True Value Investing in Credits through Machine Learning[1]

22 March 2024

Patrick Houweling is a managing director in the Quantitative Investments Division at Robeco in Rotterdam, The Netherlands. p.houweling@robeco.nl.

Philip Messow is a senior quantitative researcher in the Quantitative Investments Division at Robeco in Rotterdam, The Netherlands. p.messow@robeco.nl.

Robbert-Jan ‘t Hoen is a quantitative researcher in the Quantitative Investments Division at Robeco in Rotterdam, The Netherlands. r.t.hoen@robeco.nl.

Abstract

Value investing in the credit market aims to identify mispricings by determining whether a bond’s credit spread adequately compensates for its risk. To assess how successful value approaches are in capturing mispricings, we use a decomposition of credit returns into a risk and a repricing component. We show that existing value factors from the literature earn not only from mispricings but also from taking more risk. To better control for risk, we construct a novel machine learning based value factor and find that it outperforms existing value factors while earning less from risk taking and more from mispricings.

Key highlights

1. We successfully replicate a large set of value factors from the literature.
2. Surprisingly, we find that all existing value factors earn substantially from risk.
3. We introduce a machine learning based value factor, whose performance is driven less by risk and more by mispricings, and generates an information ratio above 3.

Keywords: corporate bonds; machine learning; mispricing; value investing

JEL Classification: G11, G12, G14

[1] Robbert-Jan ‘t Hoen is the corresponding author. The authors thank participants at the Robeco research seminars for valuable comments and suggestions. All remaining errors are our own. The views expressed in this paper are not necessarily shared by Robeco or its subsidiaries.

Introduction

Academic studies have shown that value investing is a successful investment strategy, as undervalued, “cheap”, securities tend to outperform overvalued, “expensive”, securities in many asset classes (Assness, Moskowitz and Pedersen, 2013). The exact approach to determining cheap or expensive securities differs between asset classes, but the goal is the same: to identify relative mispricings. In the equity market, the valuation of a stock is often determined by comparing its market price to a fundamental anchor, such as the company’s book value (Fama and French, 1992). Similarly, in the credit market, the typical value approach aims to determine whether the bond’s market price of risk, in other words its credit spread, is a sufficient compensation for a corporate bond’s risk (L’Hoir and Boulhabel, 2010; Houweling and van Zundert, 2017).

There is an active debate on the source of the value premium in the equity market as some claim that it is indeed a reward for mispricing (Lakonishok, Shleifer and Vishny, 1994; Piotroski and So, 2012), while others argue that it is instead a compensation for bearing additional risk (Fama and French, 1993; Zhang, 2005; Choi, 2013). In the credit market, unlike in the equity market and despite the growing literature on value investing in credit, there is no standard definition of the value factor. This makes it difficult to assess "the" value premium. In addition, there has not been much discussion of the source of the value premium. Mispricing is generally thought to drive the performance of the value factor in the credit market, but the evidence is relatively sparse.

We examine the effectiveness of value investing in the credit market and compare a large set of existing credit value factors from the literature to assess their performance and identify the driving forces. To the best of our knowledge, such a study has not been done before and can therefore help to deepen our understanding of value investing. For the empirical analyses, we use monthly bond data from January 1994 to December 2022. Our dataset contains 1.2 million bond-month observations and is larger in the time-series and/or cross-section dimensions than the datasets used in prior studies.

Our paper makes three contributions to the literature. First, we review the literature on value investing in credits, categorize the wide range of approaches and replicate them; we show that, while there is large heterogeneity in the cross-section of value factors, the majority performs well in our extended sample period. Second, we decompose the credit excess return over duration-matched Treasuries into a carry and repricing component. Large carry returns are driven by higher credit spreads and thus capture risk, while high repricing returns are driven by larger credit spread changes and thus result from initial mispricings. This decomposition allows us to show that all existing value factors beat the market not only by earning from mispricings but also from taking more risk. Third, recognizing the need for better risk controls, we propose a new value factor based on machine learning (ML), specifically a Gradient Boosted Regression tree. This new value factor earns 85% of its performance from repricing, better controls for risk, and yields an information ratio of 3.13, higher than any of the existing approaches. These results also hold when several robustness checks are applied.

Our results highlight the importance of performance attribution in assessing the alignment between ex-ante investment intent and ex-post investment outcome. More specifically, our results suggest that investors who engage in value strategies should carefully consider whether or not they are "true" value investors.

In the following sections, we provide an overview of value investing in credits, introduce a decomposition of the credit return into carry and repricing, discuss the data used, analyze the performance of existing value factors based on this decomposition and introduce an ML-driven value factor. Based on these insights, we then discuss the performance of the new value factor when compared to the existing ones and finish with several robustness checks.

Value investing in credits

Value investing has been studied in a variety of asset classes. The exact implementation of the value factor varies depending on the specifics of the asset. In the equity market, the book-to-price ratio (B/P) of Fama and French (1992) is widely accepted as the standard value factor definition. In the credit market, there is no such widely accepted value factor definition (yet). In the absence of a standard approach, several definitions have been proposed over the past two decades.

The earliest work on the value factor in the credit market focuses on the valuation properties of the (option-adjusted) credit spread, which is the spread of a corporate bond over the yield of duration-matched Treasuries adjusted for any embedded options. The "spread valuation" factor of Hottinga, van Leeuwen, and van IJserloo (2001, henceforth "HLIJ") is simply the credit spread and finds bonds with the widest spreads to be the most attractive due to their higher initial yields and greater upside potential.[2]

Thereafter, studies adopt the equity market approach of comparing the market price of risk to a fundamental anchor or "fair value" to understand whether bonds are cheap for a reason (value traps) or not (mispricings). Here, we can discern two streams. Firstly, there are studies that compare the credit spread with an individual bond or company characteristic and then use the resulting ratio as a value factor. Correia, Richardson, and Tuna (2012, CRT) use the Distance-to-Default (DtD) of the issuer to account for the default risk of a firm. Secondly,

[2] Bonds trading close to par (low spread) have only limited room left for a price increase (spread decrease) while there is much more downside potential.

there are studies that estimate a fair spread in a (regularized) regression on the credit spread with bond characteristics (Houweling and van Zundert, 2017, HZ), firm characteristics (L'Hoir and Boulhabel, 2010, LHB) or a combination of bond and firm controls (Kaufmann, Messow, and Wisser, 2023, KMW). In all cases, the residual of the regression serves as the value measure, being the difference between the market spread and the estimated fair spread.

Several studies recognize the need for better risk control and adopt a multi-stage approach. Israel, Palhares, and Richardson (2018, IPR) combine two value signals and thereafter neutralize the Duration-Times-Spread (DTS) exposure of their signal. Heckel, Amghar, Haik, Laplènie and de Carvalho (2020, HAHLC) combine multiple value signals and then neutralize the sector, spread, duration and size dimensions using their local scoring methodology. Shen, Pathammavong, and Chen (2019, SPC) estimate their value signal within sectors. Polbennikov, Desclèe, and Dubois (2021, PDD) sequentially run bond and firm level regressions to estimate their value signal.

Finally, there are also studies that try to more directly transfer the B/P approach from the equity market to the credit market. Bektic, Wenzler, Wegener, Schiereck and Spielmann (2019, BWWSS) simply use the B/P ratio of the stock as their value factor for the bond. Bartram, Grinblatt, and Nozawa (2020, BGN) introduce a bond market equivalent that compares the book value of a bond (par value adjusted for premium/discount at issuance) to its market value (bond price).

This is an extensive set of eleven value factors, but what is the right or "true" way to approach value investing in the credit market? To answer this question, we replicate and compare the value factors from the literature based on the information disclosed in the papers (see Appendix A for more details on the replication). In the remainder of the paper, we refer to the value factors from the different papers by the abbreviations introduced above. Before showing the replication results, we first present our method for decomposing the return of each value factor into two driving forces.

Credit return decomposition

The goal of value investing is to identify mispriced securities, whose market price will revert to their fundamental value over time as relevant information is priced in by market participants. There is a lively debate in the equity market about the source of the value premium. Some argue that it is actually a reward for mispricing, such as overreaction to new information (Lakonishok, et al., 1994). Others argue that it is instead a compensation for taking on higher risk, such as low earnings on assets (Fama and French, 1993) or high financial leverage (Choi, 2005). The unobservability of the risk premium in the equity market complicates this discussion.

