PORTFOLIO
MANAGEMENT
RESEARCH
By With Intelligence

the journal of PORTFOLIO management

QUANTITATIVE STRATEGIES: *FACTOR INVESTING*

7TH EDITION

volume 48 *number* 2

JANUARY 2022

jpm.pm-research.com

Factor Investing in Sovereign Bond Markets: Deep Sample Evidence

Guido Baltussen, Martin Martens, and Olaf Penninga

Robeco is an international asset manager offering an extensive range of active investments, from equities to bonds. Research lies at the heart of everything we do, with a 'pioneering but cautious' approach that has been in our DNA since our foundation in Rotterdam in 1929. We believe strongly in sustainability investing, quantitative techniques and constant innovation.

ROBECO
The Investment Engineers

Factor Investing in Sovereign Bond Markets: Deep Sample Evidence

Guido Baltussen, Martin Martens, and Olaf Penninga

Guido Baltussen
is a professor in finance at Erasmus University Rotterdam and executive director at Robeco Institutional Asset Management in Rotterdam, The Netherlands.
g.baltussen@robeco.nl

Martin Martens
is director at Robeco Institutional Asset Management in Rotterdam, The Netherlands.
m.martens@robeco.nl

Olaf Penninga
is executive director at Robeco Institutional Asset Management in Rotterdam, The Netherlands.
o.penninga@robeco.nl

KEY FINDINGS

- Value, momentum, and low-risk factors offer attractive premiums in bond markets between 1800 and 2020.
- Bond factor premiums are robust in-sample and out-of-sample, across periods of rising or declining yields, and in other market or macroeconomic states.
- A combined multifactor bond strategy delivers strong value-added to a passive portfolio.

ABSTRACT

The authors examine government bond factor premiums in a deep global sample from 1800 to 2020, spanning the major markets and maturities. Bond factors (value, momentum, low-risk) offer attractive premiums that do not decay across samples, are persistent over time, and are consistent across various market and macroeconomic scenarios. The factor premiums are diversified to each other, as well as to bond or equity market risks. A combined multifactor bond strategy provides the strongest risk-adjusted returns. These results strongly show a consistent added value of government bond factor premiums over a passive bond portfolio.

Several studies have showed that factor premiums are persistent phenomena in markets. Many of these studies examined equity factors (e.g., Fama and French 1992, 2015; Blitz 2012); however, recently, several papers also have showed individual factors to work well in credit markets (e.g., Houweling and Van Zundert 2017) or across asset classes (e.g., Baltussen, Swinkels, and Van Vliet 2021). Meanwhile, the size of factor investments has grown tremendously in the industry.[1] However, to date, relatively less is known about factor premiums in government bonds markets, with investors having been slower to adopt factor investing, this while government bonds are one of the major asset classes in the world, with their size representing about 30% of overall market capitalizations across asset classes (Doeswijk, Lam, and Swinkels 2020). Which factors are present in government bond

[1] Estimates of the amount of money invested in factor strategies vary from one source to another, ranging from $1 trillion to $2 trillion globally in most cases. In a report published in 2017, Morgan Stanley estimated that almost $1.5 trillion was invested in smart beta, quant, and factor-based strategies and that assets under management have been growing by 17% per year on average since 2010. According to a survey by FTSE Russell, 58% of asset owners worldwide had implemented smart beta—in other words, factor-based strategies—in their portfolios in 2019.

markets? Are they persistent over time? What is their added value? In this article, we extensively examine global bond factor premiums and answer these questions over a deep sample spanning 221 years across the major government bond markets.

