	[0.000]	[0.020]	[0.000]	[0.000]	[0.000]
(8) winner_cm								−0.040***	−0.063***	−0.035**	−0.051***	−0.005***	−0.009***	−0.060***	0.897***	1.002***	0.930***
								[0.004]	[0.006]	[0.013]	[0.006]	[0.001]	[0.000]	[0.003]	[0.000]	[0.000]	[0.000]
(9) AQR_fx									−0.023**	0.005*	−0.011*	0.035***	0.031***	−0.020	0.937***	1.042***	0.970***
									[0.049]	[0.078]	[0.095]	[0.006]	[0.002]	[0.101]	[0.000]	[0.000]	[0.000]
(10) winner_fx										0.028	0.011	0.058**	0.054***	0.003	0.960***	1.065***	0.993***
										[0.104]	[0.251]	[0.026]	[0.009]	[0.448]	[0.000]	[0.000]	[0.000]
(11) AQR_eqi											−0.016	0.030**	0.026***	−0.025	0.932***	1.037***	0.965***
											[0.183]	[0.013]	[0.003]	[0.110]	[0.000]	[0.000]	[0.000]
(12) winner_eqi												0.046**	0.043***	−0.009	0.948***	1.053***	0.981***
												[0.010]	[0.002]	[0.120]	[0.000]	[0.000]	[0.000]
(13) AQR_govtb													−0.004	−0.055**	0.902***	1.007***	0.935***
													[0.289]	[0.019]	[0.000]	[0.000]	[0.000]
(14) winner_govtb														−0.051***	0.906***	1.011***	0.938***
														[0.003]	[0.000]	[0.000]	[0.000]
(15) AMP_across															0.957***	1.062***	0.990***
															[0.000]	[0.000]	[0.000]
(16) winner3_across																0.105***	0.033*
																[0.006]	[0.076]
(17) winner2_across																	−0.072**
																	[0.011]

Table 4.8: Out-of-Sample Sharpe Ratios

This table shows the in- and out-of-sample performance of the existing factor models (see the list in Table 4.2) and the winning models from the model scan (see Panel B of Table 4.4 and Panel A of Table 4.6) within the same asset class and across all asset classes. The first column shows the full-sample monthly Sharpe ratio of the tangency portfolios implied by the models. The remainder of the table shows the results for out-of-sample tests where the initial estimation period for the factor weights in the tangency portfolio is half the sample period ($T/2$) or two-thirds the sample period ($2T/3$). In each case, EST shows the in-sample Sharpe ratio of the estimation period, PERF shows the in-sample Sharpe ratio of the remaining period, and PERFw shows the actual out-of-sample Sharpe ratio based on the weights from the first in-sample estimation period.

Model	Sample SR	T/2			2T/3		
		EST	PERF	PERFw	EST	PERF	PERFw
winner_useq	**0.418**	**0.346**	**0.488**	**0.486**	**0.385**	**0.481**	**0.481**
CAPM_useq	0.182	0.104	0.310	0.310	0.152	0.261	0.261
FF3_useq	0.240	0.186	0.358	0.215	0.227	0.335	0.226
C4_useq	0.240	0.188	0.380	0.205	0.227	0.342	0.226
FF5_useq	0.360	0.487	0.365	0.174	0.389	0.361	0.285
FF6_useq	0.360	0.507	0.389	0.140	0.392	0.377	0.272
HXZ4_useq	0.293	0.246	0.441	0.193	0.252	0.420	0.341
HXZ5_useq	0.428	0.381	0.525	0.437	0.397	0.554	0.477
DHS_useq	0.313	0.255	0.442	0.298	0.264	0.432	0.393
SY_useq	0.368	0.344	0.432	0.366	0.341	0.471	0.404
winner_inteq	**0.675**	**0.789**	**0.644**	**0.542**	**0.834**	**0.604**	**0.365**
CAPM_inteq	0.084	0.074	0.108	0.108	0.098	0.048	0.048
FF3_inteq	0.216	0.213	0.225	0.214	0.282	0.086	0.081
C4_inteq	0.336	0.340	0.352	0.322	0.405	0.247	0.183
FF5_inteq	0.650	0.816	0.547	0.461	0.831	0.496	0.306
FF6_inteq	0.652	0.820	0.590	0.442	0.831	0.506	0.306
HXZ4_inteq	0.684	0.841	0.575	0.512	0.835	0.553	0.389
SY_inteq	0.552	0.561	0.586	0.526	0.651	0.485	0.332
winner_cb	**0.756**	**0.977**	**0.819**	**0.503**	**0.811**	**0.881**	**0.615**
CAPM_cb	0.288	0.312	0.271	0.271	0.273	0.356	0.356
FF3_cb	0.343	0.422	0.303	0.237	0.360	0.371	0.305
augFF3_cb	0.434	0.542	0.377	0.300	0.488	0.399	0.310
FF5_cb	0.309	0.320	0.396	0.223	0.324	0.412	0.243
BBW	0.401	0.469	0.351	0.325	0.428	0.444	0.347
BSW	0.434	0.676	0.347	0.153	0.541	0.388	0.157
IRP	0.738	0.922	0.814	0.448	0.782	0.932	0.568
KPP	0.634	0.927	0.693	0.360	0.767	0.767	0.379
winner_cm	**0.273**	**0.291**	**0.321**	**0.187**	**0.249**	**0.442**	**0.262**
CAPM_cm	0.021	0.030	0.120	-0.120	0.001	0.095	0.095
BGR_cm	0.135	0.209	0.213	0.027	0.140	0.270	0.073
AQR4_cm	0.277	0.328	0.441	0.169	0.265	0.556	0.222
winner_fx	**0.060**	**0.027**	**0.112**	**0.112**	**0.026**	**0.165**	**0.165**
CAPM_fx	0.019	0.104	0.117	-0.117	0.032	0.020	-0.020
AQR_fx	0.200	0.192	0.265	0.212	0.162	0.381	0.346
LRV2011	0.138	0.147	0.177	0.090	0.135	0.144	0.144
Verdelhan2018	0.149	0.175	0.245	0.077	0.135	0.191	0.175
winner_eqi	**0.124**	**0.107**	**0.463**	**-0.463**	**0.005**	**0.466**	**-0.466**
CAPM_eqi	0.115	0.043	0.248	0.248	0.092	0.188	0.188
AQR_eqi	0.242	0.209	0.556	-0.050	0.220	0.523	0.127
winner_govtb	**0.242**	**0.246**	**0.240**	**0.240**	**0.230**	**0.281**	**0.281**
CAPM_govtb	0.242	0.246	0.240	0.240	0.230	0.281	0.281
AQR_govtb	0.300	0.265	0.403	0.310	0.320	0.447	0.151
winner1_across	**1.053**	**1.268**	**1.059**	**0.817**	**1.232**	**1.219**	**0.750**
winner2_across	1.095	1.333	1.171	0.837	1.248	1.251	0.801
winner3_across	1.030	1.236	1.016	0.801	1.192	1.117	0.736
AMP_across	0.128	0.140	0.194	0.068	0.159	0.160	0.036

top models within an asset class generally produce in-sample and out-of-sample Sharpe ratios that are among the highest in their asset class. Note that the Sharpe ratios are not always higher than those of existing models, because the existing models may include factors that perform well, but have been eliminated as non-viable risk factors in the first step of our selection scheme.[8] Importantly, the top models for each asset class typically achieve superior performance while containing fewer or the same number of factors as the existing models.

Most importantly, the top integrated models across all asset classes are clearly superior to all others. Not only do they have higher in-sample Sharpe ratios, but they also have higher out-of-sample Sharpe ratios than all the top and existing models in each asset class. For example, the top model from U.S. equities has an out-of-sample Sharpe ratio of 0.486 for the second half of the sample period. The top models for international equities and corporate bonds perform similarly, with out-of-sample Sharpe ratios of 0.542 and 0.503, respectively. For all other asset classes, the out-of-sample Sharpe ratios are also considerably lower. The top three top models across asset classes, on the other hand, achieve out-of-sample Sharpe ratios that are at least 48% higher, all exceeding 0.8. The top models across asset classes also clearly outperform the only existing integrated competitor, the AMP global factor model.

4.5.3 Spanning Tests

In this section, we perform spanning tests to answer the following questions: How well does the top integrated model explain other factors across different asset classes? Which of the factors from which asset classes are most important? Is there any further evidence of cross-market linkages among the factors? Can factors from one asset class explain time-series variation in factors from other asset classes?

[8] This is the case for U.S. equities, for example. The HXZ5_useq model performs slightly better than the top model for U.S. equities based on the in-sample Sharpe ratio. However, all of the model's factors are eliminated by the factor protocol step because they are not sufficiently related to the covariance matrix of U.S. equity returns. For example the EG_useq factor, which has a t-statistic of the mean return of 2.58, is included in the HXZ5_useq model, but is not eligible for the top U.S. equity model.

Specifically, we regress all the factors from all asset classes that are not selected on the top integrated model.

We present the results in Table 4.9. In Panel A shows the results for U.S. equity factors. These generally load most strongly on MKT_useq, MGMT_inteq, and QMJ_inteq. Thus, cross-market linkages appear to be strongest within the equity asset class across regions. However, some of the equity factors also load on the corporate bond, currency, and equity index factors in the top model. Almost all alphas of U.S. equity factors not included in the integrated model are not statistically significant. Only the alphas of ROE_useq, EG_useq, and QMJ_useq are significant at the 5% level. The GRS test fails to reject the null hypothesis that the alphas of the U.S. equity factors not included in the top unified model are jointly zero.

Panel B reports the results for international equities (excluding the U.S.). The U.S. stock market factor and the three international equity factors appear to be important for this asset class. All but one of the excluded international equity factors are significantly related to QMJ_inteq. There are few significant loadings on the other asset class factors after controlling for the effects of U.S. and international equity factors. The alphas of HML_inteq, HMLm_inteq, and iMOM_inteq are statistically significant. Consequently, the GRS test also rejects that all alphas are jointly zero. However, recalling from the previous section of the factor protocol, these factors are not materially related to the covariance matrix of international equity returns, and thus cannot be considered viable risk factors.

The results for the next major asset class, corporate bonds, are in Panel C. The factors also load mainly on those of their own asset class as well as on MKT_useq. Only two out of 17 factors have significant alphas at the 5% level. Among them, STR_cb fails the necessary condition of the PKR factor protocol. In addition, the GRS test can not reject the null hypothesis that the alphas for all excluded corporate bond factors are jointly zero.

We now turn to the explanatory power of the top integrated model for the four remaining asset classes. The results are reported in Panels D to G. Although

Table 4.9: Spanning Regressions

This table reports the results of spanning regressions of the factors of each asset class that are not selected on the top model across asset classes from the model scanning procedure. We present the intercept from the spanning regressions (α) as well as the loadings of the excluded factors on the winning model factors from August 2006 to December 2019. The t-statistics are based on robust Newey & West (1987) standard errors with four lags. R^2 is the coefficient of determination of the simple regressions. *GRS* reports the results of the Gibbons et al. (1989) test of the null hypothesis that all alphas are jointly zero for each asset class. The GRS test statistics are shown on the bottom of each panel, along with the corresponding p-values in brackets. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: U.S. Equities

	α	MKT_useq	SMB_inteq	MGMT_inteq	QMJ_inteq	Carry_cb	MOM_cbeq	MOMfx_aqr	Carry_eqi	R^2
SMB_useq	−0.22	0.24***	0.28***	−0.02	0.00	−0.03	−0.03	−0.01	−0.14	20.3
	(−0.94)	(3.85)	(2.67)	(−0.11)	(0.03)	(−0.38)	(−0.12)	(−0.17)	(−1.56)	
HML_useq	0.19	0.08	−0.07	0.54***	−0.59***	−0.27**	−0.49	−0.09	−0.06	38.1
	(0.74)	(1.24)	(−0.54)	(3.32)	(−4.53)	(−2.60)	(−1.32)	(−1.35)	(−0.73)	
RMW_useq	0.21	−0.12***	−0.21**	−0.04	0.17**	0.08*	0.08	0.07	−0.11*	27.6
	(1.47)	(−3.48)	(−2.41)	(−0.47)	(1.98)	(1.71)	(0.55)	(1.51)	(−1.82)	
CMA_useq	0.10	−0.05	−0.04	0.33***	−0.20*	−0.06	0.14	0.08	−0.04	17.9
	(0.59)	(−1.17)	(−0.44)	(3.25)	(−1.94)	(−0.92)	(0.81)	(1.54)	(−0.93)	
MOM_useq	−0.33	0.08	0.28	0.14	1.38***	−0.55***	0.48	0.46***	−0.07	50.1
	(−0.75)	(0.58)	(1.30)	(0.41)	(3.74)	(−2.87)	(1.08)	(3.01)	(−0.46)	
STR_useq	0.24	0.23**	0.12	0.16	−0.38*	−0.13	0.03	0.23*	0.11	27.2
	(0.66)	(2.38)	(0.84)	(0.53)	(−1.76)	(−1.05)	(0.09)	(1.75)	(1.07)	
LTR_useq	−0.27	0.15**	0.12	0.59***	−0.39***	−0.09	−0.07	−0.11	−0.02	30.2
	(−1.10)	(2.19)	(0.93)	(2.73)	(−2.89)	(−0.65)	(−0.19)	(−1.24)	(−0.20)	
ME_useq	−0.10	0.25***	0.28***	0.04	−0.05	−0.09	−0.19	−0.02	−0.12	23.8
	(−0.42)	(4.17)	(2.85)	(0.24)	(−0.35)	(−1.08)	(−0.90)	(−0.27)	(−1.30)	
IA_useq	0.13	−0.04	−0.06	0.31***	−0.26**	−0.08	0.07	0.07	−0.05	17.6
	(0.75)	(−0.95)	(−0.57)	(3.11)	(−2.21)	(−1.39)	(0.43)	(1.40)	(−0.81)	
ROE_useq	0.44**	−0.10**	−0.19*	0.09	0.43***	−0.30***	−0.08	0.16**	−0.05	48.0
	(2.41)	(−2.06)	(−1.70)	(0.70)	(2.84)	(−4.09)	(−0.38)	(2.34)	(−0.79)	
EG_useq	0.37**	−0.12**	−0.17*	−0.05	0.25***	−0.02	0.15	−0.13**	−0.11**	35.9
	(2.43)	(−2.34)	(−1.93)	(−0.59)	(2.68)	(−0.22)	(1.00)	(−2.12)	(−2.23)	
PEAD_useq	0.10	0.04	−0.08	−0.15	0.50***	−0.21***	0.28	−0.02	−0.06	36.2
	(0.49)	(0.74)	(−0.87)	(−1.04)	(4.48)	(−3.34)	(1.36)	(−0.28)	(−0.85)	
FIN_useq	0.46	−0.22***	−0.33*	0.49**	−0.11	−0.13	0.27	−0.04	−0.19**	31.2
	(1.55)	(−3.70)	(−1.79)	(2.51)	(−0.64)	(−1.65)	(1.06)	(−0.48)	(−2.13)	
LIQ_useq	0.22	0.06	0.04	−0.98***	−0.08	0.00	−0.05	0.12	0.17	19.5
	(0.69)	(0.78)	(0.22)	(−3.15)	(−0.46)	(0.00)	(−0.16)	(0.97)	(1.43)	
HMLm_useq	0.27	−0.02	−0.15	0.24	−1.16***	0.28**	−0.72**	−0.19**	−0.03	55.1
	(0.82)	(−0.26)	(−0.78)	(1.00)	(−4.11)	(2.31)	(−2.01)	(−2.33)	(−0.22)	
BAB_useq	0.03	0.05	0.45***	−0.21	0.36*	0.05	0.44	0.09	0.07	17.4
	(0.12)	(0.67)	(3.14)	(−0.61)	(1.81)	(0.30)	(1.18)	(0.81)	(0.70)	
QMJ_useq	0.34**	−0.19***	−0.19*	0.19*	0.46***	−0.14*	0.40**	−0.03	−0.15**	61.5
	(2.35)	(−4.70)	(−1.93)	(1.78)	(4.83)	(−1.76)	(2.12)	(−0.49)	(−2.52)	
SMB_useqsy	−0.04	0.23***	0.19*	−0.00	−0.03	−0.05	−0.01	−0.01	−0.08	22.0
	(−0.17)	(4.24)	(1.91)	(−0.01)	(−0.27)	(−0.55)	(−0.08)	(−0.16)	(−0.98)	
MGMT_useq	0.23	0.02	−0.16*	0.86***	−0.04	−0.27***	−0.09	−0.01	−0.04	45.5
	(1.36)	(0.57)	(−1.67)	(6.56)	(−0.33)	(−4.65)	(−0.41)	(−0.26)	(−0.67)	
PERF_useq	0.50	−0.33***	−0.29	−0.73***	0.94***	−0.14	1.22***	0.25***	−0.09	58.5
	(1.65)	(−3.62)	(−1.57)	(−2.81)	(4.12)	(−1.17)	(4.04)	(2.98)	(−0.72)	
GRS	1.47									
	[0.11]									