In the credit market, the risk premium is in fact observable as the credit spread and therefore, we are able to assess whether a bond's risk premium changed after the initial investment and to consequently determine if a bond was likely to be mispriced or not. For example, for a bond with an initial credit spread of 200bps that drops to 150bps (resp. increases to 250bps) in the next period, a spread tightening (resp. widening) is observed, which indicates that the bond likely was mispriced assuming no significant market and/or curve movement in that given period. There would be no indication of a mispricing if the spread remained unchanged.

Based on this observation, we work with the following well-known decomposition of the credit return into a spread level component and a spread change component. We focus on credit excess returns over duration-matched Treasuries, rather than total returns, based on the assumption that investors in the corporate bond market primarily seek to capture risk premia driven primarily by the default premium, independent of the term risk premium (van

Binsbergen, Nozawa and Schwert, 2023). The credit return $R_{t\to t+\tau}$ of a corporate bond over the period from t to $t+\tau$ can be approximated by performing an k^{th} order Taylor expansion:

$$R_{t\to t+\tau} = S_t \times \tau - \Delta S_{t+\tau}\frac{dP_t}{dS_t} + \frac{1}{2!}(\Delta S_{t+\tau})^2\frac{d^2P_t}{dS_t^2} - \ \dots + \frac{1}{k!}(\Delta S_{t+\tau})^k\frac{d^kP_t}{dS_t^k}$$

where S_t is the spread at time t, τ is the holding period, P_t is the bond price at time t and $\Delta S_{t+\tau} = S_{t+\tau} - S_t$ is the spread change over the holding period. In this equation, we recognize the bond duration (first order), convexity (second order), and higher order derivatives of bond price to credit spread. The equation can be rewritten as:

$$R_{t\to t+\tau} = \underbrace{S_t \times \tau}_{carry} + \underbrace{\sum_{i=1}^{k}\frac{(-1)^i}{i!}(\Delta S_t)^i\frac{d^iP_t}{dS_t^i}}_{repricing}.$$

This formulation shows that the sources of credit return are (i) spread pick-up or *carry*, which is a compensation for bearing risk, and (ii) spread change or *repricing*, which results from changes in the credit spread. We demonstrate below in Table 2 that carry investing indeed comes with a lot of extra risk versus the market. The carry and repricing components are only weakly correlated[3], which means that it should be possible to earn more from one without increasing the other.

We believe that a successful value factor in the credit market should earn (most of) its outperformance from repricing, as positive repricing returns are an indication of holding bonds with reverting mispricings, while higher carry returns are merely a compensation for taking on more risk.

[3] Over the 1994-2022 sample, we find an average cross-sectional Pearson (Spearman's rank) correlation of -0.3% (6.9%) between carry and repricing return.

Data

Our data set contains monthly observations from January 1994 to December 2022. The bonds that we consider are constituents of the Bloomberg US Aggregate Corporate index (containing Investment Grade (IG) bonds) and the Bloomberg US Corporate High Yield (HY) index. As a result of the index construction rules, smaller and less liquid bonds are not included in these indices. Within this sample, we focus on senior bonds, rated between AAA and CCC-. Also, we only consider bonds that are issued by a listed company, as certain value factors require equity or accounting information for their construction. We exclude bonds at the start of the month if they were not rated, or if their credit spread or market value is missing. In addition, bonds are excluded when there is missing coverage for any of the inputs for the value factor replication.

Bond data is sourced from Bloomberg and includes bond characteristics and analytics such as market value, time to maturity, credit rating, credit spread and returns. If a bond defaults, Bloomberg bases the last price on the expected recovery rate. Thus, we have no survivorship bias in this study. Firm data is obtained from Compustat and includes fundamentals such as assets, total debt and earnings and equity market data like the market cap and equity return is collected from Datastream Prices and analyst forecasts are finally obtained from I/B/E/S Estimates. As a robustness test, we also obtain returns based on actual trades from the Trade Reporting and Compliance Engine (TRACE) over the period 2005-2020, using the filters proposed in Dick-Nielsen (2009) and Dick-Nielsen (2014).

The final dataset contains a total of 1,235,195 bond-month observations. Table 1 summarizes the dataset. On average, the sample consists of 3,549 bonds per month, ranging from 1,872 at the start to 5,441 at the end of the sample. The average bond is BBB+ rated and has a credit spread of 185 bps, a time to maturity of 10.6 years, an outstanding amount of $554 mln and a credit return of 0.11% per month. Further, the total market capitalization per month is on average close to $2.1 trillion, of which 85% is IG-rated and 15% is HY-rated.

Table 1: Summary Statistics

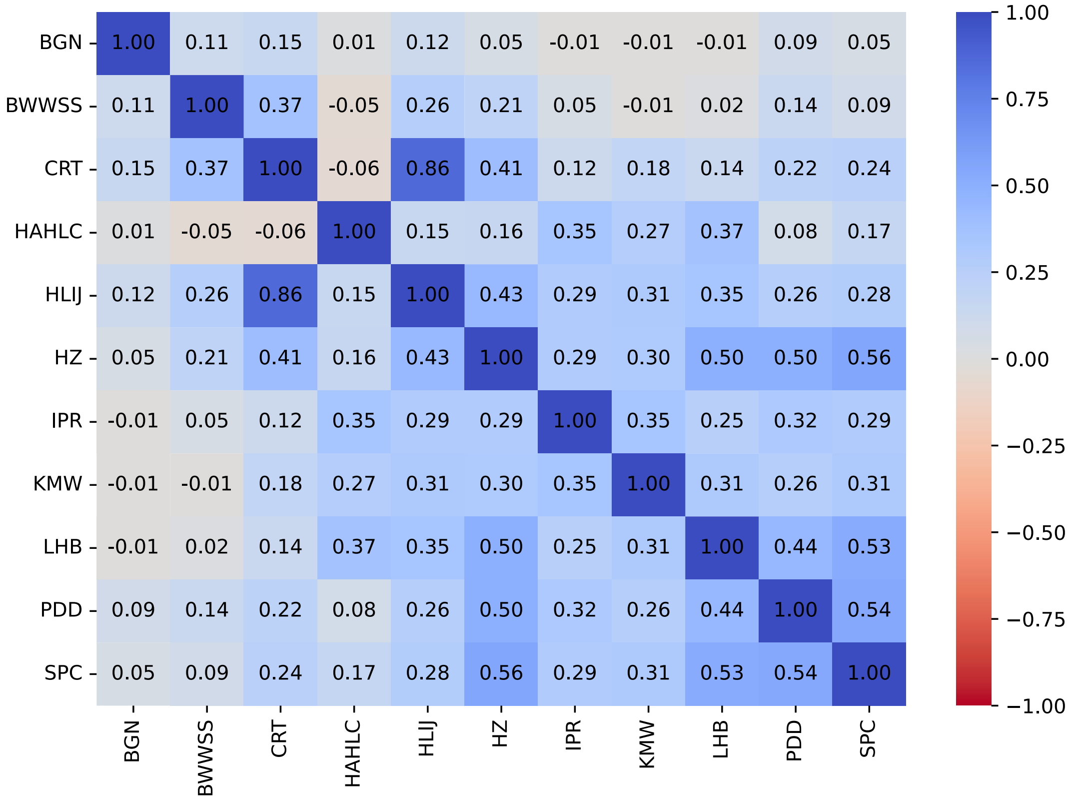
	Mean	Percentile				
		5%	25%	50%	75%	95%
Number of observations	3,549					
Rating	BBB+	AA-	A	BBB+	BBB-	BB-
Spread (bps)	185	55	94	136	207	451
Time to Maturity (in years)	10.6	1.6	3.8	6.8	14.4	28.5
Amount outstanding ($ mln)	554	210	279	403	632	1506
Credit return (per month)	0.11%	-2.06%	-0.45%	0.10%	0.68%	2.36%

Note: Sample period 1994-2022. The individual statistics are computed cross-sectionally and are then aggregated over time. The spread is the option-adjusted credit spread. The rating is the middle rating between ratings provided by Fitch, Moody's and S&P. The credit return is the excess return over duration-matched Treasuries.