To date, relatively few studies have examined government bond factor premiums. Asness, Moskowitz, and Pedersen (2013) were among the first to investigate the use of value and cross-sectional momentum in asset classes other than equities, including government bonds. Koijen et al. (2018) demonstrated that carry explains the cross-section of returns in numerous asset classes, including government bonds. These studies examine government bond factors typically over a limited sample, focus on one particular factor, or only consider a long–short perspective. For example, the aforementioned papers roughly cover the 30-year period of 1982–2012.[2] However, this sample has been unique, with few major episodes of bond market crises, economic recessions, and inflationary episodes. Since 1980, yield levels have displayed a secular decline in most markets. As a result, a key question is how bond factor premiums are influenced by falling or rising yield levels, and other episodes that are typically a concern for investors. In addition, several studies have argued that published factor premiums could be influenced by p-hacking (see Harvey 2017).[3] As a result, published findings might reflect a type I error in testing (i.e., falsely discovering predictability) and may fail to hold out-of-sample.

To address these concerns, we use an extensive historical sample that spans all major government bond markets from developed countries over a 221-year period (January 1800–December 2020). Basically, we have 190 years of additional data to put the published results to the test. In total, we have 35,784 monthly return observations in our sample, thereby providing us with sizable testing power to examine bond factor premiums. Moreover, over our sample period, global bond yields displayed several secular rates cycles, as illustrated by the development of the global average 10-year yield based on France, Germany, Japan, the United Kingdom, and the United States in Exhibit 1. Roughly post-1980, global government bond yields displayed a secular decline across the world. However, before this point, yields displayed different behavior, with also secular rises in yields. Thus, our study provides a natural robustness test of the influence of secular yield trends on bond factor premiums.

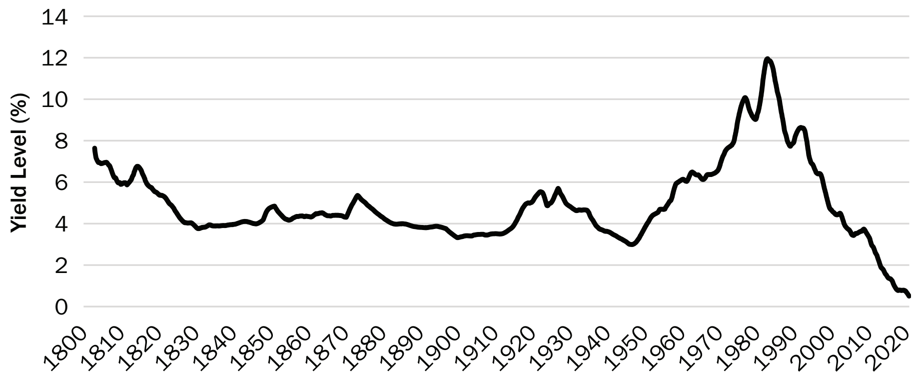
In this study, we focus on three key bond factors—value, momentum, and low risk—and apply them either across developed market bonds or within a market on the bond yield curve. These factors are typically considered key factors within the industry (see, e.g., Blitz, Baltussen, and Van Vliet 2020 and Houweling and Van Zundert 2017), have been documented in previous studies, and have sufficient coverage over a substantial part of our sample period and across the markets we study. Value and momentum are applied across bond markets, following Asness, Moskowitz, and

[2] In addition, Brooks, Palhares and Richardson (2018) and Kothe, Lohre, and Rother (2020) covered carry, momentum and value for government bonds, but for the even smaller samples of 1997–2017 and 1994–2019, respectively.

[3] P-hacking refers to the conscious or unconscious misuse of data analysis to find patterns in data. As a case in point, Harvey, Liu, and Zhu (2016) found a clear publication bias pattern in the top finance journals; of over 300 documented stock-level anomalies, many become questionable after analysis in a rigorous testing framework that allows for multiple hypotheses testing bias. P-hacking is not limited to financial economics but is mostly discussed in social sciences and medicine. *The Economist* discussed the topic in 2013 with the headline title "How Science Goes Wrong." Begley and Ellis (2012) showed that out of 53 studies on preclinical cancer, only 11% could be replicated. An open science collaboration in 2015 showed that, out of 97 significant psychological studies, only 36 could be replicated. In behavioral economics, Camerer et al. (2016) found that out of 18 laboratory studies in economics, only 11 could be replicated with similar findings.