continued on the next page

Table 4.9: Spanning Regressions (continued)

Panel B: International Equities

	α	MKT_useq	SMB_inteq	MGMT_inteq	QMJ_inteq	Carry_cb	MOM_cbeq	MOMfx_aqr	Carry_eqi	R^2
MKT_inteq	0.12	0.68***	0.02	−0.54***	−0.70***	0.30**	0.04	0.13	0.21**	80.4
	(0.41)	(10.2)	(0.15)	(−2.72)	(−5.38)	(2.36)	(0.15)	(1.38)	(2.23)	
HML_inteq	0.46***	−0.09***	−0.35***	0.25***	−0.50***	−0.02	0.14	0.05	−0.08	46.7
	(3.88)	(−3.71)	(−4.39)	(2.73)	(−8.46)	(−0.53)	(1.09)	(1.21)	(−1.50)	
HMLm_inteq	0.54***	−0.12***	−0.45***	0.09	−0.80***	0.09	0.25*	−0.07	−0.06	57.0
	(3.72)	(−3.31)	(−3.76)	(0.73)	(−10.4)	(1.44)	(1.86)	(−1.10)	(−0.90)	
RMW_inteq	0.04	0.07***	−0.19**	−0.06	0.38***	0.00	0.15*	−0.01	−0.01	49.6
	(0.44)	(3.17)	(−2.33)	(−0.97)	(6.42)	(0.00)	(1.78)	(−0.33)	(−0.46)	
ROE_inteq	0.03	0.06**	−0.21***	−0.11	0.39***	0.01	0.10	−0.06	0.01	46.3
	(0.26)	(2.49)	(−2.71)	(−1.28)	(4.63)	(0.11)	(0.75)	(−1.48)	(0.30)	
CMA_inteq	0.07	−0.08***	0.11**	0.61***	0.02	−0.00	0.12	0.03	−0.06	58.4
	(0.80)	(−2.84)	(2.13)	(5.98)	(0.29)	(−0.04)	(1.11)	(0.58)	(−1.44)	
MOM_inteq	0.20	0.06	0.22*	0.25	0.87***	−0.25*	0.01	0.34***	0.09	47.6
	(0.75)	(0.85)	(1.69)	(1.22)	(4.46)	(−1.95)	(0.02)	(2.86)	(1.01)	
iMOM_inteq	0.34***	−0.01	0.05	0.03	0.18***	−0.03	0.04	0.09*	−0.01	14.6
	(2.89)	(−0.23)	(0.64)	(0.22)	(2.73)	(−0.64)	(0.24)	(1.90)	(−0.13)	
PERF_inteq	0.07	0.05*	−0.07	0.11	0.73***	−0.06	0.09	−0.03	0.02	63.1
	(0.49)	(1.81)	(−1.06)	(1.28)	(8.35)	(−1.14)	(0.70)	(−0.44)	(0.64)	
BAB_inteq	−0.09	−0.20***	0.13	0.62***	1.00***	−0.07	0.04	−0.01	0.00	81.1
	(−0.48)	(−5.38)	(1.21)	(4.77)	(10.8)	(−1.37)	(0.25)	(−0.10)	(0.05)	
GRS	4.42***									
	[0.00]									

Panel C: Corporate Bonds

	α	MKT_useq	SMB_inteq	MGMT_inteq	QMJ_inteq	Carry_cb	MOM_cbeq	MOMfx_aqr	Carry_eqi	R^2
MKT_cb	−0.11	0.05	0.05	−0.03	0.08	0.42***	0.20	0.07	0.07*	50.6
	(−1.06)	(1.17)	(0.75)	(−0.27)	(1.37)	(6.29)	(0.87)	(1.09)	(1.96)	
TERM_cb	−0.18	−0.05	0.09	0.19	0.60***	0.22	0.66	−0.05	0.05	19.8
	(−0.57)	(−0.34)	(0.58)	(0.67)	(3.27)	(0.76)	(0.86)	(−0.32)	(0.46)	
DEF_cb	−0.25	0.12***	−0.03	−0.04	−0.12	0.42***	−0.31*	0.18**	0.01	52.0
	(−1.53)	(2.71)	(−0.30)	(−0.32)	(−1.29)	(4.64)	(−1.71)	(2.00)	(0.13)	
DRF_cb	0.05	0.11*	−0.02	0.21	0.03	0.43**	0.20	−0.23*	−0.01	20.8
	(0.17)	(1.66)	(−0.16)	(1.01)	(0.13)	(2.45)	(0.56)	(−1.91)	(−0.16)	
LRF_cb	0.12	−0.01	0.15*	−0.02	0.01	0.39***	−0.19	−0.09	0.08	44.9
	(1.08)	(−0.51)	(1.84)	(−0.31)	(0.11)	(3.70)	(−1.17)	(−1.37)	(1.61)	
CRF_cb	0.19	0.14***	0.16	−0.12	−0.26***	0.09	0.44*	−0.11	0.02	29.7
	(1.14)	(2.74)	(1.34)	(−0.77)	(−3.14)	(1.38)	(1.92)	(−1.09)	(0.30)	
Value_cb	−0.01	0.02	0.03	0.06	0.06	0.54***	0.65***	0.03	0.03	60.8
	(−0.11)	(0.91)	(0.53)	(0.83)	(0.96)	(6.93)	(4.00)	(0.69)	(0.96)	
DUR_cb	−0.44**	0.15	−0.05	−0.07	0.09	0.77***	0.67	0.14	0.08	50.6
	(−2.21)	(1.47)	(−0.47)	(−0.36)	(0.65)	(3.79)	(1.30)	(1.40)	(1.20)	
VOL_cb	−0.31*	0.16**	−0.04	−0.03	0.09	0.69***	0.34	0.12	0.07	58.6
	(−1.95)	(2.00)	(−0.47)	(−0.18)	(0.78)	(4.22)	(0.83)	(1.42)	(1.31)	
STR_cb	0.34**	−0.07	0.04	0.06	−0.01	0.04	0.05	0.10*	0.02	6.78
	(2.36)	(−1.50)	(0.73)	(0.55)	(−0.14)	(0.36)	(0.20)	(1.97)	(0.38)	
MOM_cb	−0.17	0.04	−0.02	−0.13	−0.05	−0.30**	0.58**	−0.07	0.05	26.9
	(−1.08)	(1.07)	(−0.19)	(−1.30)	(−0.74)	(−2.23)	(2.20)	(−0.79)	(0.99)	
LTR_cb	0.13	0.01	0.21*	0.16	−0.11	−0.00	−0.47**	0.03	−0.10*	13.6
	(0.71)	(0.21)	(1.97)	(1.25)	(−1.33)	(−0.02)	(−2.58)	(0.23)	(−1.75)	
VIX_cb	0.08	−0.03	0.03	−0.14**	−0.12**	0.10	0.22	−0.01	−0.00	22.4
	(1.25)	(−1.38)	(0.73)	(−2.26)	(−2.35)	(1.52)	(1.31)	(−0.38)	(−0.11)	
UNC_cb	0.06	0.04	−0.11	−0.21**	−0.14	−0.04	0.39*	−0.08	0.05	15.7
	(0.42)	(0.91)	(−1.45)	(−2.21)	(−1.59)	(−0.37)	(1.68)	(−1.54)	(1.31)	
EPU_cb	0.08	0.05**	−0.09*	−0.04	−0.06	0.02	0.05	0.01	0.01	19.0
	(0.86)	(2.04)	(−1.81)	(−0.67)	(−0.93)	(0.29)	(0.32)	(0.37)	(0.31)	
EPUtax_cb	0.07	0.03*	−0.09**	−0.11	−0.04	−0.01	0.01	−0.02	0.01	13.5
	(1.11)	(1.72)	(−2.32)	(−1.54)	(−1.18)	(−0.22)	(0.11)	(−0.40)	(0.49)	
GRS	1.46									
	[0.13]									

continued on the next page

Table 4.9: Spanning Regressions (continued)

Panel D: Commodities

	α	MKT_useq	SMB_inteq	MGMT_inteq	QMJ_inteq	Carry_cb	MOM_cbeq	MOMfx_aqr	Carry_eqi	R^2
MKT_cm	0.65*	0.25**	0.02	−1.27***	−0.65***	−0.12	−0.25	0.22*	0.47***	46.2
	(1.68)	(2.35)	(0.13)	(−4.84)	(−2.94)	(−0.78)	(−0.83)	(1.83)	(3.33)	
Value_cm	0.77	−0.20**	−0.39*	0.57*	−0.01	0.44**	−0.71	0.18	−0.30	11.5
	(1.61)	(−2.12)	(−1.79)	(1.80)	(−0.03)	(2.09)	(−1.40)	(1.11)	(−1.39)	
MOM_cm	−0.51	0.33***	0.28	−0.42	0.34	−0.23	0.98**	0.37**	−0.33*	17.5
	(−0.89)	(3.10)	(1.58)	(−1.54)	(1.34)	(−1.40)	(2.03)	(1.99)	(−1.79)	
Carry_cm	0.22	0.13	−0.00	−0.32	−0.02	0.02	0.74*	0.03	−0.09	4.61
	(0.51)	(1.25)	(−0.02)	(−1.48)	(−0.12)	(0.09)	(1.76)	(0.24)	(−0.73)	
TSMOM_cm	−0.09	0.14	0.60***	−0.39	0.64**	−0.23	0.12	0.60***	−0.12	21.3
	(−0.19)	(1.19)	(2.80)	(−1.22)	(2.57)	(−1.21)	(0.24)	(3.39)	(−0.54)	
GRS	1.87									
	[0.10]									

Panel E: Currencies

	α	MKT_useq	SMB_inteq	MGMT_inteq	QMJ_inteq	Carry_cb	MOM_cbeq	MOMfx_aqr	Carry_eqi	R^2
MKT_fx	0.12	0.15***	−0.09	−0.25***	−0.27***	0.08	−0.03	0.03	0.20***	47.2
	(0.77)	(3.34)	(−1.10)	(−2.74)	(−3.74)	(0.95)	(−0.15)	(0.49)	(3.76)	
HML_fx	0.46*	0.15***	−0.15*	−0.27**	−0.30***	−0.05	0.23	0.12	0.11*	29.2
	(1.70)	(3.04)	(−1.72)	(−2.30)	(−2.72)	(−0.48)	(1.27)	(1.47)	(1.75)	
Carry_fx	0.15	−0.17***	0.13	0.27**	0.30***	0.08	−0.22	−0.13	−0.12*	30.1
	(0.60)	(−3.35)	(1.43)	(2.25)	(2.66)	(0.77)	(−1.24)	(−1.51)	(−1.86)	
Dollar_fx	−0.04	−0.15***	0.11	0.27***	0.28***	−0.06	0.04	−0.03	−0.21***	47.3
	(−0.27)	(−3.24)	(1.24)	(2.88)	(3.74)	(−0.71)	(0.19)	(−0.55)	(−3.84)	
Value_fxaqr	0.41**	−0.05	−0.16**	0.11	−0.05	−0.03	−0.27*	−0.11	−0.11*	13.0
	(2.47)	(−1.32)	(−2.35)	(0.85)	(−0.54)	(−0.40)	(−1.66)	(−1.00)	(−1.85)	
Carry_fxaqr	0.27	0.28***	−0.02	−0.23	−0.32***	−0.09	0.25	0.19	0.17**	47.2
	(1.09)	(5.62)	(−0.19)	(−1.43)	(−2.97)	(−0.84)	(1.12)	(1.25)	(2.35)	
TSMOM_fx	0.83**	−0.28**	0.05	−0.11	0.32	0.12	−0.83	1.12***	−0.05	26.1
	(2.10)	(−2.39)	(0.21)	(−0.34)	(0.94)	(0.49)	(−0.98)	(5.05)	(−0.25)	
GRS	2.65**									
	[0.03]									

Panel F: Equity Indices

	α	MKT_useq	SMB_inteq	MGMT_inteq	QMJ_inteq	Carry_cb	MOM_cbeq	MOMfx_aqr	Carry_eqi	R^2
MKT_eqi	0.22	0.63***	−0.23**	−0.55***	−0.52***	0.17**	0.14	0.04	0.03	83.9
	(1.20)	(12.1)	(−2.59)	(−4.00)	(−6.62)	(2.46)	(0.80)	(0.56)	(0.35)	
Value_eqi	0.03	−0.13***	−0.25***	−0.04	−0.43***	0.20**	0.19	−0.16*	0.12	18.5
	(0.16)	(−2.76)	(−2.89)	(−0.27)	(−4.03)	(2.15)	(1.07)	(−1.77)	(1.43)	
MOM_cbeqi	0.03	0.04	0.11	−0.01	0.55***	−0.13	−0.04	0.18*	−0.09	22.5
	(0.14)	(0.70)	(0.80)	(−0.07)	(4.44)	(−1.11)	(−0.15)	(1.84)	(−0.77)	
Defensive_eqi	−0.18	0.13**	0.06	0.43***	0.46***	−0.03	−0.00	0.05	−0.06	15.6
	(−0.82)	(2.31)	(0.42)	(2.78)	(3.82)	(−0.31)	(−0.02)	(0.60)	(−0.51)	
TSMOM_cbeqi	−0.18	0.21	−0.01	1.09	0.40	0.13	0.57	0.49	−0.52	7.34
	(−0.21)	(0.84)	(−0.03)	(1.62)	(0.92)	(0.39)	(0.57)	(1.61)	(−1.65)	
GRS	0.38									
	[0.86]									