The cross-section of value factors

In a first analysis, we determine how similar (or different) the value factors are. Figure 1 shows the rank correlation between each pair of value factors. Somewhat surprisingly, we find a relatively low mean (median) pairwise correlation of 36% (29%) and only one correlation above 60% (between CRT and HLIJ). The lack of high correlations and the wide variety of approaches in the set of value factors suggest that the value factors are not that similar after all, which hints at the existence of a value "factor zoo" in the credit market.

Figure 1: Rank correlation matrix of all replicated value factors



Note: Sample period 1994 – 2022. Spearman's rank correlations are computed cross-sectionally and then aggregated over time.

To assess the viability of value investing in the credit market and to compare the different value factors, we perform sorted-portfolio backtests by constructing quintile portfolios based on each of the value factors and holding the bonds until the portfolio is rebalanced. For the main results, we analyze equally weighted portfolios (while the market is value-weighted) and a one-month holding period, but the results are also robust to alternative specifications such as value-weighted portfolios or a twelve-month holding period (see below).

Table 2 presents summary statistics for the backtested value portfolios. Panel A shows that the top portfolios of all value factors outperform the market while at the same time being more volatile. Nevertheless, all value factors, except for BWWSS, achieve a higher Sharpe ratio than the market. We find that HLIJ shows both the highest return and the highest volatility for the top portfolio, at 4.17% and 10.39%, respectively. Since HLIJ is a pure carry strategy, i.e. it buys the bonds with the highest spreads, this demonstrates that carry comes with a lot of risk. BWWSS, on the other hand, has the lowest top portfolio return of all value factors, indicating that simply using the de-facto definition of the equity market value factor is not very successful in the credit market and that having a credit-specific definition is warranted. Panel B shows that all the value factors, except for BWWSS, have positive top-minus-bottom information ratios (IRs), with IPR being the highest with an IR of 1.92 and a median IR of 0.67. Panel C shows CAPM regression results for the top-minus-bottom portfolios. All value factors tend to be risk-seeking as CAPM betas are positive and range from 0.08 to 1.06. Despite their risk tilts, 9 (7) out of 11 value factors show positive and significant CAPM alpha's at the 5% (1%) significance level, with the highest being IPR with an alpha of 2.71% (t-stat of 6.92) and a median CAPM alpha of 1.71%. Note that these results include (considerable) out-of-sample periods for most value factors, which makes the positive results strong evidence in favor of value investing in the credit market.

Table 2: Performance statistics of all replicated value factors

	Market	**BGN**	**BWWSS**	**CRT**	**HAHLC**	**HLIJ**	**HZ**	**IPR**	**KMW**	**LHB**	**PDD**	**SPC**
Panel A: Top												
Return	1.04%	3.43%	1.13%	3.22%	2.83%	4.17%	2.56%	3.96%	2.89%	3.77%	3.92%	4.02%
Volatility	4.73%	9.61%	8.38%	10.38%	5.30%	10.39%	7.39%	6.08%	5.72%	7.93%	9.16%	8.54%
Sharpe ratio	0.22	0.36	0.13	0.31	0.53	0.40	0.35	0.65	0.51	0.48	0.43	0.47
Panel B: Top-minus-bottom												
Outperformance		3.05%	-0.19%	2.95%	3.04%	4.13%	2.47%	4.42%	2.96%	3.56%	3.58%	4.05%
Tracking error		6.68%	5.23%	8.45%	1.96%	8.83%	4.77%	2.30%	2.15%	3.93%	5.38%	4.11%
Information ratio		0.46	-0.04	0.35	1.56	0.47	0.52	1.92	1.37	0.91	0.67	0.98
Panel C: Top-minus-bottom CAPM statistics												
CAPM alpha		1.48%	-0.60%	1.08%	1.71%	2.03%	1.02%	2.71%	1.67%	2.10%	2.02%	2.22%
t-stat		2.13*	-1.10	1.61	6.34**	2.97**	2.30*	6.92**	6.65**	5.02**	3.50**	4.74**
CAPM beta		0.87	0.66	1.06	0.08	1.06	0.48	0.21	0.18	0.61	0.82	0.73

Note: Sample period 1994-2022. The backtest uses quintile portfolios and a one-month holding period. Return, volatility, outperformance, tracking error and CAPM alpha statistics in Panels A, B and C are given in annualized percentages. CAPM statistics in Panel C are estimated by regressing the top-minus-bottom portfolio return for each value factor on a constant and the market return. The market return series is the market-value weighted return of all bonds in the sample at each point in time. Robust standard errors are used for the t-statistic calculation. * denotes statistical significance at the 5% level and ** at the 1% level.

The top-minus-bottom performance of the value factor is driven by the buying and shorting of bonds. However, shorting bonds can be difficult in practice. Therefore, to assess the relative importance of the long and short sides, we look at the contribution of both sides for all value factors in Table 3. In general, we find that about 2/3th of the performance comes from the long side and 1/3th from the short side. This means that the value factors are generally able to detect bonds with increasing spreads that do not reverse, and benefit from shorting them, but that they do not completely drive the top-minus-bottom results.

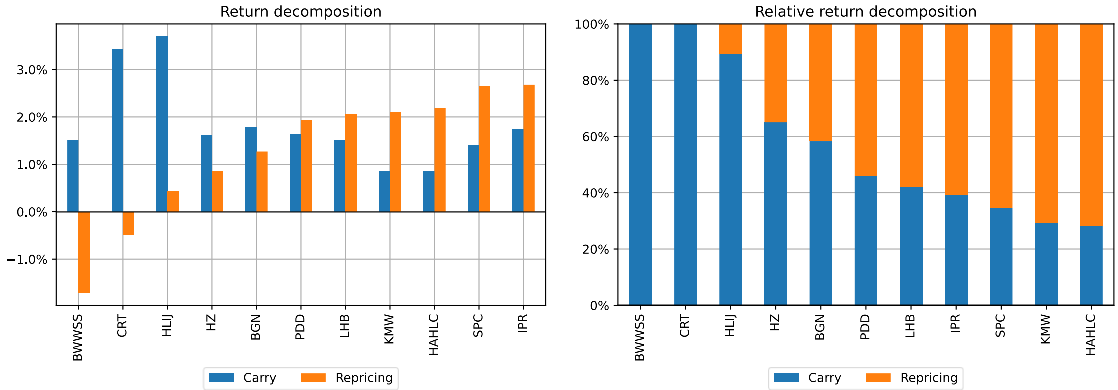
Table 3: Long versus short side contribution

	BGN	**BWWSS**	**CRT**	**HAHLC**	**HLIJ**	**HZ**	**IPR**	**KMW**	**LHB**	**PDD**	**SPC**
Panel A: Top-minus-market											
Outperformance	2.39%	0.09%	2.18%	1.79%	3.12%	1.52%	2.92%	1.85%	2.73%	2.88%	2.98%
Tracking error	5.56%	4.27%	6.16%	1.49%	6.19%	3.28%	2.32%	1.58%	3.65%	4.97%	4.26%
Information ratio	0.43	0.02	0.35	1.20	0.50	0.46	1.26	1.17	0.75	0.58	0.70
Panel B: Bottom-minus-market											
Outperformance	-0.66%	0.27%	-0.77%	-1.25%	-1.01%	-0.96%	-1.49%	-1.10%	-0.82%	-0.70%	-1.07%
Tracking error	1.93%	1.40%	2.56%	1.47%	2.96%	2.07%	2.18%	1.65%	1.94%	1.70%	1.52%
Information ratio	-0.34	0.20	-0.30	-0.85	-0.34	-0.46	-0.69	-0.67	-0.42	-0.41	-0.71

Note: Sample period 1994-2022. The backtest uses quintile portfolios and a one-month holding period. Outperformance and tracking error are given in annualized percentages. The market return series is the market-value weighted return of all bonds in the sample over time.

So far, we have evaluated the value factors based on their overall performance. In order to understand how much of the performance of the value factors is actually due to the repricing of mispriced bonds and how much due to earning a higher carry, we decompose the top-minus-bottom returns according to the credit return decomposition introduced above. Figure 2 shows the performance decomposition of the top-minus-bottom portfolios from an absolute perspective (left) and a relative (right) perspective. We find that all value factors except for BWWSS and CRT earn from repricing, with a median (maximum) contribution of 54% (72%). At the same time, we find that all value factors earn from carry, with a median (minimum) contribution of 46% (28%). Taken together, this shows that the value factors do not earn all of their performance from mispricings, but that risk is also a significant driver of performance. In addition, there is considerable dispersion across the value factors, with BWWSS and CRT earning nothing from repricing, KMW and HAHLC earning most of their performance from repricing, and the other value factors falling in between.

Figure 2: Carry and repricing attribution



Note: Sample period 1994-2022. The carry and repricing return are calculated in accordance with the credit return decomposition. Each plot is sorted in ascending order on its "repricing" value, i.e. on the repricing return level for the figure on the left and on the repricing return fraction for the figure on the right. The y-limits in the right figure are capped between 0% and 100% for visibility reasons.

Next, we examine whether the larger contribution of repricing is actually beneficial for the (risk-adjusted) performance of the value factors. Table 4 shows the regressions of the fraction of performance attributable to repricing on the outperformance, tracking error and information ratio of the value factor top-minus-bottom portfolios. We find that value factors with a higher mispricing fraction do not have significantly higher returns, but they do have significantly tracking error and higher information ratios as a 10% increase in mispricing fraction reduces the tracking error by 0.70% (t-stat of -3.57) per year and increases the information ratio by 0.17 (t-stat of 6.07). These results suggest that earning more from repricing is indeed beneficial from an investment perspective as the risk-adjusted performance improves.

Table 4: Repricing benefit regression

		Outperformance	Tracking error	Information ratio
Constant	coef	2.13%	7.86%	0.10
	t-stat	2.31*	8.27**	0.98
Repricing %	coef	2.25%	-6.99%	1.73
	t-stat	1.33	-3.57**	6.07**
# Obs		11	11	11
R^2		0.25	0.66	0.66

Note: Sample period 1994-2022. Outperformance, tracking error and information ratio top-minus-bottom portfolio statistics are separately regressed on a constant and the respective repricing fractions. Repricing fractions are capped between 0 and 100%. Robust standard errors and a small sample correction are used for the t-statistic calculation. * denotes statistical significance at the 5% level and ** at the 1% level.

All the value approaches tested above are linear in nature. These linear approaches implicitly assume that the spread is linearly related to the risk controls and that there are no significant non-linear effects or interaction effects. Several value factors attempt to control for such risks by introducing a separate risk overlay (e.g. IPR and HAHLC).[4] We believe that modeling the non-linear and interaction effects directly instead of through a multi-step approach could be a more effective way to constructing a “true” value factor that earns most of its performance from repricing returns. To this end, we explore a machine learning (ML) approach next.

Machine Learning based Value factor

So far, we have found that the value factors proposed in the literature achieve an attractive (risk-adjusted) return, but are also associated with certain limitations. As a result, the existing value factors earn both through repricing and carry. By using more flexible models in the form of Gradient Boosted Regression (GBR) trees, we address the structural issues. GBR can take into account non-linearities, as well as interaction effects, to better model dependencies, thereby reducing the risk bias and making the return earned more dependent on the actual

[4] The risk overlays are successful, as the same value factors with their risk overlay removed have significantly higher risk exposure and lower risk-adjusted performance. The results are available upon request and are not included here because discussing the different risk overlays in detail is outside the scope of this paper.

spread change and less dependent on structurally taking more risk exposure through a higher carry. It should be mentioned that it is not the aim of this study to run a horse race to find the best ML model. Our choice of a tree-based model is consistent with, for example, Hanauer, Kononova, and Rapp (2021), who show that a tree-based ML value factor outperforms traditional value factors in the equity market.

Trees sequentially partition explanatory variables into non-overlapping domains. Each observation lies in a single domain, and each domain is associated with an expected value of the dependent variable. The algorithm looks for splits that define these areas in a way that maximizes the explained variation of the dependent variable in the context of "training" data. Boosted trees follow the idea of using a large number of very simple sub models – so-called weak learners. The very simple trees are trained sequentially to explain the variation that has not been explained by previously estimated trees. The impact of a single tree (and, therefore, the overall model complexity) is subsequently mitigated by multiplying the explained variation by a "decay rate". The decay rate is one of the "hyperparameters" that have to be chosen and that define the complexity of the model (i.e. the number of nodes per tree and number of interactions between features). The other two common hyperparameters are the number of trees and the depth of the single trees. We run our model with hyperparameters [number of trees, learning rate, max depth] = [400, 0.1, 5]. We show later that our results do not depend on the exact specification of the hyperparameters.

As input for the boosted regression trees, we use a broad set of features, which have already been used in the studies reported above. The features used are listed in the Appendix. These features are a mix of market data, balance sheet data, ratings and sectors. Table 6 shows the associated quintile performance for the ML based value factor (henceforth "HMH") from 1994 to 2022. We can see that the top quintile portfolio (Q1) is also associated with the highest returns and the highest risk-adjusted performance. The top quintile achieves an outperformance of 2.47% and an IR of 2.11 while the top-minus-bottom performance is 4.35% with an IR of 3.13. In addition, outperformances and information ratios are monotonically increasing across the quintiles, indicating attractive properties of our proposed value factor.

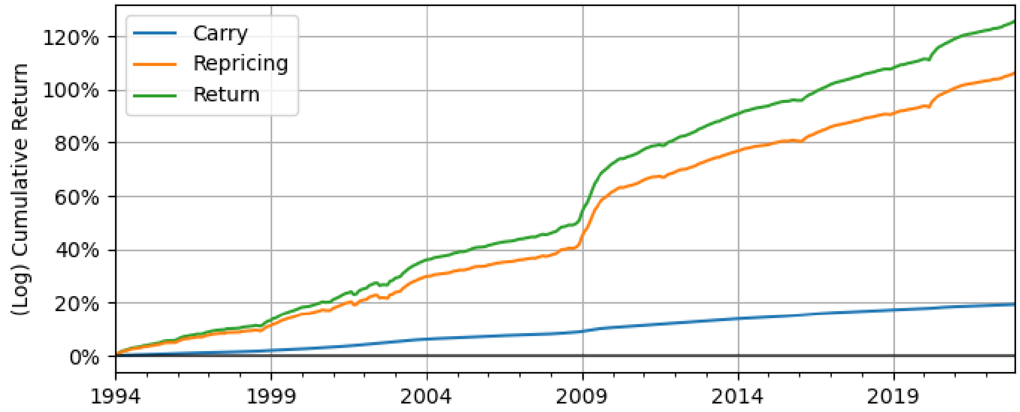
Table 6: Performance statistics of HMH

	Q1	Q2	Q3	Q4	Q5	Q1 – Q5
Panel A: Top-minus-market						
Outperformance	2.47%	1.09%	0.36%	-0.53%	-1.88%	4.35%
Tracking error	1.17%	1.41%	1.95%	0.82%	1.19%	1.39%
Information ratio	2.11	0.78	0.18	-0.64	-1.58	3.13
Panel B: CAPM statistics						
CAPM Alpha	2.44%	0.93%	0.08%	-0.57%	-1.73%	4.16%
t-stat	11.24**	4.19**	0.30	-3.78**	-9.63**	20.20**
CAPM Beta	0.03	0.16	0.26	0.04	-0.15	0.18

Note: Sample period 1994-2022. The backtest uses quintile portfolios and a one-month holding period. Outperformance, tracking error and CAPM alpha statistics in Panels A and B are given in annualized percentages. CAPM statistics in Panel B are estimated by regressing the portfolio returns on a constant and the market return. The market return series is the market-value weighted return of all bonds in the sample at each point in time. Robust standard errors are used for the t-statistic calculation. * denotes statistical significance at the 5% level and ** at the 1% level.

Figure 3 plots the top-minus-bottom performance decomposition over time for HMH. We can see that the repricing and carry contributions are stable over time. This is remarkable for the repricing component because it means that HMH is able to identify true mispricings regardless of the underlying market environment.

Figure 3: Top-minus-bottom outperformance decomposition over time



Note: Sample period 1994-2022. The carry and repricing return are calculated in accordance with the credit return decomposition. Log cumulative returns are used.

Comparing HMH to existing Value factors

In the previous section we have seen that the HMH signal is a powerful value factor on its own. In this section, we shed light on the question whether it also improves upon the existing value factors that were presented above.

Table 7 summarizes the backtest results for all value factors including HMH. We can see, that HMH outperforms the other value approaches on a risk-adjusted basis with an IR of 3.13. The higher IR is not only accomplished by a high outperformance, but importantly by the lowest tracking error, which implies that it is most successful in controlling for risks. In addition, the CAPM alpha is the most significant for HMH resulting in a t-statistic of 20.20.