Panel G: Government Bonds

	α	MKT_useq	SMB_inteq	MGMT_inteq	QMJ_inteq	Carry_cb	MOM_cbeq	MOMfx_aqr	Carry_eqi	R^2
MKT_govtb	0.04	−0.03	0.02	0.12	0.20***	0.13	0.26	−0.05	0.06	18.2
	(0.25)	(−0.68)	(0.32)	(1.09)	(2.65)	(1.33)	(1.04)	(−0.85)	(1.16)	
Value_govtb	0.13	−0.03	−0.12**	0.02	−0.25***	0.07	0.14*	−0.05	0.03	16.6
	(1.14)	(−1.27)	(−2.03)	(0.27)	(−4.29)	(1.58)	(1.68)	(−1.27)	(0.66)	
MOM_govtb	−0.07	0.04	0.02	0.10	0.25***	−0.00	−0.00	0.19***	−0.13***	20.1
	(−0.58)	(1.39)	(0.30)	(1.63)	(3.51)	(−0.06)	(−0.04)	(4.07)	(−2.74)	
Carry_govtb	0.19*	0.00	−0.04	−0.05	−0.17**	0.04	0.15*	−0.06	0.05	9.48
	(1.75)	(0.01)	(−0.84)	(−0.57)	(−2.15)	(0.83)	(1.88)	(−1.64)	(0.82)	
Defensive_govtb	0.17**	0.03	−0.05	−0.00	−0.12**	−0.04	−0.10	−0.01	0.05	14.8
	(2.31)	(1.08)	(−1.10)	(−0.04)	(−2.10)	(−1.10)	(−0.68)	(−0.25)	(1.33)	
TSMOM_govtb	0.19	−0.12	−0.25	−0.06	1.57***	0.42	0.98	0.43	0.04	22.4
	(0.25)	(−0.54)	(−0.58)	(−0.11)	(3.77)	(1.11)	(1.01)	(1.42)	(0.13)	
GRS	0.62									
	[0.72]									

the optimal integrated model does not include any commodity factors, it can explain them quite well. They load on all the factors in the top model. None of the alphas are significant at the 5% level and the GRS test does not reject its null hypothesis. For currency returns, MKT_useq, QMJ_inteq, and Carry_eqi seem to be important, as many currency factors load on them. The currency factors have little to no exposure to the two corporate bond factors and, interestingly, only TSMOM_fx has a significant loading on the only currency factor included in the benchmark model (MOM_fxaqr). The GRS test rejects the null hypothesis that all alphas from the regressions of the currency factors on the optimal model are jointly zero. However, this is mainly because the model cannot explain the Value_fxaqr and TSMOM_fx factors, both of which do not satisfy the necessary condition of being viable risk factors.

Similar to the international equities, the factors of the equity indices load mainly on the three international equity factors and the U.S. market factor. All individual alphas of the equity index factors are not significantly different from zero. Interestingly, all sovereign bond factors load significantly on QMJ_inteq, while they seem to be unrelated to the U.S. stock market and the two corporate bond factors. The GRS tests for both equity indices and government bonds fail to reject the null that all factor alphas are jointly zero.

Overall, an integrated model consisting of eight factors from five different asset classes can explain the majority of the remaining prominent existing factors from seven major asset classes (only eleven out of 77 factors have significant alphas with $t >= 1.96$; the GRS tests of five out of seven asset classes cannot reject the null that the corresponding alphas are jointly zero). Not surprisingly, many factors have exposures to the U.S stock market. Consistent with the results based on the cumulative posterior probabilities, we find that among the factors that are selected in the optimal unified model, MKT_useq, SMB_inteq, MGMT_inteq, QMJ_inteq, and Carry_cb play important roles in explaining many factors across asset classes. The evidence on integration is again split. On the one hand, the factors in the top model of an asset class generally explain the variation in other factors in the same asset class best. However, there are some notable cross-market linkages, suggesting that the integrated model may perform as well as or better than the asset-class-specific models, even when it comes to pricing different assets in that class.

4.5.4 Model Performance with Test Assets

Next, we thus turn to an left-hand-side (LHS) approach and examine the empirical performance of different factor models based on test assets for the different asset classes. Following Lewellen et al. (2010), we use a wide range of test portfolios that are not simply directly related to the risk characteristics used to construct these factors. In particular, we use the following sets of test portfolios for the different asset classes: the thirteen aggregated theme anomaly long-short portfolios of Jensen et al. (2022) (JKP) for U.S. equities and international equities, 29 long-short characteristic portfolios for corporate bonds as in Chapter 3, 23 individual commodities, 23 government bonds, 43 equity indices, and 48 global value and momentum portfolios across asset classes and markets.

Table 4.10 summarizes the results. In general, the top integrated factor model generates low GRS statistics and a small number of significant alphas at the 5% level. In particular, the GRS test does not reject the null hypothesis that all the alphas of the test asset portfolios are jointly zero for the U.S. equity test portfolios, commodities, equity indices, and global bond government bonds. It can also capture a reasonable amount of time-series variation in returns across asset classes, yielding time-series R^2s that are comparably large to those of many existing asset-class-specific models. Finally, also the average absolute alphas are comparatively low.

Finally, we directly compare this top model to another integrated factor model: the AMP global three-factor model (consisting of a global market factor and cross-asset-class value and momentum factors) in explaining the 48 high, medium, and low VME value and momentum portfolios across asset classes. While 17 of the 48 portfolios have significant alphas with respect to the AMP global

Table 4.10: Time-Series Asset Pricing Tests with Test Assets

This table reports the results of time-series asset pricing tests of the existing pricing models, as well as the asset-class specialized and integrated winning models from the model scan (see Table 4.2, Panel B of Table 4.4, and Panel A of Table 4.6). In Panels A through F, we examine the models' performance on asset class specific test portfolios. Finally, in Panel G, we report the results for the global models across all asset classes. *GRS* reports the results for the Gibbons et al. (1989) test of the null hypothesis that all alphas for a model are jointly zero, with the corresponding *p*-value in brackets. $A|\alpha_i|$ is the average absolute alpha of the test portfolios. $\#sig\ \alpha_i$ reports how many test portfolios have significant alphas at the 5% level. We use Newey & West (1987) standard errors with four lags. $\frac{A|\alpha_i|}{A|r_i|}$ is the ratio of average absolute alpha to average absolute portfolio return. $\frac{A\alpha_i^2}{Ar_i^2}$ is the ratio of the corresponding squares. $\frac{As^2(\alpha_i)}{A\alpha_i^2}$ is the ratio of the average squared standard error of the alphas to the average squared alpha. $A(R^2)$ is the average adjusted R^2 of the regressions (in percentage points). $SH^2(f)$ is the squared Sharpe ratio of the optimal portfolio from the model factors, and $SH^2(\alpha)$ is the squared Sharpe ratio achievable with the alphas of the test assets. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	GRS	[*p*-value]	$A\|\alpha_i\|$	$\#sig\ \alpha_i$	$\frac{A\|\alpha_i\|}{A\|r_i\|}$	$\frac{A\alpha_i^2}{Ar_i^2}$	$\frac{As^2(\alpha_i)}{A\alpha_i^2}$	$A(R^2)$	$SH^2(f)$	$SH^2(\alpha)$
Panel A: U.S. Equities - 13 JKP anomaly theme portfolios										
winner_useq	2.35***	[0.01]	0.10	2	0.83	0.61	1.06	0.21	0.17	0.22
winner_across	1.57	[0.11]	0.10	1	0.87	0.62	1.44	0.33	1.10	0.27
AMP_across	2.85***	[0.00]	0.13	2	1.04	1.22	0.33	0.30	0.02	0.23
CAPM_useq	3.82***	[0.00]	0.20	5	1.69	2.73	0.21	0.13	0.03	0.32
FF3_useq	3.41***	[0.00]	0.15	4	1.26	1.48	0.30	0.38	0.06	0.29
C4_useq	3.38***	[0.00]	0.15	3	1.22	1.40	0.21	0.47	0.06	0.29
FF5_useq	2.51***	[0.01]	0.11	3	0.94	0.74	0.51	0.45	0.13	0.23
FF6_useq	2.50***	[0.01]	0.11	5	0.91	0.71	0.33	0.53	0.13	0.23
HXZ4_useq	3.56***	[0.00]	0.13	5	1.09	1.09	0.40	0.39	0.09	0.31
HXZ5_useq	2.62***	[0.00]	0.11	4	0.88	0.62	0.61	0.40	0.18	0.25
DHS_useq	3.44***	[0.00]	0.17	6	1.41	1.71	0.26	0.32	0.10	0.30
SY_useq	2.48***	[0.01]	0.11	2	0.93	0.68	0.55	0.49	0.13	0.23
Panel B: International Equities - 13 JKP anomaly theme portfolios										
winner_inteq	3.09***	[0.00]	0.11	7	0.75	0.73	0.40	0.50	0.45	0.20
winner_across	3.96***	[0.00]	0.10	5	0.65	0.52	0.44	0.52	1.10	0.38
AMP_across	18.6***	[0.00]	0.16	9	1.07	1.21	0.13	0.34	0.02	0.98
CAPM_inteq	6.07***	[0.00]	0.18	8	1.17	1.46	0.14	0.15	0.01	0.32
FF3_inteq	10.5***	[0.00]	0.19	8	1.23	1.94	0.08	0.38	0.05	0.49
C4_inteq	15.8***	[0.00]	0.14	8	0.96	1.13	0.09	0.51	0.11	0.91
FF5_inteq	2.56**	[0.02]	0.09	4	0.62	0.47	0.67	0.46	0.42	0.16
FF6_inteq	1.81*	[0.08]	0.10	7	0.65	0.46	0.21	0.58	0.42	0.13
HXZ4_inteq	1.67	[0.11]	0.09	3	0.61	0.43	1.00	0.37	0.47	0.13
SY_inteq	11.8***	[0.00]	0.14	7	0.91	0.86	0.24	0.46	0.30	0.80

continued on the next page

Table 4.10: Time-Series Asset Pricing Tests with Test Assets (continued)

	GRS	[p-value]	$A\|\alpha_i\|$	$\#sig\ \alpha_i$	$\frac{A\|\alpha_i\|}{A\|r_i\|}$	$\frac{A\alpha_i^2}{Ar_i^2}$	$\frac{As^2(\alpha_i)}{A\alpha_i^2}$	$A(R^2)$	$SH^2(f)$	$SH^2(\alpha)$
Panel C: Corporate Bonds - 29 LS characteristic portfolios										
winner_cb	3.76***	[0.00]	0.09	7	0.41	0.17	0.29	0.55	0.57	0.52
winner_across	2.35***	[0.01]	0.11	3	0.50	0.30	0.44	0.44	1.10	0.47
AMP_across	6.25***	[0.00]	0.18	11	0.87	0.77	0.14	0.18	0.02	0.65
CAPM_cb	5.65***	[0.00]	0.12	4	0.56	0.35	0.24	0.21	0.08	0.53
FF3_cb	5.91***	[0.00]	0.10	5	0.48	0.32	0.24	0.34	0.12	0.53
augFF3_cb	6.30***	[0.00]	0.11	9	0.53	0.29	0.19	0.40	0.19	0.66
FF5_cb	6.75***	[0.00]	0.11	4	0.53	0.42	0.22	0.35	0.10	0.59
BBW	4.71***	[0.00]	0.13	8	0.59	0.29	0.24	0.31	0.16	0.48
BSW	4.46***	[0.00]	0.10	4	0.47	0.26	0.31	0.27	0.19	0.46
IRP	5.27***	[0.00]	0.07	6	0.32	0.09	0.57	0.57	0.54	0.59
KPP	4.13***	[0.00]	0.08	4	0.36	0.11	0.46	0.56	0.40	0.38
Panel D: Commodities										
winner_commodity	1.63**	[0.04]	0.55	2	1.07	0.92	0.81	0.03	0.07	0.35
winner_across	1.36	[0.13]	0.87	4	1.69	1.97	0.63	0.14	1.10	0.58
AMP_across	2.04***	[0.00]	0.59	2	1.15	1.09	0.55	0.10	0.02	0.42
CAPM_cm	1.91***	[0.01]	0.52	5	1.01	1.00	0.45	0.28	0.00	0.38
BGR_cm	1.80**	[0.02]	0.50	4	0.98	0.92	0.50	0.30	0.02	0.37
AQR4_cm	1.60**	[0.04]	0.51	5	1.00	0.84	0.57	0.30	0.08	0.35
Panel E: Equity Indices										
winner_eqindices	0.85	[0.73]	0.43	0	1.15	1.25	1.39	0.01	0.02	0.32
winner_across	1.04	[0.43]	0.50	5	1.33	2.28	0.65	0.54	1.10	0.81
AMP_across	0.93	[0.61]	0.28	3	0.75	0.62	1.21	0.57	0.02	0.34
CAPM_eqi	0.96	[0.55]	0.33	7	0.89	0.88	0.90	0.51	0.01	0.37
AQR_eqi	0.81	[0.79]	0.26	1	0.69	0.51	1.54	0.53	0.06	0.32
Panel F: Government Bonds										
winner_govtb	0.50	[0.93]	0.14	0	0.51	0.37	2.31	0.08	0.06	0.05
winner_across	0.50	[0.94]	0.16	0	0.60	0.49	2.26	0.20	1.10	0.11
AMP_across	1.05	[0.41]	0.16	1	0.59	0.39	1.61	0.22	0.02	0.11
CAPM_govtb	0.50	[0.93]	0.14	0	0.51	0.37	2.31	0.08	0.06	0.05
AQR_govtb	0.47	[0.95]	0.08	0	0.30	0.13	6.52	0.19	0.09	0.05
Panel G: Global 48 VME portfolios										
winner_across	1.35	[0.11]	0.29	5	0.73	0.68	0.46	0.59	1.10	0.98
AMP_across	6.01***	[0.00]	0.23	17	0.58	0.36	0.43	0.61	0.02	2.02

model, only five do so with respect to the top integrated model. The GRS test favors the top integrated model over the AMP global model, as it fails to the null hypothesis that the alphas are all jointly zero.