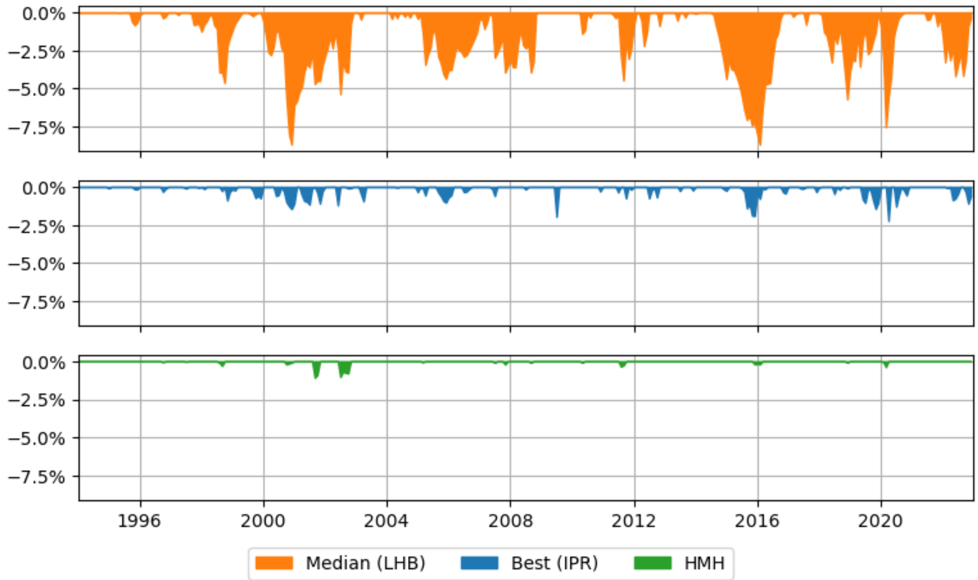
Table 7: Value top-minus-bottom performance

	BGN	BWWSS	CRT	HAHLC	HLIJ	HZ	IPR	KMW	LHB	PDD	SPC	HMH
Panel A: Top-minus-bottom												
Outperformance	3.05%	-0.19%	2.95%	3.04%	4.13%	2.47%	4.42%	2.96%	3.56%	3.58%	4.05%	4.35%
Tracking error	6.68%	5.23%	8.45%	1.96%	8.83%	4.77%	2.30%	2.15%	3.93%	5.38%	4.11%	1.39%
Information ratio	0.46	-0.04	0.35	1.56	0.47	0.52	1.92	1.37	0.91	0.67	0.98	3.13
Panel B: CAPM statistics												
CAPM Alpha	2.04%	-1.06%	1.35%	2.87%	2.44%	1.60%	4.60%	2.71%	2.95%	2.67%	3.29%	4.16%
t-stat	2.25*	-1.68	1.68	8.62**	3.05**	3.25**	11.49**	7.92**	5.67**	4.17**	7.81**	20.20**
CAPM Beta	0.96	0.84	1.54	0.17	1.63	0.84	-0.17	0.24	0.58	0.87	0.73	0.18

Note: Sample period 1994-2022. The backtest uses quintile portfolios and a one-month holding period. Outperformance, tracking error and CAPM alpha statistics in Panels A and B are given in annualized percentages. CAPM statistics in Panel B are estimated by regressing the top-minus-bottom portfolio return for each value factor on a constant and the market return. The market return series is the market-value weighted return of all bonds in the sample at each point in time. Robust standard errors are used for the t-statistic calculation. * denotes statistical significance at the 5% level and ** at the 1% level.

We also show the associated drawdowns of HMH compared to the second best signal (IPR) and the median signal (LHB) in Figure 4. We see that LHB suffers from serious drawdowns at several points in time, IPR already substantially improves over this and HMH hardly suffers from any drawdowns. This shows that HMH provides strong downside risk protection compared to the other value factors, i.e. HMH is best at avoiding the “value traps".

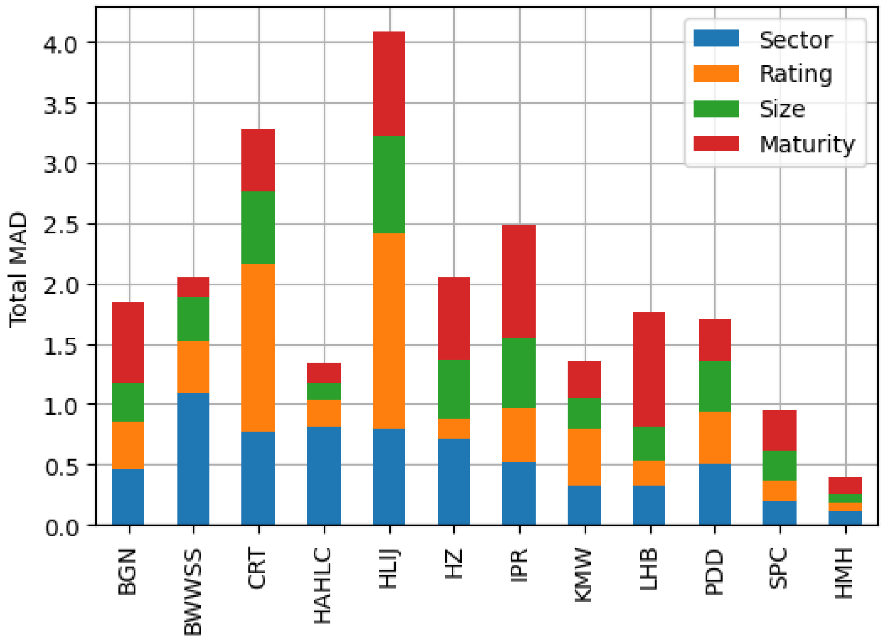
Figure 4: Drawdowns over time for selected value factors



Note: Sample period 1994-2022. Drawdowns are calculated for top-minus-bottom-bottom portfolios of each value factor. Results are only shown for three selected value factors for visibility reasons. The “median” (“best”) here refers to the value factor with the median (best) performance among all value factors excluding HMH.

As another way to compare the value factors, we focus on the exposure to traditional corporate bond risk exposures. For all value factors, we measure the top and bottom portfolio exposure to credit ratings, sectors (Bloomberg Class3), issuer size quintile buckets (ranging from small to large) and time to maturity quintile buckets (ranging from short to long) for each month. Then, to determine the overall exposure to a particular risk dimension, we calculate the “total MAD” as the sum of the median absolute deviation (MAD) between the top and bottom portfolio exposure for each sub-category (i.e. AA rated bonds or large issuers) of a risk dimension (i.e. rating or issuer size). A lower (higher) total MAD value indicates that the top and bottom portfolios are better (worse) able to stay in line with each other for a particular dimension. Figure 5 shows the combined results for all the risk dimensions. We find that there is clear dispersion in the exposure to risk dimensions amongst the different value factors. HLIJ has the highest exposure to the risk dimensions, which showcases that a carry strategy inherently takes large bets on all risk dimensions, while HMH has the lowest exposure to all risk dimensions and any of the risk dimensions individually, which shows that by better controlling for risk in the value estimation it is possible to significantly reduce the exposure to relevant risk dimensions. This reduction in risk exposures makes HMH less exposed to systemic shocks and therefore less risky than the other value factors.