Figure 4.2 plots the actual sample average returns of the 48 VME portfolios across asset classes against their predicted expected returns. On the left are the fitted expected returns from the AMP global three-factor model, while on the right are those from the top integrated model. In addition, we plot the 45-degree line through the origin to highlight the magnitude of the pricing errors. If a model works well, then these dots representing the test assets should line up well along the 45-degree line (i.e., data and model are in agreement). For the AMP global model, we see that the dots in the left graph are mostly clustered along the x-axis dimension and spread out along the y-axis dimension. On the other hand, for the top unified model, the dots are spread out along both dimensions, with a fairly good alignment (with some dispersion) along the 45-degree line. Thus, the top integrated model across asset classes can explain the returns of these test portfolios across asset classes quite well.

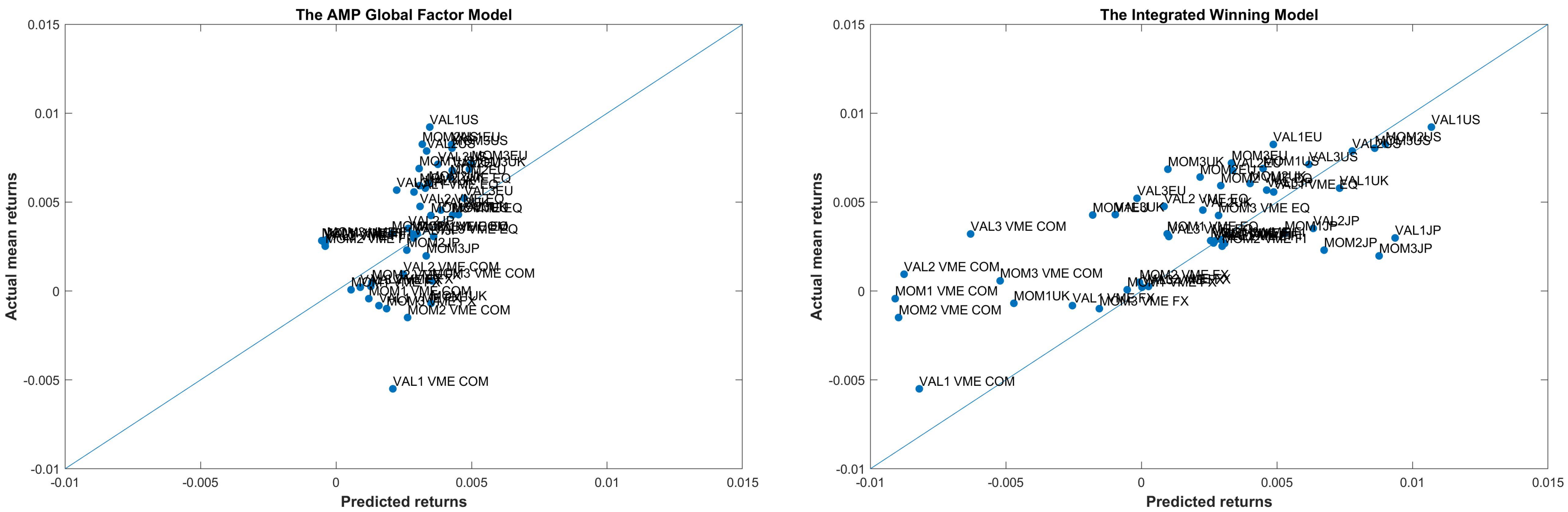
Taking all the evidence from both the right-hand-side (RHS) and LHS tests together, we find that an integrated model that includes factors from different asset classes performs quite well. The integrated factor model should definitely be used as a benchmark to evaluate multi-asset-class strategies. However, it can also be used for single-asset classes, where it performs on a par with the best specialized models. The fact that factors also help to explain returns across asset classes underscores the notion that asset classes are not entirely disintegrated.

4.6 Conclusion

Our study is motivated by the proliferation of factors across asset classes. However, different asset classes are typically analyzed in isolation. Little is known about the cross-market linkages of these factors. Therefore, we analyze market integration from a multi-asset, multi-factor perspective. We find that the different asset classes

Figure 4.2: The Model Performance Explaining 48 VME Portfolios

This figure plots the model-implied returns versus the realized average excess returns of the 48 high, medium, and low value and momentum portfolios across markets and asset classes of Asness et al. (2013) (48 VME portfolios). Each dot represents one portfolio. A 45-degree line is added to highlight the pricing errors given by the vertical distances from the line. The predicted returns on the left-hand-side are from the Asness et al. (2013) (AMP) three-factor model across asset classes (consisting of a global market factor and across-asset-class value and momentum factors), and those on the right-hand-side are from the optimal integrated model based on eight selected factors across asset classes using the two-step approach {MKT_useq SMB_inteq MGMT_inteq QMJ_inteq Carry_eqi Carry_cb MOM_cbeq MOM_fxaqr}. The sample period is from August 2006 through December 2019.



are far from fully integrated: models for one asset class typically fail to explain factors from other asset classes. Thus, there are multiple systematic drivers of returns across asset classes.

There are also some cross-market linkages, though. We thus distill and analyze a unified factor model that can describe returns across asset classes. Of the 77 factors across the seven different asset classes, the U.S. stock market, the size, management and quality factors of international equities, and the corporate bond carry factor appear to be the most important components of an optimal unified model. An integrated model consisting of these five factors, plus the corporate bond equity momentum factor, the currency momentum factor, and the equity index carry factor, performs quite well across asset classes. It achieves high in-sample and out-of-sample Sharpe ratios. Finally, the top integrated model subsumes a long list of factors and performs well in pricing assets across different asset classes.

C Appendix

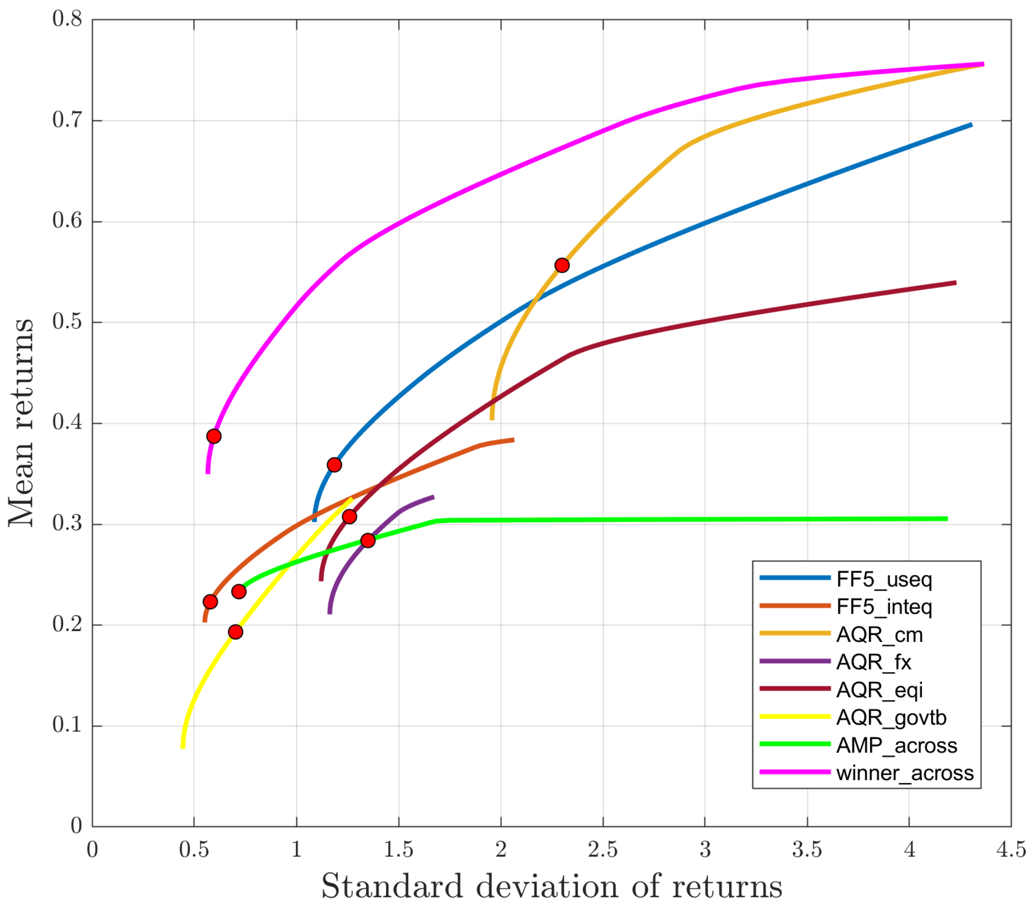
C.1 Robustness Check with a Longer Sample Period

In this section, we test the robustness of the main results using a longer sample period starting in 1991, but excluding corporate bonds. The optimal integrated factor set includes MKT_useq, QMJ_useq, SMB_inteq, MGMT_inteq, QMJ_inteq, MKT_govtb, Value_cm, Carry_cm. Consistent with the main analysis, we find that the same factors from U.S. and international equities make it into the integrated model. Not surprisingly, without corporate bonds, a government bond factor (MKT_govtb) is required in the optimal model. Interestingly, the optimal model also includes commodity factors for the longer sample period.

As can be seen in Figure C.1 for the longer sample excluding corporate bonds, the efficient frontier of the chosen combination of factors from the scanning approach is furthest to the northwest compared to those of the leading existing models in each asset class. The optimal combination of factors from multiple asset classes provides better investment opportunities than just the multiple styles from single asset classes.

Figure C.1: Efficient Frontiers - A Longer Sample Period

This figure plots the efficient frontiers and the tangency portfolios (red dots) for the representative models from each asset class: FF5_useq (U.S. equities), FF5_inteq (international equities, excluding the U.S.), AQR_cm (commodities), AQR_fx (currencies), AQR_eqi (equity indices), and AQR_govtb (government bond indices). In addition, it also plots the optimal integrated model based on eight selected factors across asset classes though the two-step approach {MKT_useq, QMJ_useq, SMB_inteq, MGMT_inteq, QMJ_inteq, MKT_govtb, Value_cm, Carry_cm} as a multi-asset, multi-factor investment strategy. To construct the efficient frontier, we require that all weights are positive and sum to one. The analysis is performed over the sample period from January 1991.



C.2 Model Selection Method Implementation Details

In this section, we describe the Bayesian marginal-likelihood-based model comparison approach introduced by Barillas & Shanken (2018) and revisited by Chib et al. (2020). This method allows a simultaneous comparison of all possible models based on subsets of the factor space under investigation. We compute their log marginal likelihoods to perform the prior–posterior update. The final model ranking is based on the posterior probabilities.

More specifically, for a set of K (traded) potential risk factors, there are generally $J = 2^K - 1$ possible factor combinations. The model space is thus $\mathcal{M} = \{\mathbb{M}_1, \mathbb{M}_2, ..., \mathbb{M}_J\}$. $\mathbb{M}_j$ is a possible model defined by the vector of included factors $\tilde{f}_j$ and the vector of excluded factors f_j^*.

Each factor model has a $L_j \times 1$ vector of included factors $\tilde{f}_j$ and a $(K - L_j) \times 1$ vector of excluded factors f_j^*. The data generating process of model j is thus given by

$$\tilde{f}_{j,t} = \tilde{\alpha}_j + \tilde{\epsilon}_{j,t}, \tag{C.1}$$

and

$$f_{j,t}^* = B_{j,f}^* \tilde{f}_{j,t} + \epsilon_{j,t}^*. \tag{C.2}$$

$\tilde{\alpha}_j$ is a $L_j \times 1$ parameter vector, and $\tilde{\epsilon}_{j,t}$ is a multivariate normally distributed residual vector. $B_{j,f}^*$ is a $(K - L_j) \times L_j$ parameter matrix. $\epsilon_{j,t}^*$ is also a multivariate normally distributed residual vector. A special case is when all factors are included in f_j.

The log marginal likelihood of a model $\mathbb{M}_j$ $(j \neq J)$ with y given the sample data of the factors over T time periods in closed form is

$$\log \tilde{m}(y|\mathbb{M}_j) = \log \tilde{m}(\tilde{f}|\mathbb{M}_j) + \log \tilde{m}(f^*|\mathbb{M}_j). \tag{C.3}$$

The first term on the RHS of Equation (C.3) is

$$- \frac{(K - L_j)L_j}{2} \log 2 - \frac{\tilde{T} L_j}{2} \log \pi - \frac{L_j}{2} \log (\tilde{T} k_j + 1)$$
$$- \frac{(\tilde{T} + L_j - K)}{2} \log |\psi_j| + \log \Gamma_{L_j} \left(\frac{\tilde{T} + L_j - K}{2} \right).$$

The second term on the RHS of Equation (C.3) is

$$\frac{(K-L_j)L_j}{2}\log 2-\frac{(K-L_j)(\tilde{T}-L_j)}{2}\log\pi$$
$$-\frac{(K-L_j)}{2}\log|W_j^*|-\frac{\tilde{T}}{2}\log|\psi_j^*|+\log\Gamma_{K-L_j}\left(\frac{\tilde{T}}{2}\right),$$

where $\tilde{T}=T-n_t$ and

$$W_j^* = \sum_{t=n_t+1}^{T}\tilde{f}_{j,t}\tilde{f}_{j,t}',$$
$$\psi_j = \sum_{t=n_t+1}^{T}(\tilde{f}_{j,t}-\hat{\tilde{\alpha}}_j)(\tilde{f}_{j,t}-\hat{\tilde{\alpha}}_j)'+\frac{\tilde{T}}{\tilde{T}k_j+1}\left(\hat{\tilde{\alpha}}_j-\tilde{\alpha}_{j0}\right)\left(\hat{\tilde{\alpha}}_j-\tilde{\alpha}_{j0}\right)'$$
$$\psi_j^* = \sum_{t=n_t+1}^{T}(f_{j,t}^*-\hat{B}_{j,f}^*\tilde{f}_{j,t})(f_{j,t}^*-\hat{B}_{j,f}^*\tilde{f}_{j,t})'.$$

$\Gamma_d(.)$ denotes the d-dimensional multivariate gamma function. All other variables are as previously defined. Hats on the parameters indicate that they are the estimates obtained by linear regressions of equations (C.1) and (C.2).

Following the recommendation of Chib et al. (2020), we use this model along with the model-specific prior $\tilde{\alpha}_j|M_j \sim \mathcal{N}(\tilde{\alpha}_{j0}, k_j\Sigma_j)$ with

$$\tilde{\alpha}_{j0} = n_t^{-1}\sum_{t=1}^{n_t}\tilde{f}_{j,t},$$

where $n_t = tr \times T$ is the size of the training sample, which we set to $tr = 10\%$ of the data, as in Chib et al. (2020). The model-specific multiplier k_j can be calculated as

$$k_j = \frac{1-tr}{tr}\times L_j^{-1}sum(diag(V_{j0})/diag(\hat{\Sigma}_{j0})),$$

where V_{j0} is the negative inverse Hessian over $\tilde{\alpha}_j$ and $\hat{\Sigma}_{j0}$ the estimate of the covariance matrix Σ_j in the training sample.