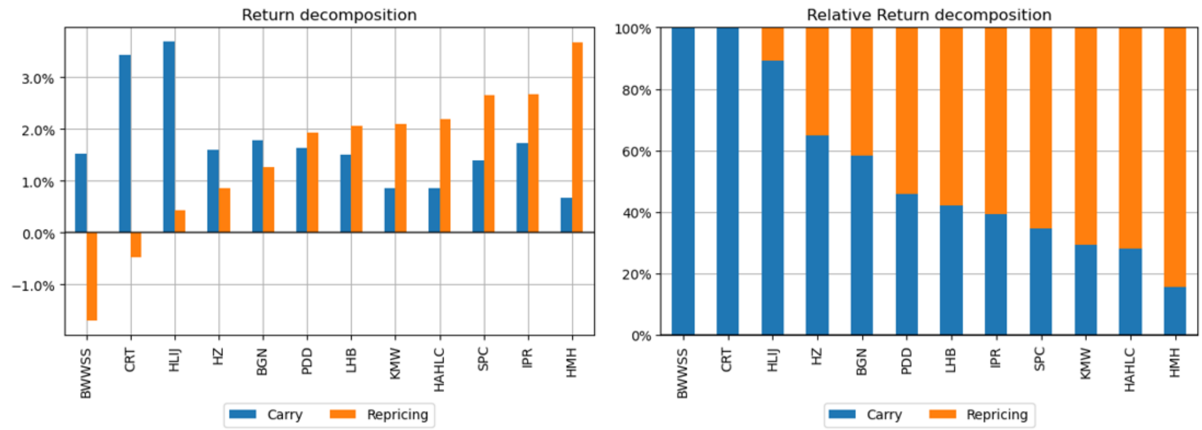
Figure 5: Traditional risk exposures



Note: Sample period 1994-2022. The total MAD is calculated for credit rating, sector (Bloomberg Class3), size (quintile buckets on issuer size) and maturity (quintile buckets on time to maturity). All calculations are done for the top-minus-bottom portfolios of each value factor.

Next, we assess what the lower exposure to risk dimensions means for the HMH performance decomposition compared to the other value factors. Figure 6 shows the decomposition of the carry and repricing returns in absolute (left) and relative (right) terms. From the left chart, we can see that HMH earns the least from carry and the most from repricing, which is the desirable outcome for a value factor. This assessment also holds when looking at the relative decomposition on the right. HMH earns 85% of its outperformance from repricing and only 15% from carry. Thus, we can see that HMH behaves more like a “true” value factor compared to the existing value factors.

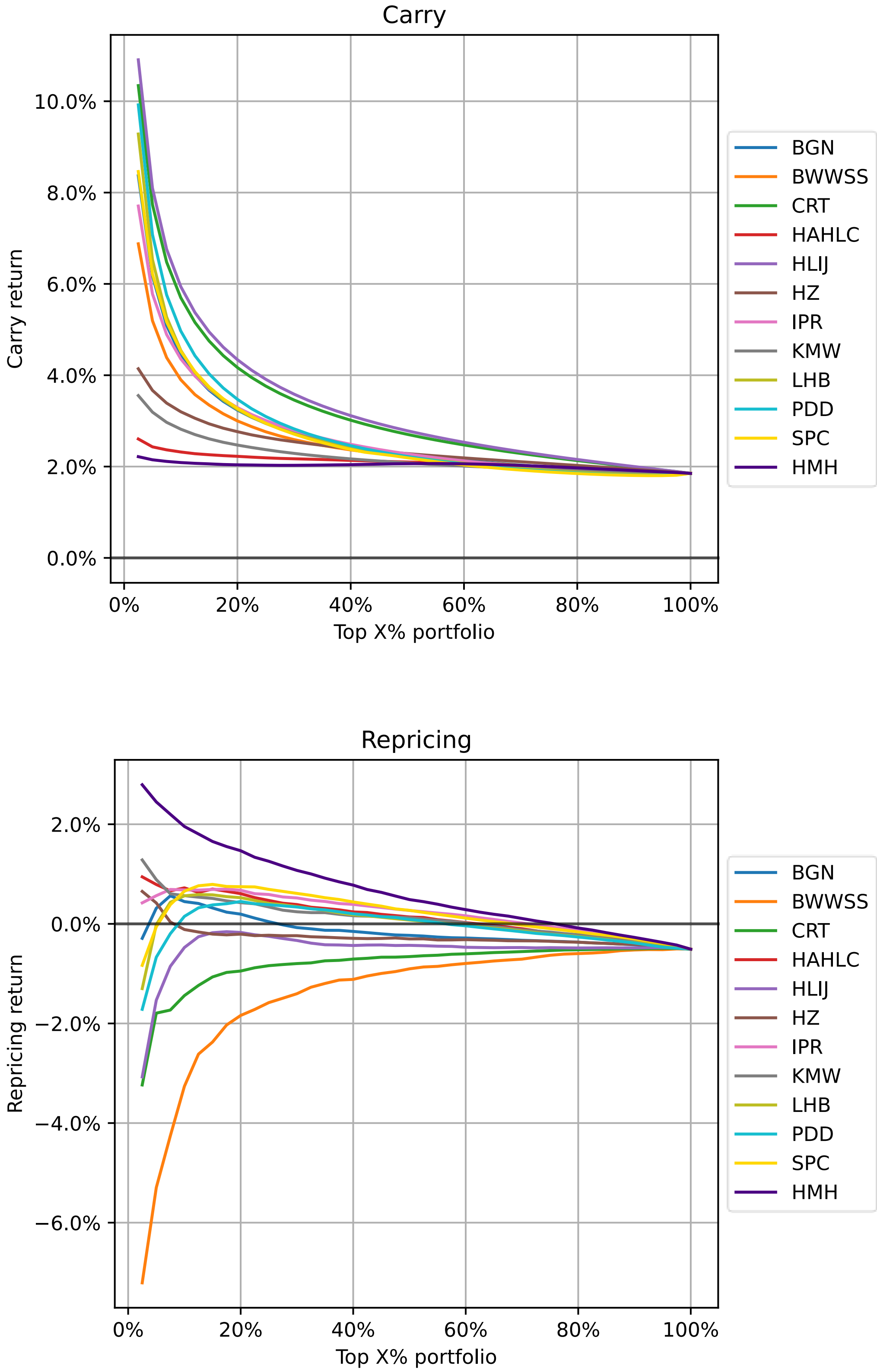
Figure 6: Carry and repricing attribution



Note: Sample period 1994-2022. The carry and repricing return are calculated in accordance with the credit return decomposition. All calculations are done for the top-minus-bottom portfolios of each value factor. Each plot is sorted in ascending order on its “repricing” value, i.e. on the repricing return level for the figure on the left and on the repricing return fraction for the figure on the right. The y-limits in the right figure are capped between 0% and 100% for visibility reasons.

It is also expected that a more flexible model that is closer to a “true” value factor should be able to correctly identify more extreme mispricings. Figure 7 shows the repricing and carry returns of portfolios based on the top x% of scores, ranging in small steps from 2.5% (highly concentrated) to 100% (the entire market) for each value factor. Indeed, for most value factors, we observe that the repricing return initially rises, but then suddenly falls as the portfolio becomes more concentrated. At the same time, the carry return shoots up. This means that the “deep value” opportunities identified by many value factors are actually fairly priced or even expensive. In contrast, for HMH, we find that the repricing return continues to rise as the portfolios become more concentrated, while the carry return barely changes, confirming that it is able to successfully navigate even the more complex cases.

Figure 7: Carry and repricing attribution for concentrated portfolios



Note: Sample period 1994-2022. A grid of backtests with increasingly concentrated portfolios is run, ranging in small steps from 100% (the entire market) to 2.5% (highly concentrated) for each value factor, and the carry and repricing return is tracked for each value factor. The carry and repricing return are calculated in accordance with section "Credit return decomposition" from above. The x-axis shows the portfolio formation rule, e.g. 20% means that the best 20% scores of the respective value factor were used to create the portfolio.

As a final check, we run spanning regressions of HMH on all other value factors and the market return for the top-minus-bottom portfolios. This regression yields a statistically and economically significant alpha of 2.27% per year (t-stat of 14.6), indicating that HMH adds value on top of the existing value factors.

Robustness

In this section, we perform a battery of checks to evaluate the robustness of our results. The main results are labeled as "Base Case". Specifically, we look at the following dimensions:

- Deciles: Use deciles instead of quintiles portfolios.
- 12m: Twelve-month holding period instead of one-month holding period (originated by Jegadeesh and Titman, 1993).
- Top-Market: Top-minus-market portfolios instead of top-minus-bottom portfolios.
- Value-weighted: Value-weighted portfolios instead of equal-weighted portfolios.
- IG only: Evaluate on the IG part of the universe only.
- HY only: Evaluate on the HY part of the universe only.
- TRACE: Use TRACE transaction prices instead of Bloomberg index prices to calculate returns following Bessembinder, Kahle, Maxwell, and Xu (2008) and Dick-Nielsen (2009).