The end-product of the scanning process is a ranking of models

$$\{\mathbb{M}_1*, \mathbb{M}_2*, ..., \mathbb{M}_J*\} \tag{C.4}$$

by

$$\tilde{m}(y|\mathbb{M}_1*) > \tilde{m}(y|\mathbb{M}_2*) > ... > \tilde{m}(y|\mathbb{M}_J*). \tag{C.5}$$

$\mathbb{M}_1*$ denotes the winning model, identified as the model with the highest posterior probability. The remaining terms in the posterior probability calculation can be summarized by just a normalization constant. Thus, the ranking of the posterior probabilities is equivalent to the ranking of the marginal likelihoods $\tilde{m}(y|\mathbb{M}_j)$.

C.3 Additional Figures and Tables

Figure C.2: The Correlation Matrix of Monthly Factor Returns

This figure shows the correlation matrix for our set of candidate factors from seven different asset classes (see the list in Table 4.1) for their common sample period from August 2006 to December 2019.

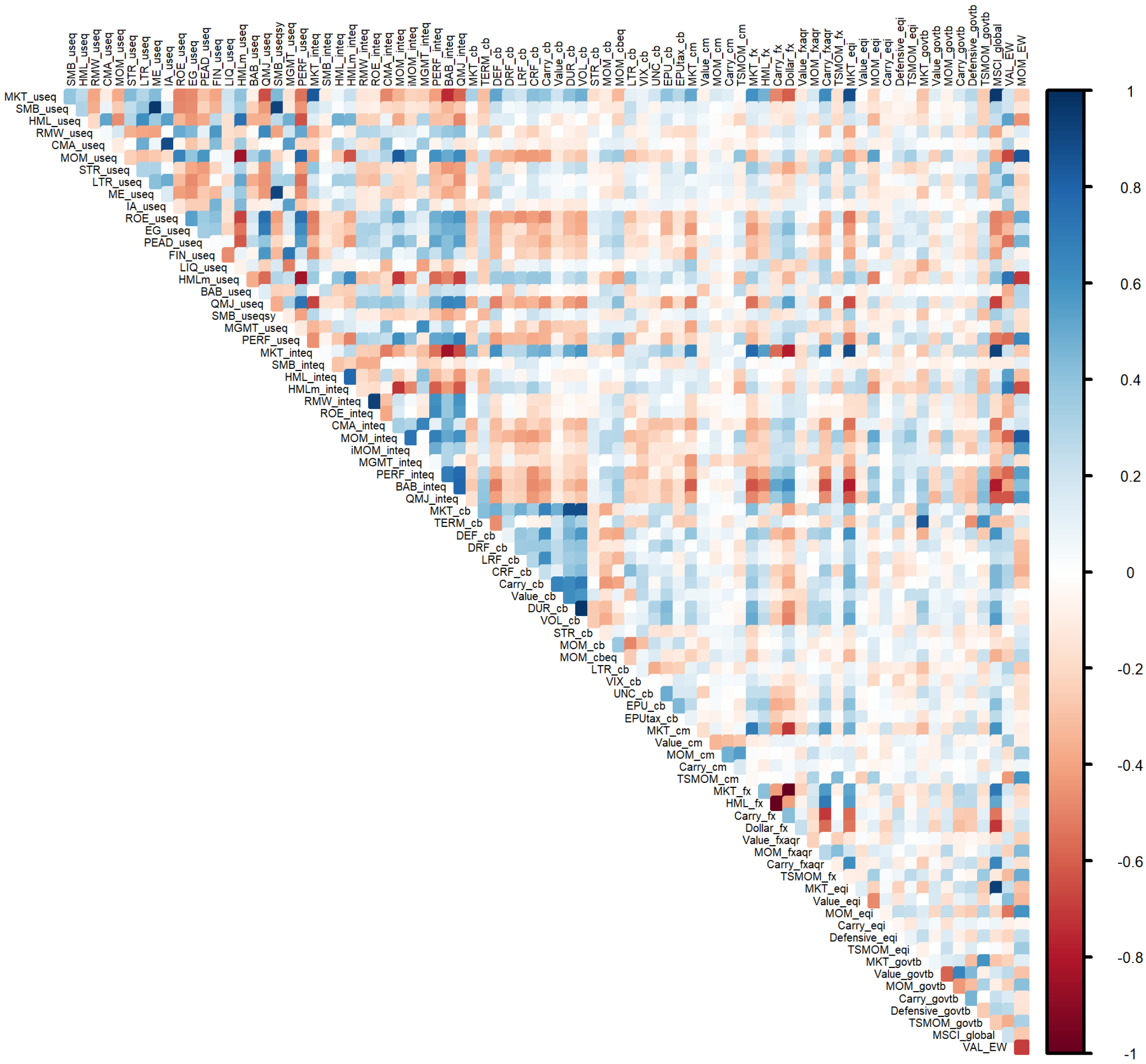


Table C.1: Data Sources

This table lists the URLs of the web data sources used in this chapter. We use the Welch & Goyal (2008) data to calculate the TERM and DEF factors. For other corporate bond factors for which data are not readily available, we use the Kelly & Pruitt (2022) dataset to calculate them ourselves as in Chapter 3. All other factors presented in Table 4.1 are available from the sources listed in the body of this table.

Authors/Sources	**URL to download data**
Fama & French (FF)	`https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`
Hou et al. (2015,2021) (HXZ)	`http://global-q.org/factors.html`
Stambaugh & Yuan (2017) (SY)	`https://finance.wharton.upenn.edu/~stambaug/`
Daniel et al. (2020) (DHS)	`http://www.kentdaniel.net/data.php`
Ilmanen et al. (2021) (AQR), Asness & Frazzini (2013), Frazzini & Pedersen (2014), Asness et al. (2019), Moskowitz et al. (2012)	`https://www.aqr.com/Insights/Datasets`
Pástor & Stambaugh (2003)	`https://faculty.chicagobooth.edu/-/media/faculty/lubos-pastor/research/liq_data_1962_2020.txt`
Hanauer (2020)	`https://www.globalfactorpremia.org/`
Jensen et al. (2022)	`https://jkpfactors.com/?country=world_ex_us&factor=all_factors&weighting=vw`
Bai et al. (2019) (BBW)	`https://sites.google.com/a/georgetown.edu/turan-bali/data-working-papers`
Welch & Goyal (2008)	`https://sites.google.com/view/agoyal145`
Kelly et al. (2023) (KPP)	`https://sethpruitt.net/research/`
Lustig et al. (2011) (LRV) Verdelhan (2018)	`http://web.mit.edu/adrienv/www/Data.html`
Chen & Zimmermann (2021)	`https://www.openassetpricing.com/`

Table C.2: Summary Statistics of the Factors

This table shows the descriptive statistics of the factors used in this chapter. The factors come from seven major asset classes (see the list in Table 4.1). We present the summary statistics of the factors for their common sample period from August 2006 to December 2019. The t-statistics (in parentheses) are based on Newey & West (1987) standard errors with four lags. In addition to the (monthly) mean return, we also present the median, standard deviation (SD), skewness, and kurtosis of the monthly factor returns.

	Mean	(t-statistic)	Median	SD	Skew	Kurt
U.S. Equities						
MKT_useq	0.78**	(2.08)	1.36	4.28	−0.77	4.66
SMB_useq	0.02	(0.13)	0.18	2.37	0.30	3.05
HML_useq	−0.25	(−1.14)	−0.41	2.65	0.16	5.03
RMW_useq	0.25*	(1.91)	0.27	1.56	0.11	3.28
CMA_useq	0.01	(0.10)	0.00	1.46	0.30	2.68
MOM_useq	0.03	(0.07)	0.22	4.65	−2.57	20.5
STR_useq	0.12	(0.55)	−0.06	2.99	−0.02	5.99
LTR_useq	−0.26	(−1.17)	−0.36	2.54	0.17	3.04
ME_useq	0.04	(0.25)	0.20	2.40	0.21	3.04
IA_useq	−0.01	(−0.11)	−0.18	1.54	0.20	2.57
ROE_useq	0.24	(1.16)	0.47	2.23	−1.60	12.6
EG_useq	0.41**	(2.58)	0.27	1.80	−0.11	5.33
PEAD_useq	0.19	(1.16)	0.21	2.09	−0.25	5.64
FIN_useq	0.28	(1.30)	−0.02	2.82	0.55	3.85
LIQ_useq	−0.07	(−0.24)	0.14	3.46	−0.46	4.90
HMLm_useq	−0.13	(−0.40)	−0.52	3.71	2.48	19.3
BAB_useq	0.40*	(1.76)	0.55	2.64	−0.17	7.01
QMJ_useq	0.42*	(1.73)	0.29	2.49	0.24	4.53
SMB_useqsy	0.14	(0.86)	0.15	2.16	0.25	3.01
MGMT_useq	0.14	(0.81)	0.04	2.05	0.03	4.33
PERF_useq	0.59	(1.42)	0.39	4.69	−0.43	5.97
International Equities						
MKT_inteq	0.42	(0.88)	0.65	4.96	−0.78	5.75
SMB_inteq	0.20	(1.43)	0.24	1.52	0.38	4.94
HML_inteq	0.17	(1.40)	0.17	1.45	−0.02	2.93
HMLm_inteq	0.16	(1.01)	−0.11	1.87	0.62	5.20
RMW_inteq	0.26***	(3.08)	0.17	1.10	−0.05	4.11
ROE_inteq	0.23**	(2.25)	0.10	1.19	0.18	3.68
CMA_inteq	0.23	(1.64)	0.19	1.25	0.78	8.88
MOM_inteq	0.46*	(1.69)	0.74	2.71	−2.34	16.6
iMOM_inteq	0.41***	(4.00)	0.52	1.16	−0.13	3.65
MGMT_inteq	0.25**	(2.07)	0.06	1.29	0.97	5.34
PERF_inteq	0.43***	(2.82)	0.53	1.75	−0.85	6.01
BAB_inteq	0.36	(1.30)	0.38	3.03	−0.21	5.08
QMJ_inteq	0.48***	(2.75)	0.53	1.92	−0.32	5.35

continued on the next page

Table C.2: Summary Statistics of the Factors (continued)

	Mean	(t-statistic)	Median	SD	Skew	Kurt
Corporate Bonds						
MKT_cb	0.39***	(3.31)	0.41	1.36	0.15	10.5
TERM_cb	0.50**	(2.02)	0.23	3.27	0.47	5.35
DEF_cb	0.07	(0.44)	0.07	2.19	−0.47	6.65
DRF_cb	0.69***	(2.90)	0.59	2.36	0.87	8.06
LRF_cb	0.47***	(2.78)	0.28	1.39	3.81	29.5
CRF_cb	0.38*	(1.86)	0.26	1.91	−0.32	7.96
Carry_cb	0.97***	(4.86)	1.00	2.12	0.97	7.11
Value_cb	0.71***	(5.82)	0.81	1.37	−0.33	5.79
DUR_cb	0.55**	(2.46)	0.60	2.66	−0.00	7.72
VOL_cb	0.55***	(2.85)	0.57	2.32	0.38	7.55
STR_cb	0.35***	(2.74)	0.41	1.33	0.45	6.64
MOM_cb	−0.36**	(−2.40)	−0.19	1.65	−2.92	20.5
MOM_cbeq	0.22***	(3.78)	0.14	0.84	−0.05	8.61
LTR_cb	0.07	(0.46)	−0.09	1.73	1.80	12.6
VIX_cb	0.12	(1.64)	0.06	0.71	2.18	16.2
UNC_cb	−0.01	(−0.12)	0.03	1.20	−1.46	11.9
EPU_cb	0.09	(1.14)	0.13	0.84	−1.14	8.43
EPUtax_cb	0.02	(0.24)	0.05	0.65	−1.60	11.0
Commodities						
MKT_cm	−0.09	(−0.22)	0.20	4.47	−0.58	6.20
Value_cm	0.98**	(2.60)	1.10	4.79	−0.02	3.32
MOM_cm	−0.11	(−0.29)	0.50	4.71	−0.60	3.49
Carry_cm	0.43	(1.51)	0.29	4.08	−0.52	4.39
TSMOM_cm	0.11	(0.29)	0.25	4.60	−0.20	5.25
Currencies						
MKT_fx	0.03	(0.22)	0.19	1.88	−0.42	4.32
HML_fx	0.30	(1.65)	0.41	2.23	−0.52	3.44
Carry_fx	0.33*	(1.82)	0.28	2.26	0.69	3.91
Dollar_fx	0.07	(0.43)	−0.08	1.90	0.59	4.53
Value_fxaqr	0.29**	(2.30)	0.08	1.62	0.34	4.23
MOM_fxaqr	−0.13	(−0.95)	−0.02	2.12	−0.41	3.34
Carry_fxaqr	0.19	(0.85)	0.35	2.52	−0.41	5.28
TSMOM_fx	0.55	(1.35)	0.52	5.20	0.75	6.00
Equity Indices						
MKT_eqi	0.47	(1.18)	1.09	4.09	−0.97	5.81
Value_eqi	−0.11	(−0.76)	−0.01	2.02	−0.20	3.07
MOM_eqi	0.22	(1.33)	0.06	2.36	0.27	2.77
Carry_eqi	−0.24	(−1.53)	−0.24	1.96	0.53	3.83
Defensive_eqi	0.24	(1.54)	0.11	1.98	0.10	3.31
TSMOM_eqi	0.77	(1.24)	0.70	7.07	0.01	3.29
Government Bonds						
MKT_govtb	0.32***	(3.06)	0.36	1.32	0.06	3.13
Value_govtb	0.06	(0.69)	0.04	1.13	−0.69	7.41
MOM_govtb	0.12	(1.19)	0.09	1.33	0.58	5.42
Carry_govtb	0.15*	(1.73)	0.15	1.15	−1.17	11.6
Defensive_govtb	0.05	(0.76)	0.14	0.88	−0.81	4.85
TSMOM_govtb	1.35**	(2.14)	0.52	7.67	0.55	3.82

Table C.3: Detailed Factor Identification Protocol Results

This table shows the results of the first-step factor identification protocol. We regress each of the canonical variates on a constant and the set of candidate factors. For each factor, we report the average t-statistic from regressions with all canonical variates and the average t-statistic from regressions with those variates associated with significant (at 5%) canonical correlations. We also report the number of significant t-statistics in the two halves of our sample period. The requirement for passing the first step of the factor identification protocol is that the average of the absolute t-statistics associated with the significant canonical correlations exceeds 1.96 and the average number of significant t-statistics over the two periods is more than 25% of the number of canonical variates (Pukthuanthong et al., 2019). We mark with a "✓" those factors that pass the first-stage screening implied by these two conditions. The factors that fail this step of the factor identification protocol are marked with a "–".

Panel A: U.S. Equities

	MKT_useq	SMB_useq	HML_useq	RMW_useq	CMA_useq	MOM_useq	STR_useq	LTR_useq
Avg. t-stat	17.5	2.33	2.64	1.06	1.18	1.23	1.62	1.41
Avg. t-stat sig.	24.8	2.81	3.26	1.34	1.23	1.39	1.74	1.77
Sig. t-stats first half	7.0	2.0	6.0	1.0	2.0	1.0	4.0	3.0
Sig. t-stats second half	4.0	6.0	4.0	3.0	2.0	2.0	3.0	3.0
Avg.	5.5	4.0	5.0	2.0	2.0	1.5	3.5	3.0
Pass first step?	✓	✓	✓	–	–	–	–	–
	ME_useq	IA_useq	ROE_useq	EG_useq	PEAD_useq	FIN_useq	LIQ_useq	HMLm_useq
Avg. t-stat	1.20	1.00	1.19	1.59	1.06	1.11	1.21	1.78
Avg. t-stat sig.	1.25	0.96	1.33	1.79	1.10	1.30	1.39	2.26
Sig. t-stats first half	1.0	2.0	2.0	2.0	3.0	3.0	3.0	1.0
Sig. t-stats second half	3.0	1.0	2.0	2.0	2.0	2.0	2.0	4.0
Avg.	2.0	1.5	2.0	2.0	2.5	2.5	2.5	2.5
Pass first step?	–	–	–	–	–	–	–	–
	BAB_useq	QMJ_useq	SMB_useqsy	MGMT_useq	PERF_useq			
Avg. t-stat	1.70	1.70	1.23	1.35	1.70			
Avg. t-stat sig.	1.96	2.09	1.46	1.57	2.07			
Sig. t-stats first half	4.0	3.0	0.0	3.0	3.0			
Sig. t-stats second half	2.0	3.0	3.0	0.0	2.0			
Avg.	3.0	3.0	1.5	1.5	2.5			
Pass first step?	✓	✓	–	–	–			

continued on the next page

Table 3.3: Detailed Factor Identification Protocol Results (continued)

Panel B: International Equities

	MKT_inteq	SMB_inteq	HML_inteq	HMLm_inteq	RMW_inteq	ROE_inteq	CMA_inteq	MOM_inteq
Avg. t-stat	7.11	4.31	1.45	1.17	1.18	1.00	2.25	3.47
Avg. t-stat sig.	8.83	5.26	1.66	1.26	1.34	1.12	2.69	4.17
Sig. t-stats first half	6.0	3.0	4.0	1.0	1.0	3.0	6.0	6.0
Sig. t-stats second half	7.0	6.0	2.0	2.0	3.0	2.0	4.0	2.0
Avg.	6.5	4.5	3.0	1.5	2.0	2.5	5.0	4.0
Pass first step?	✓	✓	–	–	–	–	✓	✓
	iMOM_inteq	MGMT_inteq	PERF_inteq	BAB_inteq	QMJ_inteq			
Avg. t-stat	1.37	2.51	1.48	3.94	3.49			
Avg. t-stat sig.	1.51	2.96	1.72	4.71	4.19			
Sig. t-stats first half	6.0	4.0	4.0	5.0	6.0			
Sig. t-stats second half	1.0	4.0	3.0	7.0	5.0			
Avg.	3.5	4.0	3.5	6.0	5.5			
Pass first step?	–	✓	–	✓	✓			

Panel C: Corporate Bonds

	MKT_cb	TERM_cb	DEF_cb	DRF_cb	LRF_cb	CRF_cb	Carry_cb	Value_cb
Avg. t-stat	4.45	2.29	1.87	1.74	1.62	1.00	2.75	1.57
Avg. t-stat sig.	4.90	2.52	2.01	1.90	1.71	1.07	2.93	1.71
Sig. t-stats first half	7.0	6.0	5.0	3.0	6.0	0.0	3.0	3.0
Sig. t-stats second half	7.0	4.0	3.0	3.0	1.0	2.0	6.0	2.0
Avg.	7.0	5.0	4.0	3.0	3.5	1.0	4.5	2.5
Pass first step?	✓	✓	✓	–	–	–	✓	–
	DUR_cb	VOL_cb	STR_cb	MOM_cb	MOM_cbeq	LTR_cb	VIX_cb	UNC_cb
Avg. t-stat	3.58	2.01	1.19	2.27	2.38	1.99	2.25	1.92
Avg. t-stat sig.	3.95	2.15	1.22	2.37	2.59	2.11	2.39	2.02
Sig. t-stats first half	4.0	3.0	2.0	3.0	5.0	3.0	2.0	2.0
Sig. t-stats second half	5.0	4.0	1.0	5.0	3.0	5.0	7.0	6.0
Avg.	4.5	3.5	1.5	4.0	4.0	4.0	4.5	4.0
Pass first step?	✓	✓	–	✓	✓	✓	✓	✓
	EPU_cb	EPUtax_cb						
Avg. t-stat	1.06	1.27						
Avg. t-stat sig.	1.08	1.37						
Sig. t-stats first half	2.0	2.0						
Sig. t-stats second half	2.0	2.0						
Avg.	2.0	2.0						
Pass first step?	–	–						

continued on the next page

Table 3.3: Detailed Factor Identification Protocol Results (continued)

Panel D: Commodities

	MKT_cm	Value_cm	MOM_cm	Carry_cm	TSMOM_cm
Avg. t-stat	12.1	1.99	1.20	2.31	1.96
Avg. t-stat sig.	19.4	2.35	0.95	3.00	2.49
Sig. t-stats first half	2.0	2.0	2.0	3.0	3.0
Sig. t-stats second half	1.0	2.0	0.0	1.0	1.0
Avg.	1.5	2.0	1.0	2.0	2.0
Pass first step?	✓	✓	–	✓	✓

Panel E: Currencies

	MKT_fx	HML_fx	Carry_fx	Dollar_fx	Value_fxaqr	MOM_fxaqr	Carry_fxaqr	TSMOM_fx
Avg. t-stat	1.75	1.44	0.87	1.10	1.33	3.05	2.16	2.07
Avg. t-stat sig.	2.94	2.02	0.87	1.62	1.83	5.43	3.57	3.22
Sig. t-stats first half	3.0	1.0	0.0	2.0	3.0	3.0	3.0	1.0
Sig. t-stats second half	2.0	2.0	0.0	1.0	0.0	5.0	3.0	2.0
Avg.	2.5	1.5	0.0	1.5	1.5	4.0	3.0	1.5
Pass first step?	✓	–	–	–	–	✓	✓	–

Panel F: Equity Indexes

	MKT_eqi	Value_eqi	MOM_eqi	Carry_eqi	Defensive_eqi	TSMOM_eqi
Avg. t-stat	5.71	1.88	1.17	1.78	1.95	1.36
Avg. t-stat sig.	11.0	2.91	1.34	2.28	3.55	1.37
Sig. t-stats first half	2.0	2.0	1.0	3.0	3.0	2.0
Sig. t-stats second half	1.0	1.0	1.0	2.0	1.0	2.0
Avg.	1.5	1.5	1.0	2.5	2.0	2.0
Pass first step?	–	–	–	✓	✓	–

Panel G: Government Bonds

	MKT_govtb	Value_govtb	MOM_govtb	Carry_govtb	Defensive_govtb	TSMOM_govtb
Avg. t-stat	4.49	1.45	2.00	1.86	1.47	1.34
Avg. t-stat sig.	8.44	2.00	2.17	2.67	1.84	1.63
Sig. t-stats first half	2.0	2.0	1.0	3.0	3.0	1.0
Sig. t-stats second half	2.0	1.0	2.0	1.0	1.0	2.0
Avg.	2.0	1.5	1.5	2.0	2.0	1.5
Pass first step?	✓	–	–	✓	–	–

Table C.4: Detailed Model Scan Results

This table summarizes the main results of the model scan algorithm. For each asset class, we examine all possible asset pricing models that can be formed, subject to the restrictions on factor correlations. The model scan applies the BS–CZZ approach to all candidate factors that survive the first step of factor identification. Finally, we identify the optimal factor model across asset classes from the pool of factors in the best models for each asset class. We report the log marginal likelihoods ($\log \tilde{m}(y|M_j)$), posterior probabilities ($Pr(M_j|y)$), and ratios of posterior to prior probabilities ($\frac{Pr(M_j|y)}{Pr(M_j)}$) of the top models. We report the top three models in each asset class.

Risk Factors	$\log \tilde{m}(y\|M_j)$	$Pr(M_j\|y)$	$\frac{Pr(M_j\|y)}{Pr(M_j)}$
A. U.S. Equities			
MKT_useq, QMJ_useq	1,591.56	57.3	17.8
MKT_useq, BAB_useq, QMJ_useq	1,590.55	20.9	6.47
MKT_useq, SMB_useq, QMJ_useq	1,589.51	7.37	2.28
B. International Equities			
MKT_inteq, SMB_inteq, MGMT_inteq, QMJ_inteq	2,739.61	54.8	69.6
MKT_inteq, SMB_inteq, MOM_inteq, MGMT_inteq, QMJ_inteq	2,738.39	20.1	15.9
MKT_inteq, SMB_inteq, CMA_inteq, MGMT_inteq, QMJ_inteq	2,737.29	6.65	5.25
C. Corporate Bonds			
TERM_cb, Carry_cb, DUR_cb, MOM_cbeq	4,833.81	15.6	160
Carry_cb, MOM_cbeq	4,832.98	6.82	69.7
DEF_uscb, Carry_cb, MOM_cbeq	4,832.73	5.32	54.5
D. Commodities			
Value_cm, Carry_cm	968.838	40.0	6.00
Value_cm	968.614	32.0	4.80
Value_cm, Carry_cm, TSMOM_cm	966.917	5.86	0.88
E. Currencies			
MOMfx_aqr	1,074.92	34.3	2.40
Carryfx_aqr	1,074.73	28.4	1.99
MKT_fx	1,074.63	25.6	1.79
F. Equity Indices			
Carry_eqi	724.408	67.3	2.02
Defensive_eqi	723.341	23.2	0.69
Carry_eqi, Defensive_eqi	722.451	9.52	0.29
G. Government Bonds			
MKT_govtb	849.844	82.4	2.47
MKT_govtb, Carry_govtb	848.157	15.3	0.46
Carry_govtb	846.281	2.34	0.07

continued on the next page

Table C.4: Detailed Model Scan Result (continued)

Risk Factors	$\log \tilde{m}(y\|M_j)$	$Pr(M_j\|y)$	$\frac{Pr(M_j\|y)}{Pr(M_j)}$
H. Top models across asset classes			
MKT_useq, SMB_inteq, MGMT_inteq, QMJ_inteq, Carry_cb, MOM_cbeq, MOMfx_aqr, Carry_eqi	5,406.31	3.82	705
MKT_useq, SMB_inteq, MGMT_inteq, QMJ_inteq, Carry_cb, DUR_cb, MOM_cbeq, Carry_eqi, MKT_govtb	5,406.25	3.58	660
MKT_useq, SMB_inteq, MGMT_inteq, QMJ_inteq, Carry_cb, MOM_cbeq, Carry_eqi	5,405.91	2.56	471

Chapter 5

Conclusion and Further Research

5.1 Summary and Conclusion

This dissertation investigates the use of factor models to measure performance by investors in corporate bond markets, systematically examines proposed factors in the corporate bond literature to identify an optimal factor model for corporate bond returns, and finally provides a comprehensive analysis of factor pricing and market integration across asset classes.

Chapter 2 comprehensively investigates how corporate bond mutual fund investors measure performance by analyzing the relation between mutual fund flows and different performance measures. Specifically, we run a horse race among different performance measures, ranging from the simple raw return and the Sharpe ratio to alphas estimated by using single and different multi-factor models. Our empirical analysis reveals that the Sharpe ratio explains the net flows into actively managed U.S. corporate bond mutual funds better than any of these alternatives. Morningstar ratings appear to explain an even larger share of investor fund flows, but the Sharpe ratio has important explanatory power within the Morningstar ratings groups. It thus seems that most investors do not use any factor model at all. We point out the potential harmful consequences caused by such investors' reliance of the Sharpe ratio and Morningstar ratings as primary performance measures.

Chapter 3 systematically examines a large set of the most prominent corporate bond factors in the literature to separate useful factors from redundant ones. First, we check whether the factors systematically move corporate bond prices. We find that many prominent recent factors, such as those of Bai et al. (2019), fail this first screening test for viable factors. In the second step, we adopt a Bayesian marginal likelihood-based approach proposed by Barillas & Shanken (2018) and Chib et al. (2020) to simultaneously compare all possible models that can be formed as subsets of the factors that pass the first step. The main finding that emerges from our analysis is that the best factor model for corporate bond returns is based on the combination of carry, duration, stock momentum, and term structure factors. The result indicates that only a small subset of the 23 considered factors really matters for corporate bond pricing. We show the outperformance of the optimal model from the Bayesian model scan relative to the prominent existing factor models. Further analysis shows the winning model overall explains reasonably well the time-series and cross-sectional variation of corporate bond returns (represented by various test assets). This model can be used as a benchmark model for future research, for investors in corporate bond markets to implement factor-investing strategies, and to evaluate performance.

Chapter 4 studies factor pricing across asset classes and examines market integration from a multi-asset, multi-factor perspective, moving beyond the convention of analyzing different asset classes separately. We find that single-asset-class-specialized models typically fail to explain factors from other asset classes, indicating that the different asset classes are far from perfectly integrated. However, there are also some cross-market linkages. We thus uncover and analyze an integrated factor model that can effectively explain returns across asset classes and provide a useful benchmark for multi-asset, multi-factor investing. Of the 77 factors across the seven different asset classes, the U.S. stock market, the size, management and quality factors of international equities, and the corporate bond carry factor appear to be the most important components of an optimal unified model. An integrated model consisting of these five factors, plus the corporate bond equity momentum factor, the currency momentum factor, and the equity index carry factor, has reasonably good performance across asset classes. It achieves higher in-sample and out-of-sample Sharpe ratios than any existing models that specialize in one asset class. Finally, the top integrated model subsumes a long list of factors and performs well in pricing assets across different asset classes.

5.2 Suggestions for Further Research

This section suggests several potentially interesting research avenues that are not yet explored within the scope of the three studies in this dissertation and left for future work.

Over the past 20 years, academic research, and the environmental and social trend toward "responsible" investing in financial markets have had a significant influence on the investment strategies and tastes of investors. Factor investing and more quantitative approaches have been widely adopted. A growing number of large institutional investors have declared to make environmental, social, and governance (ESG) performance an important objective in their investment decision-making process.