Table 8 shows the corresponding results in terms of risk-adjusted performance and the percentage of return that can be explained by repricing. The results for HMH are strong regardless of the analysis. Looking at the combined average of all seven scenarios, HMH delivers the highest risk-adjusted return and the largest share of repricing. Notably, HMH performs even better with more concentrated portfolios (deciles). This is in contrast to the signal with the best alternative performance (IPR), where the quintile performance was better than the decile performance. Furthermore, the performance and the repricing percentage do not depend on whether the portfolios are value- or equal-weighted. In fact, the risk-adjusted performance of HMH with value-weighting of 3.17 is even better than the base case of 3.13. It is also interesting to see that the results for both IG and HY hold up. The performance of most value factors is substantially worse in the HY sample compared to the IG sample, while for HMH the performance in the HY sample is only slightly below the IG sample. It is also worth noting that the results for the top-minus-market and 12-month holding period scenarios are also satisfactory. These scenarios align the strategy with portfolios that are more realistic, and again, the risk-adjusted performance and repricing percentage are highest for HMH. Finally, we can see that the results for HMH are also superior with TRACE returns compared to the other value factors, which shows that the results are not dependent on the exact return series that are used.

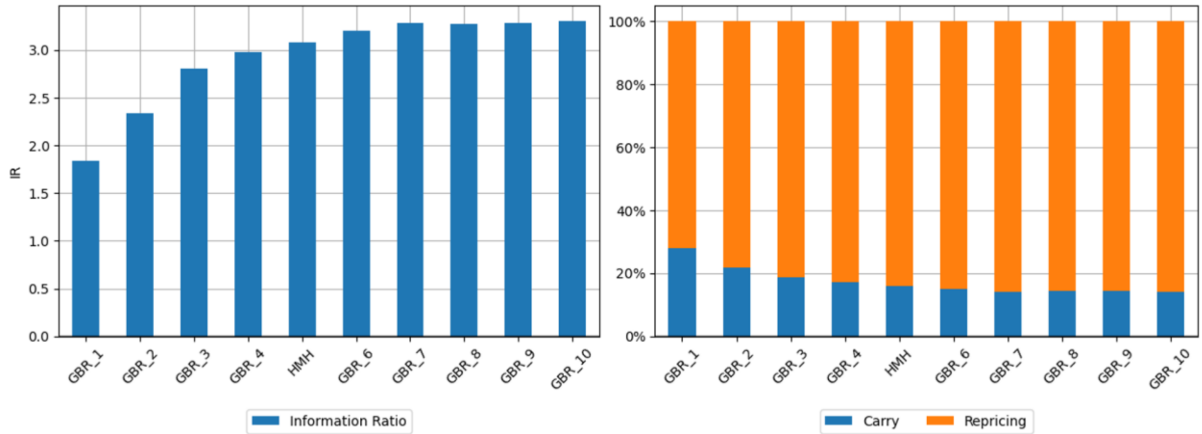
Table 8: Robustness of top-minus-bottom performance

		BGN	BWWSS	CRT	HAHLC	HLIJ	HZ	IPR	KMW	LHB	PDD	SPC	HMH
Information ratio	Base Case	0.46	-0.04	0.35	1.56	0.47	0.52	1.92	1.37	0.91	0.67	0.98	**3.13**
	Deciles	0.47	-0.07	0.35	1.61	0.46	0.56	1.59	1.40	0.81	0.51	0.83	**3.50**
	12m	0.43	0.23	0.32	0.65	0.37	0.39	0.98	0.89	0.35	0.47	0.63	**1.68**
	Top-Market	0.43	0.02	0.35	1.20	0.50	0.46	1.26	1.17	0.75	0.58	0.70	**2.10**
	Value weighted	0.43	-0.10	0.27	1.42	0.43	0.45	1.75	0.96	0.74	0.66	0.85	**3.17**
	IG only	0.35	-0.19	0.18	1.53	0.32	0.48	1.62	0.94	0.91	0.82	0.91	**3.03**
	HY only	0.31	-0.15	0.27	0.96	0.38	0.44	0.62	0.49	0.56	0.39	0.84	**2.41**
	TRACE	0.32	0.03	0.27	1.26	0.37	0.26	1.56	1.04	0.56	0.42	0.52	**2.12**
Repricing %	Base Case	42%	-914%	-16%	72%	11%	35%	61%	71%	58%	54%	65%	**85%**
	Deciles	37%	-306%	-20%	71%	5%	37%	50%	69%	48%	40%	57%	**85%**
	12m	35%	-45%	-31%	43%	-15%	13%	33%	42%	7%	29%	38%	**65%**
	Top-Market	32%	-1,453%	-17%	66%	13%	25%	43%	54%	41%	36%	45%	**83%**
	Value weighted	54%	-127%	-32%	76%	10%	38%	66%	62%	62%	60%	66%	**85%**
	IG only	76%	0%	-43%	**89%**	17%	46%	84%	71%	70%	67%	73%	88%
	HY only	-7%	-96%	-58%	54%	-22%	-4%	-3%	**118%**	11%	-7%	37%	79%
	TRACE	51%	-717%	-18%	73%	13%	10%	59%	78%	62%	59%	68%	**85%**

Note: Sample period 1994-2022. Backtests with different configurations are run and the information ratio and repricing fraction of the top-minus-bottom portfolios is tracked for each value factor. The repricing return is calculated in accordance with the credit return decomposition. All tests cover the 1994-2022 period, except for the TRACE robustness test, which covers the 2015-2020 period. The highest value in each row is shown in bold.

Next, we want to show that the HMH results are robust to changing the hyperparameters of the underlying ML model. For this purpose, we keep the number of trees at 400 and the learning rate at 0.1 and vary the tree depth from 1 to 10, since the three hyperparameters are somewhat interchangeable: increasing the number of trees or the tree depth increases the complexity, while increasing the learning rate reduces the number of steps needed to arrive at a more complex model. Figure 8 shows the corresponding results. We can see that the results are robust to the exact specification and that we have not simply chosen the best model. There is initially a trade-off between model complexity and the risk-adjusted performance and the repricing fraction: the more complex the model, the higher the IR and the repricing fraction. However, this trade-off flattens out as the tree depth increases, indicating that the model begins to overfit the training data. Blitz et al. (2023) recommend in such cases to choose the simplest model that achieves the required level of complexity for the modeling task, which here means a tree depth much higher than 1 but lower than 10.

Figure 8: HMH performance sensitivity to changing tree depth



Note: Sample period 1994-2022. Backtests for variants of HMH with tree depth ranging from 1 to 10 are run and the information ratio and repricing fraction of the top-mins-bottom portfolio is tracked. The repricing return is calculated in accordance with the credit return decomposition.

It can be concluded that the presented results are robust as they hold up to changes in portfolio construction, changes in the sample and model changes.

Conclusion

This paper examines the effectiveness of value investing in the credit market. Given the lack of a standard value factor definition for credit, we replicate a comprehensive set of value factors and evaluate their performance. We find heterogeneity in the cross-section of value factors in terms of their definitions and signal correlations. Nevertheless, we can replicate results of earlier papers and find that most value factors perform well in our extended sample as 9 out of 11 value factors show positive and significant CAPM alpha's at the 5% significance level.

To understand what drives the performance of the value factors, we utilize a decomposition of credit returns into a carry component, which is mainly driven by taking more risk, and a repricing component, which is mainly driven by initial mispricings. This decomposition shows that existing value factors earn substantially from both components. We also show that value factors that earn more from the repricing component tend to have higher risk-adjusted returns. This supports the idea that the repricing component is closely related to mispricing. Therefore, we conclude that a "true" value factor should derive (most of) its performance from repricing.

In addition to measuring the degree of repricing, we also explore how to further increase (decrease) the repricing (carry) return of the value factor. To this end, we go beyond linear models (with risk overlays) and introduce a machine learning-based value factor that is able to outperform existing value factors and better identify mispricings: 85% of its performance comes from repricing, it controls better for risk, it achieves an IR of 3.13 and earns a CAPM alpha of 4.17% (t-stat of 20.21) per year. These results hold up under several robustness checks.

Our study shows that investors who use systematic strategies for corporate bonds should take a close look at whether they are actually value investors or carry investors in disguise (or a mix of both). Fortunately, our introduced value factor based on state-of-the-art machine learning techniques helps bringing investors closer to harvesting “true” value in the credit market.

Bibliography

Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). “Value and momentum everywhere.” *The Journal of Finance* 68 (3): 929-985.