Given the growing interest of investors into ESG financial products, the investigation in Chapter 2 can be extended to the case of investors engaging in ESG. How do they evaluate and trade off financial and non-financial performance when making investment decisions in bond markets? In particular, which performance measures do they currently use and which should they use? Future work on these questions could yield helpful insights, as to date the lack of consistency and comparability of ESG ratings from different rating providers creates a barrier for the proper integration of sustainability concerns into investment decisions, as well as for the accuracy of the performance assessment in both financial and sustainable aspects of funds engaging in ESG. On the one hand, asset managers' decisions on security selection and weighting could be driven not only by the ESG score inputs but also by the choice of rating provider. On the other hand, this issue makes it difficult and confusing for investors to distinguish between outperformers and laggards. Relying on a misleading metric to judge the sustainability of a financial product may lead to negative consequences on financial returns, without actually improving the sustainability of the portfolio.

The existing literature reports widely mixed empirical evidence and a lot of disagreement among views across academics and practitioners about the relationship between ESG and asset returns, and whether ESG enhances or harms financial performance. Using the research setup and methods in Chapters 3 and 4, one can check whether ESG factors add incremental information to explain the cross-section of corporate bond returns, as well as returns across asset classes and markets, or they can be subsumed by other existing factors (such as quality factors, credit risk, tail risk). Such an analysis could provide further useful evidence in this debate, and may establish some common ground to explain the inconclusiveness of existing findings.

Academic studies on asset pricing factors are generally criticized for ignoring real-world concerns, such as transaction costs and trading frictions. Therefore, how best to translate academic factors into realistic factor-based investing strategies across asset classes and markets is a pivotal and practical question. A comprehensive study and comparison of the impact of real-world investment constraints and transaction costs on the implementability and performance of academic factors across asset classes in practice may carry useful implications for the optimal choices related to implementation design and portfolio optimization (such as rebalancing frequency, portfolio sorting, weighting scheme). Investigating how well asset pricing models explain the cross-section of returns taking transaction costs into account could provide different perspectives on the practicality and robustness about model performance. For example, a recent study by Detzel, Novy-Marx, & Velikov (2023) highlights the effect of omitting trading costs when assessing factor models for U.S. stocks. Examining this issue beyond equities for other asset classes such as bonds, currencies and commodities, etc. and in the context of asset allocation across asset classes and markets would be useful to provide a comprehensive insight.

References

Adrian, T., Etula, E., & Muir, T. (2014). Financial intermediaries and the cross-section of asset returns. *Journal of Finance*, *69*(6), 2557–2596. [cited on p. 133]

Agarwal, V., Green, T. C., & Ren, H. (2018). Alpha or beta in the eye of the beholder: What drives hedge fund flows? *Journal of Financial Economics*, *127*(3), 417–434. [cited on p. 14, 15, 32]

Ahmed, S., Bu, Z., & Tsvetanov, D. (2019). Best of the best: A comparison of factor models. *Journal of Financial and Quantitative Analysis*, *54*(4), 1713–1758. [cited on p. 99]

Ambastha, M., Dor, A. B., Dynkin, L., Hyman, J., & Konstantinovsky, V. (2010). Empirical duration of corporate bonds and credit market segmentation. *Journal of Fixed Income*, *20*(1), 5–27. [cited on p. 32]

Aragon, G. O., Li, L., & Qian, J. (2019). The use of credit default swaps by bond mutual funds: Liquidity provision and counterparty risk. *Journal of Financial Economics*, *131*(1), 168–185. [cited on p. 53]

Artavanis, N. T., Eksi, A., & Kadlec, G. B. (2019). Downside risk and mutual fund flows. *Tulane University Working Paper*. [cited on p. 39]

Asness, C., & Frazzini, A. (2013). The devil in HML's details. *Journal of Portfolio Management*, *39*(4), 49–68. [cited on p. 150, 152, 189]

Asness, C. S., Frazzini, A., & Pedersen, L. H. (2019). Quality minus junk. *Review of Accounting Studies*, *24*(1), 34–112. [cited on p. 149, 150, 152, 161, 189]

Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. *Journal of Finance*, *68*(3), 929–985. [cited on p. 145, 149, 154, 155, 168, 181]

Bai, J., Bali, T. G., & Wen, Q. (2019). Common risk factors in the cross-section of corporate bond returns. *Journal of Financial Economics*, *131*(3), 619–642. [cited on p. 14, 20, 21, 51, 84, 94, 95, 98, 99, 103, 108, 110, 129, 136, 138, 139, 140, 141, 142, 151, 152, 154, 189, 198]

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics*, *131*(4), 1593–1636. [cited on p. 139]

Bakshi, G., Gao, X., & Rossi, A. G. (2019). Understanding the sources of risk underlying the cross section of commodity returns. *Management Science*, *65*(2), 619–641. [cited on p. 149, 154]

Bali, T. G., Brown, S. J., & Tang, Y. (2017a). Is economic uncertainty priced in the cross-section of stock returns? *Journal of Financial Economics*, *126*(3), 471–489. [cited on p. 94, 139, 142]

Bali, T. G., Subrahmanyam, A., & Wen, Q. (2017b). Return-based factors for corporate bonds. *UCLA Working Paper*. [cited on p. 94, 98, 99, 103, 139, 140, 142, 151, 152, 154]

Bali, T. G., Subrahmanyam, A., & Wen, Q. (2021a). Long-term reversals in the corporate bond market. *Journal of Financial Economics*, *139*(2), 656–677. [cited on p. 94, 98, 108, 139, 142]

Bali, T. G., Subrahmanyam, A., & Wen, Q. (2021b). The macroeconomic uncertainty premium in the corporate bond market. *Journal of Financial and Quantitative Analysis*, *56*(5), 1653–1678. [cited on p. 94, 98, 139, 142, 151, 152]

Bao, J., Pan, J., & Wang, J. (2011). The illiquidity of corporate bonds. *Journal of Finance*, *66*(3), 911–946. [cited on p. 138]

Barber, B. M., Huang, X., & Odean, T. (2016). Which factors matter to investors? Evidence from mutual fund flows. *Review of Financial Studies*, *29*(10), 2600–2642. [cited on p. 3, 11, 14, 15, 24, 26, 27, 39, 44, 50, 54]

Barillas, F., Kan, R., Robotti, C., & Shanken, J. (2020). Model comparison with Sharpe ratios. *Journal of Financial and Quantitative Analysis*, *55*(6), 1840–1874. [cited on p. 96, 106, 107, 118, 147, 166]

Barillas, F., & Shanken, J. (2017). Which alpha? *Review of Financial Studies*, *30*(4), 1316–1338. [cited on p. 99]

Barillas, F., & Shanken, J. (2018). Comparing asset pricing models. *Journal of Finance*, *73*(2), 715–754. [cited on p. 5, 95, 100, 104, 113, 117, 120, 136, 146, 159, 161, 168, 185, 198]

Bekaert, G., & De Santis, R. A. (2021). Risk and return in international corporate bond markets. *Journal of International Financial Markets, Institutions and Money*, *72*, 101338. [cited on p. 12, 18, 19, 29, 33, 38, 61, 63, 66, 69, 71, 73, 76, 77, 79, 81, 83, 85, 87, 89, 91, 126]

Bektić, D., Wenzler, J.-S., Wegener, M., Schiereck, D., & Spielmann, T. (2019). Extending Fama–French factors to corporate bond markets. *Journal of Portfolio Management*, *45*(3), 141–158. [cited on p. 95, 98, 143]

Ben-David, I., Li, J., Rossi, A., & Song, Y. (2022). What do mutual fund investors really care about? *The Review of Financial Studies*, *35*(4), 1723–1774. [cited on p. 12, 13, 14, 15, 27, 39, 46, 47, 48, 54]

Ben-Rephael, A., Choi, J., & Goldstein, I. (2021). Mutual fund flows and fluctuations in credit and business cycles. *Journal of Financial Economics*, *139*(1), 84–108. [cited on p. 30]

Berk, J. B., & Van Binsbergen, J. H. (2015). Measuring skill in the mutual fund industry. *Journal of Financial Economics*, *118*(1), 1–20. [cited on p. 18]

Berk, J. B., & Van Binsbergen, J. H. (2016). Assessing asset pricing models using revealed preference. *Journal of Financial Economics*, *119*(1), 1–23. [cited on p. xi, 3, 11, 14, 15, 39, 40, 75]

Blocher, J., & Molyboga, M. (2017). The revealed preference of sophisticated investors. *European Financial Management*, *23*(5), 839–872. [cited on p. 14, 15]

Bollen, N. P., & Pool, V. K. (2008). Conditional return smoothing in the hedge fund industry. *Journal of Financial and Quantitative Analysis*, *43*(2), 267–298. [cited on p. 16]

Bollen, N. P., & Pool, V. K. (2009). Do hedge fund managers misreport returns? Evidence from the pooled distribution. *Journal of Finance*, *64*(5), 2257–2288. [cited on p. 15, 16]

Borri, N., & Verdelhan, A. (2011). Sovereign risk premia. *MIT Working Paper*. [cited on p. 149]

Boudoukh, J., Richardson, M., Subrahmanyam, M., & Whitelaw, R. F. (2002). Stale prices and strategies for trading mutual funds. *Financial Analysts Journal*, *58*(4), 53–71. [cited on p. 58]

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, *29*(2), 238–249. [cited on p. 27]

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, *52*(1), 57–82. [cited on p. 6, 154]

Chalmers, J., Edelen, R., & Kadlec, G. (2001). On the perils of security pricing by financial intermediaries: The wildcard option in transacting mutual fund shares. *Journal of Finance*, *56*(6), 2209–2236. [cited on p. 58]

Chen, A. Y., & Zimmermann, T. (2021). Open source cross-sectional asset pricing. *Critical Finance Review, Forthcoming*. [cited on p. 153, 189]

Chen, Q., Goldstein, I., & Jiang, W. (2010a). Payoff complementarities and financial fragility: Evidence from mutual fund outflows. *Journal of Financial Economics*, *97*(2), 239–262. [cited on p. 17, 31, 36]

Chen, Y., Ferson, W., & Peters, H. (2010b). Measuring the timing ability and performance of bond mutual funds. *Journal of Financial Economics*, *98*(1), 72–89. [cited on p. 57]

Chen, Y., & Qin, N. (2016). The behavior of investor flows in corporate bond mutual funds. *Management Science*, *63*(5), 1365–1381. [cited on p. 17, 32, 36]

Chen, Z., Lookman, A. A., Schürhoff, N., & Seppi, D. J. (2014). Rating-based investment practices and bond market segmentation. *Review of Asset Pricing Studies*, *4*(2), 162–205. [cited on p. 32]

Chib, S., & Zeng, X. (2020). Which factors are risk factors in asset pricing? A model scan framework. *Journal of Business & Economic Statistics*, *38*(4), 771–783. [cited on p. 105]

Chib, S., Zeng, X., & Zhao, L. (2020). On comparing asset pricing models. *Journal of Finance*, *75*(1), 551–577. [cited on p. 5, 95, 104, 113, 117, 136, 144, 146, 159, 161, 185, 186, 198]

Choi, J., Hoseinzade, S., Shin, S. S., & Tehranian, H. (2020). Corporate bond mutual funds and asset fire sales. *Journal of Financial Economics*, *138*(2), 432–457. [cited on p. 36]

Choi, J., & Kim, Y. (2018). Anomalies and market (dis) integration. *Journal of Monetary Economics*, *100*, 16–34. [cited on p. 4, 97, 125, 149]

Choi, J., Kronlund, M., & Oh, J. Y. J. (2021). Sitting bucks: Stale pricing in fixed income funds. *Journal of Financial Economics, Forthcoming*. [cited on p. 15, 23, 58]

Chordia, T., Goyal, A., Nozowa, Y., Subrahmanyam, A., & Tong, Q. (2017). Are capital market anomalies common to equity and corporate bond markets? *Journal of Financial and Quantitative Analysis*, *52*(4), 1301. [cited on p. 4, 97, 125]

Chung, K. H., Wang, J., & Wu, C. (2019). Volatility and the cross-section of corporate bond returns. *Journal of Financial Economics*, *133*(2), 397–417. [cited on p. 51, 84, 94, 98, 140, 143, 151, 152]

Cici, G., Gibson, S., & Merrick Jr, J. J. (2011). Missing the marks? Dispersion in corporate bond valuations across mutual funds. *Journal of Financial Economics*, *101*(1), 206–226. [cited on p. 15, 55, 57]

Cochrane, J. (2009). *Asset pricing: Revised edition*. Princeton University Press. [cited on p. 6, 145]

Cochrane, J. H. (2005). *Asset Pricing*. Princeton, NJ: Princeton University Press. [cited on p. 133]

Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance*, *66*(4), 1047–1108. [cited on p. 6]

Connor, G., & Korajczyk, R. A. (1988). Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics*, *21*(2), 255–289. [cited on p. 103, 158]

Cooper, I., Mitrache, A., & Priestley, R. (2022). A global macroeconomic risk model for value, momentum, and other asset classes. *Journal of Financial and Quantitative Analysis*, *57*(1), 1–30. [cited on p. 149]

Correia, M., Richardson, S., & Tuna, İ. (2012). Value investing in credit markets. *Review of Accounting Studies*, *17*(3), 572–609. [cited on p. 139]

Daniel, K., Hirshleifer, D., & Sun, L. (2020). Short-and long-horizon behavioral factors. *Review of Financial Studies*, *33*(4), 1673–1736. [cited on p. 100, 150, 152, 154, 189]

Del Guercio, D., & Tkac, P. A. (2002). The determinants of the flow of funds of managed portfolios: Mutual funds vs. pension funds. *Journal of Financial and Quantitative Analysis*, *37*(4), 523–557. [cited on p. 31]

Del Guercio, D., & Tkac, P. A. (2008). Star power: The effect of Monrningstar ratings on mutual fund flow. *Journal of Financial and Quantitative Analysis*, *43*(4), 907–936. [cited on p. 17, 26, 27]

Detzel, A., Novy-Marx, R., & Velikov, M. (2023). Model comparison with transaction costs. *Journal of Finance*. [cited on p. 201]

Dick-Nielsen, J., Feldhütter, P., & Lando, D. (2012). Corporate bond liquidity before and after the onset of the subprime crisis. *Journal of Financial Economics*, *103*(3), 471–492. [cited on p. 133]

Elton, E. J., & Gruber, M. J. (2013). Mutual funds. In *Handbook of the Economics of Finance*, vol. 2, (pp. 1011–1061). Elsevier. [cited on p. 18]

Elton, E. J., Gruber, M. J., Agrawal, D., & Mann, C. (2001). Explaining the rate spread on corporate bonds. *Journal of Finance*, *56*(1), 247–277. [cited on p. 19]

Elton, E. J., Gruber, M. J., & Blake, C. R. (1995). Fundamental economic variables, expected returns, and bond fund performance. *Journal of Finance*, *50*(4), 1229–1256. [cited on p. 12, 15, 19, 29, 33, 38, 61, 63, 66, 69, 71, 73, 74, 75, 76, 77, 79, 81, 83, 85, 87, 89, 91]

Evans, R. B. (2010). Mutual fund incubation. *Journal of Finance*, *65*(4), 1581–1611. [cited on p. 18]

Evans, R. B., & Fahlenbrach, R. (2012). Institutional investors and mutual fund governance: Evidence from retail–institutional fund twins. *Review of Financial Studies*, *25*(12), 3530–3571. [cited on p. 31]

Evans, R. B., & Sun, Y. (2021). Models or stars: The role of asset pricing models and heuristics in investor risk adjustment. *Review of Financial Studies*, *34*(1), 67–107. [cited on p. 27, 46]

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56. [cited on p. 2, 6, 12, 19, 46, 48, 51, 54, 84, 94, 98, 103, 108, 113, 141, 142, 150, 151, 152, 154]

Fama, E. F., & French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics*, *43*(2), 153–193. [cited on p. 127, 129, 131]

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22. [cited on p. 2, 6, 29, 33, 38, 52, 61, 63, 66, 69, 71, 73, 76, 77, 79, 81, 83, 84, 85, 87, 89, 91, 95, 98, 99, 110, 143, 154, 163]

Fama, E. F., & French, K. R. (2018). Choosing factors. *Journal of Financial Economics*, *128*(2), 234–252. [cited on p. 98, 99, 150, 152, 154]

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, *81*(3), 607–636. [cited on p. 14, 27, 47, 48, 49, 76]

Fan, Z., Londono, J. M., & Xiao, X. (2022). Equity tail risk and currency risk premiums. *Journal of Financial Economics*, *143*(1), 484–503. [cited on p. 149]

Feng, G., Giglio, S., & Xiu, D. (2020). Taming the factor zoo: A test of new factors. *Journal of Finance*, *75*(3), 1327–1370. [cited on p. 100]

Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, *111*(1), 1–25. [cited on p. 149, 150, 152, 189]

Friewald, N., & Nagler, F. (2019). Over-the-counter market frictions and yield spread changes. *Journal of Finance*, *74*(6), 3217–3257. [cited on p. 133]

Galvani, V., & Li, L. (2020). Outliers and momentum in the corporate bond market. *University of Alberta Working Paper*. [cited on p. 108]

Gebhardt, W. R., Hvidkjaer, S., & Swaminathan, B. (2005a). The cross-section of expected corporate bond returns: Betas or characteristics? *Journal of Financial Economics*, *75*(1), 85–114. [cited on p. 19, 108, 116, 140]

Gebhardt, W. R., Hvidkjaer, S., & Swaminathan, B. (2005b). Stock and bond market interaction: Does momentum spill over? *Journal of Financial Economics*, *75*(3), 651–690. [cited on p. 20]

Getmansky, M., Lo, A. W., & Makarov, I. (2004). An econometric model of serial correlation and illiquidity in hedge fund returns. *Journal of Financial Economics*, *74*(3), 529–609. [cited on p. 14, 15, 55, 56, 57]

Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, *57*(5), 1121–1152. [cited on p. 96, 123, 124, 127, 148, 157, 173, 178]

Goetzmann, W., Ingersoll, J., Spiegel, M., & Welch, I. (2007). Portfolio performance manipulation and manipulation-proof performance measures. *Review of Financial Studies*, *20*(5), 1503–1546. [cited on p. 14, 18, 55, 56]

Goetzmann, W. N., Ivković, Z., & Rouwenhorst, K. G. (2001). Day trading international mutual funds: Evidence and policy solutions. *Journal of Financial and Quantitative Analysis*, *36*(3), 287–309. [cited on p. 58]

Goldstein, I., Jiang, H., & Ng, D. T. (2017). Investor flows and fragility in corporate bond funds. *Journal of Financial Economics*, *126*(3), 592–613. [cited on p. 13, 17, 18, 36, 52]

Gospodinov, N., & Robotti, C. (2021). Common pricing across asset classes: Empirical evidence revisited. *Journal of Financial Economics*, *140*(1), 292–324. [cited on p. 113, 149, 151, 153]

Green, J., Hand, J. R., & Zhang, X. F. (2017). The characteristics that provide independent information about average U.S. monthly stock returns. *Review of Financial Studies*, *30*(12), 4389–4436. [cited on p. 95]

Greene, J. T., & Hodges, C. W. (2002). The dilution impact of daily fund flows on open-end mutual funds. *Journal of Financial Economics*, *65*(1), 131–158. [cited on p. 58]

Hanauer, M. X. (2020). A comparison of global factor models. *TU Munich Working Paper*. [cited on p. 151, 152, 189]

Hansen, L. P., & Jagannathan, R. (1991). Implications of security market data for models of dynamic economies. *Journal of Political Economy*, *99*(2), 225–262. [cited on p. 6]

Harvey, C. R., & Liu, Y. (2021). Lucky factors. *Journal of Financial Economics*, *141*(2), 413–435. [cited on p. 100]

Harvey, C. R., Liu, Y., & Zhu, H. (2016). ... and the cross-section of expected returns. *Review of Financial Studies*, *29*(1), 5–68. [cited on p. 2, 95]

He, Z., Kelly, B., & Manela, A. (2017). Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics*, *126*(1), 1–35. [cited on p. 133, 149, 151]

He, Z., Khorrami, P., & Song, Z. (2022). Commonality in credit spread changes: Dealer inventory and intermediary distress. *Review of Financial Studies, Forthcoming*. [cited on p. 133]

He, Z., & Krishnamurthy, A. (2013). Intermediary asset pricing. *American Economic Review*, *103*(2), 732–70. [cited on p. 133]

Hollstein, F. (2022). Local, regional, or global asset pricing? *Journal of Financial and Quantitative Analysis*, *57*(1), 291–320. [cited on p. 163]

Hou, K., Mo, H., Xue, C., & Zhang, L. (2019). Which factors? *Review of Finance*, *23*(1), 1–35. [cited on p. 100]

Hou, K., Mo, H., Xue, C., & Zhang, L. (2021). An augmented q-factor model with expected growth. *Review of Finance*, *25*(1), 1–41. [cited on p. 150, 152, 154, 189]

Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. *Review of Financial Studies*, *28*(3), 650–705. [cited on p. 52, 84, 99, 150, 152, 154, 189]

Hou, K., Xue, C., & Zhang, L. (2020). Replicating anomalies. *Review of Financial Studies*, *33*(5), 2019–2133. [cited on p. 95]

Hu, G. X., Pan, J., & Wang, J. (2013). Noise as information for illiquidity. *Journal of Finance*, *68*(6), 2341–2382. [cited on p. 133]

Huang, J. C., Wei, K. D., & Yan, H. (2012). Investor learning and mutual fund flows. *University of Texas Working Paper*. [cited on p. 17]

Hwang, S., & Rubesam, A. (2020). Bayesian selection of asset pricing factors using individual stocks. *Journal of Financial Econometrics*, *20*(4), 716–761. [cited on p. 100]

Ilmanen, A., Israel, R., Moskowitz, T. J., Thapar, A. K., & Lee, R. (2021). How do factor premia vary over time? A century of evidence. *Yale University Working Paper*. [cited on p. 151, 152, 154, 189]

Israel, R., Palhares, D., & Richardson, S. A. (2018). Common factors in corporate bond returns. *Journal of Investment Management*, *16*(2), 17–46. [cited on p. 94, 95, 96, 98, 103, 108, 136, 138, 139, 141, 142, 154]

James, C., & Karceski, J. (2006). Investor monitoring and differences in mutual fund performance. *Journal of Banking & Finance*, *30*(10), 2787–2808. [cited on p. 32]

Jensen, T. I., Kelly, B. T., & Pedersen, L. H. (2022). Is there a replication crisis in finance? *Journal of Finance, Forthcoming*. [cited on p. 151, 152, 153, 177, 189]

Jiang, G. J., & Yuksel, H. Z. (2017). What drives the "smart-money" effect? Evidence from investors' money flow to mutual fund classes. *Journal of Empirical Finance*, *40*, 39–58. [cited on p. 17]

Jiang, H., Li, D., & Wang, A. (2021). Dynamic liquidity management by corporate bond mutual funds. *Journal of Financial and Quantitative Analysis*, *56*(5), 1622–1652. [cited on p. 36]

Jostova, G., Nikolova, S., Philipov, A., & Stahel, C. W. (2013). Momentum in corporate bond returns. *Review of Financial Studies*, *26*(7), 1649–1693. [cited on p. 20, 51, 94, 108, 139, 140]

Jurado, K., Ludvigson, S. C., & Ng, S. (2015). Measuring uncertainty. *American Economic Review*, *105*(3), 1177–1216. [cited on p. 139]

Kan, R., Robotti, C., & Shanken, J. (2013). Pricing model performance and the two-pass cross-sectional regression methodology. *Journal of Finance*, *68*(6), 2617–2649. [cited on p. 129, 131]

Kan, R., Wang, X., & Zheng, X. (2022). In-sample and out-of-sample Sharpe ratios of multi-factor asset pricing models. *University of Toronto Working Paper*. [cited on p. 120]

Kandel, S., & Stambaugh, R. F. (1995). Portfolio inefficiency and the cross-section of expected returns. *Journal of Finance*, *50*(1), 157–184. [cited on p. 130]

Karolyi, G. A., & Wu, Y. (2021). Is currency risk priced in global equity markets? *Review of Finance*, *25*(3), 863–902. [cited on p. 149]

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. [cited on p. 116]

Kelly, B., Palhares, D., & Pruitt, S. (2023). Modeling corporate bond returns. *Journal of Finance*, *78*(4), 1967–2008. [cited on p. 94, 96, 98, 103, 108, 126, 137, 139, 140, 141, 143, 149, 151, 152, 154, 189]

Kelly, B. T., & Pruitt, S. (2022). Reconciling TRACE bond returns. *Yale Working Paper*. [cited on p. 101, 103, 126, 127, 131, 153, 189]

Kim, S. H. (2021). Do mutual funds manipulate star ratings? Evidence from portfolio pumping. *University of Texas Working Paper*. [cited on p. 16]

Koijen, R. S., Moskowitz, T. J., Pedersen, L. H., & Vrugt, E. B. (2018). Carry. *Journal of Financial Economics*, *127*(2), 197–225. [cited on p. 117, 145, 149, 161]

Korsaye, S. A., Trojani, F., & Vedolin, A. (2023). The global factor structure of exchange rates. *Journal of Financial Economics*, *148*(1), 21–46. [cited on p. 149]

Kozak, S., Nagel, S., & Santosh, S. (2018). Interpreting factor models. *Journal of Finance*, *73*(3), 1183–1223. [cited on p. 6]

Lee, K. (2022). Which uncertainty measures matter for the cross-section of corporate bond returns? Evidence from the US during 1973–2020. *Finance Research Letters*, *48*, 102913. [cited on p. 94, 95, 98, 139, 142, 151, 152]

Lewellen, J., Nagel, S., & Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics*, *96*(2), 175–194. [cited on p. 99, 126, 130, 177]

Lin, H., Wang, J., & Wu, C. (2011). Liquidity risk and expected corporate bond returns. *Journal of Financial Economics*, *99*(3), 628–650. [cited on p. 19, 20]

Linnainmaa, J. T., & Roberts, M. R. (2018). The history of the cross-section of stock returns. *Review of Financial Studies*, *31*(7), 2606–2649. [cited on p. 95]

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *Journal of Finance*, *20*(4), 587–615. [cited on p. 1]

Ljung, G. M., & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303. [cited on p. 105]

Ludvigson, S. C., & Ng, S. (2009). Macro factors in bond risk premia. *Review of Financial Studies*, *22*(12), 5027–5067. [cited on p. 14, 51, 84]

Lustig, H., Roussanov, N., & Verdelhan, A. (2011). Common risk factors in currency markets. *Review of Financial Studies*, *24*(11), 3731–3777. [cited on p. 151, 152, 154, 155, 189]

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, *7*(1), 77–91. [cited on p. 7, 145]

McLean, R. D., & Pontiff, J. (2016). Does academic research destroy stock return predictability? *Journal of Finance*, *71*(1), 5–32. [cited on p. 95, 151]

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica*, *41*(5), 867–887. [cited on p. 2]

Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics*, *104*(2), 228–250. [cited on p. 149, 151, 152, 189]

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica*, (pp. 768–783). [cited on p. 1]

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708. [cited on p. 48, 76, 105, 109, 123, 124, 127, 135, 157, 162, 165, 173, 178, 190]

Novy-Marx, R. (2016). Testing strategies based on multiple signals. *University of Rochester Working Paper*. [cited on p. 101]

Pástor, L., & Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, *111*(3), 642–685. [cited on p. 20, 150, 152, 189]

Pástor, L., Stambaugh, R. F., & Taylor, L. A. (2015). Scale and skill in active management. *Journal of Financial Economics*, *116*(1), 23–45. [cited on p. 18]

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies*, *22*(1), 435–480. [cited on p. 27]

Pukthuanthong, K., Roll, R., & Subrahmanyam, A. (2019). A protocol for factor identification. *Review of Financial Studies*, *32*(4), 1573–1607. [cited on p. 5, 95, 100, 103, 104, 110, 111, 133, 146, 158, 159, 160, 192]

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, *13*(3), 341–360. [cited on p. 2]

Shanken, J. (1992). On the estimation of beta-pricing models. *Review of Financial Studies*, *5*(1), 1–33. [cited on p. 129, 131]

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, *19*(3), 425–442. [cited on p. 1]

Sharpe, W. F. (1998). Morningstar's risk-adjusted ratings. *Financial Analysts Journal*, *54*(4), 21–33. [cited on p. 26]

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business*, *74*(1), 101–124. [cited on p. 139]

Stambaugh, R. F., & Yuan, Y. (2017). Mispricing factors. *Review of Financial Studies*, *30*(4), 1270–1315. [cited on p. 52, 84, 99, 150, 152, 154, 189]

Tao, X., Wang, B., Wang, J., & Wu, C. (2022). Economic policy uncertainty and the cross-section of corporate bond returns. *Journal of Fixed Income*, *32*(1), 6–44. [cited on p. 94, 95, 98, 139, 142, 151, 152]

Treynor, J. L. (1961). *Toward a theory of market value of risky assets*. [cited on p. 1]

Verdelhan, A. (2018). The share of systematic variation in bilateral exchange rates. *Journal of Finance*, *73*(1), 375–418. [cited on p. 151, 152, 154, 189]

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, *21*(4), 1455–1508. [cited on p. 95, 189]

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 817–838. [cited on p. 105]

Zaremba, A., & Czapkiewicz, A. (2017). The cross section of international government bond returns. *Economic Modelling*, *66*, 171–183. [cited on p. 155]

Zhao, X. (2005). Determinants of flows into retail bond funds. *Financial Analysts Journal*, *61*(4), 47–59. [cited on p. 17]

Zitzewitz, E. (2003). Who cares about shareholders? Arbitrage-proofing mutual funds. *Journal of Law, Economics, and Organization*, *19*(2), 245–280. [cited on p. 58]