Bartram, Söhnke, Mark Grinblatt, and Yoshio Nozawa. 2023. “Book-to-Market, Mispricing, and the Cross-Section of Corporate Bond Returns”, *Working Paper*.

Bektic, Demir, Josef-Stefan Wenzler, Michael Wegener, Dirk Schiereck, and Timo Spielmann. 2019. “Extending Fama-French Factors to Corporate Bond Markets.” *The Journal of Portfolio Management* 45 (3): 141-158.

Bessembinder, H., Kahle, K. M., Maxwell, W. F., & Xu, D. (2008). Measuring abnormal bond performance. *The Review of Financial Studies* 22 (10): 4219-4258.

van Binsbergen, J. H., Nozawa, Y., & Schwert, M. (2023). “Duration-based valuation of corporate bonds.” *Working Paper*.

Blitz, D., Hoogteijling, T., Lohre, H., & Messow, P. (2023). “How can machine learning advance quantitative asset management?” *The Journal of Portfolio Management* 49 (7): 1-18.

Choi, J. (2013). “What drives the value premium?: The role of asset risk and leverage.” *The Review of Financial Studies* 26 (11): 2845-2875.

Correia, Maria, Scott Richardson, and Irem Tuna. 2012. “Value Investing in Credit Markets.” *Review of Accounting Studies* 17 (3): 572-609.

Dick-Nielsen, Jens. 2009. “Liquidity biases in TRACE.“ *The Journal of Fixed Income* 19 (2): 43-55.

Dick-Nielsen, Jens. 2014. “How to Clean Enhanced TRACE Data.“ *Working Paper*.

Fama, Eugene and Kenneth French. 1992. “The Cross-Section of Expected Stock Returns.” *The Journal of Finance* 47 (2): 427-465.

Fama, Eugene and Kenneth French. 1993. "Common risk factors in the returns on stocks and bonds." *Journal of financial economics* 33 (1): 3-56.

Hanauer, Matthias, Marina Kononova, and Marc Steffen Rapp. 2022. “Boosting agnostic fundamental analysis: Using machine learning to identify mispricing in European stock markets.” *Finance Research Letters* 48: 1-10.

Heckel, Thomas, Zine Amghar, Isaac Haik, Olivier Laplènie, and Raul Leote de Carvalho. 2020. “Factor Investing in Corporate Bond Markets: Enhancing Efficacy Through Diversification and Purification!” *The Journal of Fixed Income* 29 (3): 6-21.

Hottinga, Jouke, Erik van Leeuwen, and Judith van IJserloo. 2001. “Successful Factors to Select Outperforming Corporate Bonds.” *The Journal of Portfolio Management* 28 (1): 88-101.

Houweling, Patrick and Jeroen van Zundert. 2017. “Factor Investing in the Corporate Bond Market.” *Financial Analysts Journal* 73 (2): 100-115.

Israel, Ronen, Diogo Palhares, and Scott Richardson. 2018. “Common Factors in Corporate Bond Returns.” *Journal of Investment Management* 16 (2): 17-46.

Jegadeesh, Narasimhan and Sheridan Titman. 1993. “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency.” *The Journal of Finance* 48 (1): 65-91.

Kaufmann, Hendrik, Philip Messow, and Frederik Wisser. 2023. “Putting Credit Factor Investing into Practice.” *The Journal of Portfolio Management* 49 (2): 188-215..

L’Hoir, Mathieu and Mustafa Boulhabel. 2010. “A Bond-Picking Model for Corporate Bond Allocation.” *The Journal of Portfolio Management* 36 (3): 131-139.

Lakonishok, J., Shleifer, A., & Vishny, R. W. (1994). “Contrarian investment, extrapolation, and risk.” *The Journal of Finance* 49 (5): 1541-1578.

Piotroski, J. D., & So, E. C. (2012). “Identifying expectation errors in value/glamour strategies: A fundamental analysis approach.” *The Review of Financial Studies* 25 (9): 2841-2875.

Polbennikov, Simon, Albert Desclèe, and Mathieu Dubois. 2021. “Implementing Value and Momentum Strategies in Credit Portfolios.” *The Journal of Portfolio Management* 47 (2): 82-98.

Shen, Shawn, Arom Pathammavong, and Alex Chen. 2019. “Fixed-Income Value Factor.” *The Journal of Fixed Income* 29 (1): 21-43.

Zhang, L. (2005). “The value premium.” *The Journal of Finance* 60 (1): 67-103.

Appendix A: Value factor replication

We perform a replication study of a large set of value factors in the credit market. Table A1 provides an overview of the eleven selected papers from the literature that are included in the replication study.

Table A1: Literature on value investing in the corporate bond market

Acronym	Year	Title
HLIJ	2001	Successful Factors to Select Outperforming Corporate Bonds
LHB	2010	A Bond-Picking Model for Corporate Bond Allocation
CRT	2012	Value Investing in Credit Markets
HZ	2017	Factor Investing in the Corporate Bond Market
IPR	2018	Common Factors in Corporate Bond Returns
BWWSS	2019	Extending Fama-French Factors to Corporate Bond Markets
SPC	2019	Fixed-Income Value Factor
HAHLC	2020	Factor Investing in Corporate Bond Markets: Enhancing Efficacy Through Diversification and Purification!
PDD	2021	Implementing Value and Momentum Strategies in Credit Portfolios
KMW	2022	Putting Credit Factor Investing Into Practice
BGN	2023	Book-to-Market, Mispricing, and the Cross-Section of Corporate Bond Returns

In order to replicate the value factors from the different papers as closely as possible, we use the following systematic process:

- First, we thoroughly read the paper and note down all the relevant implementation details.
- Second, we label each step in the methodology as “signal construction” or “portfolio construction”.
- Third, we implement the value factor by following the signal construction steps as closely as possible. We chose not to implement the portfolio construction steps as our aim is to compare the standalone value factors and not the wide variety of implementations, which would add another layer of complexity to the comparison.

It is important to mention that the papers, alongside different portfolio construction methodologies, use a variety of security types (corporate bonds and credit default swaps), data providers (e.g. Bloomberg Barclays and ICE BofA), selection criteria (e.g. excluding financials, subordinated bonds, high yield bonds, or privates) and sample periods. This means that it is not trivial to exactly replicate the tables and statistics from the original papers to assess the success of the replication. Moreover, the value factors also might not have been specifically designed for the sample used in our paper, resulting in potentially different results than on the original sample. These differences in samples, however, can also be viewed as positive, as it allows us to compare the value factors using considerable out-of-sample data, making this replication study a particularly good test on the robustness and effectiveness of value investing in the credit market.

Appendix B: Machine Learning based Value factor

Table B1: Features used for the estimation of HMH

Feature	Description
Bond Option Adjusted Duration	Bond Duration adjusted for embedded options
Bond Excess Return Volatility 12m	Bond Excess return volatility of the last 12month
Equity Volatility	Equity Volatility based on 130 day historical volatility
Equity Momentum 1m-12m	One-month Equity Momentum - Twelve-month Equity Momentum
B/P	Book to Price
Texas Ratio	Texas Ratio is the ratio of total non-performing assets to the sum of tangible common equity and loan loss reserves
FCF to Interest Expenses	Free Cashflow to Interest Expenses
FCF to Debt	Free Cashflow to Debt
Return on Equity	Return on Equity
Debt to Assets	Debt to Assets
Analyst Recommendation	Average of Analyst Recommendations
Retained Earnings to Assets	Retained Earnings to Total Assets
Earnings Yield	Earnings Yield: Analyst estimate of the twelve-month EPS divided by stock price.
Earnings Yield Variability	Earnings Variability: Deviation of the earnings yield from its 24 months moving average divided by its rolling standard deviation
Short Term Debt to Debt	Short Term Debt to Total Debt
External Financing	Ratio of change in external capital to Total Assets
FCF to NOA	Free Cashflow to Net Operating Assets
Cash to Assets	Cash to Total Assets
Long Term Debt	Long Term Debt
Coverage Ratio	Interest Payments to EBIT
Tier1	Tier 1 capital
FCF to NOA 1y change	Relative change in Free Cashflow to Net Operating Assets over one year
Return on Assets	Return on Assets
Gross Profits to Assets	Gross Profits to Total Assets
Rating	Rating (AAA to CCC-)
Sector	Bloomberg sector classification