EXHIBIT 1

Average 10-Year Yields 1800–2020



NOTE: The exhibit shows the three-year rolling average of the 10-year yields for France, Germany, Japan, the United Kingdom, and the United States from January 1800 to December 2020.

Pedersen (2013),[4] and low risk[5] is applied on the bond curve because Frazzini and Pedersen (2014) showed that low risk is present on the bond curve but is weak to absent in an across-bond-market setting. We keep these factors and their definitions unchanged over our out-of-sample period to have a reliable and robust assessment of bond factor premiums, even in the wake of p-hacking.

Our findings are as follows. We find that value, momentum, and low risk offer attractive factor premiums, with Sharpe ratios of 0.51, 0.24, and 0.40 over our full sample. Moreover, these factor premiums are consistent over time, being positive in 72% (momentum) to 92% (value) of 10-year rolling periods. Combining the factors into a simple multifactor portfolio gives a highly significant Sharpe ratio of 0.56 (t-statistic of 8.22) from 1800 to 2020 and is positive in 89% of the 10-year rolling periods. In other words, factor strategies in government bonds offer attractive returns and diversify each other.

Next, we show that a multifactor bond portfolio gives robust performance over various macroeconomic states that are typically a concern for investors. These include recessions and expansions, crisis and non-crisis periods, years with either rising or declining yields, years with above or below median inflation, and years with positive or negative equity returns. Note that the previously published papers cover a period with declining yields. It is therefore important to examine performance in years with rising yields and declining yields.

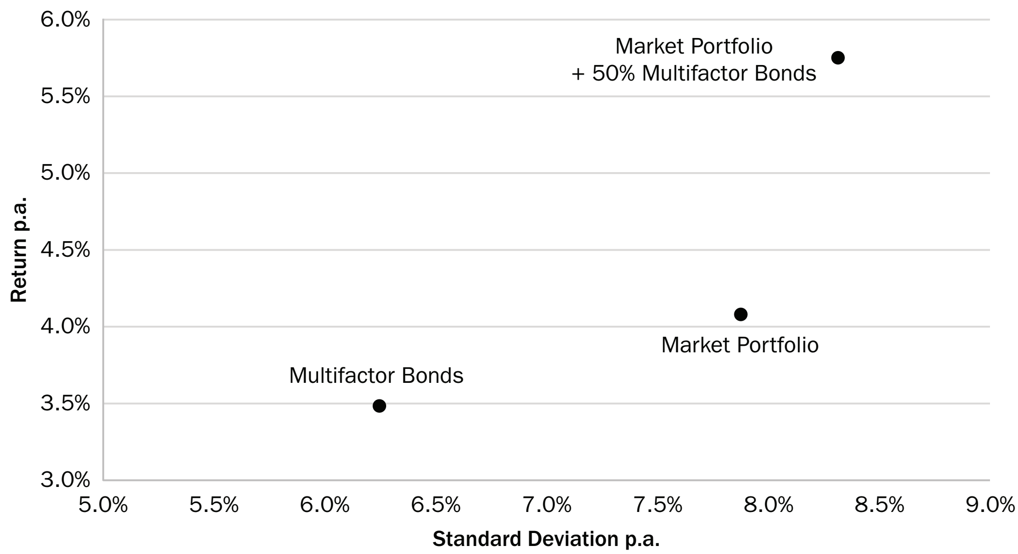
Finally, we evaluate bond factor premiums in a portfolio context. When considering a long-only bond or multi-asset investor who considers adding bond factor premiums, we find strong value-added of a multifactor bonds portfolio. Exhibit 2 summarizes the benefits. A multifactor bond portfolio has an average return of 3.48% a year at a correlation of –0.05 with the bond market, and adding a multifactor bond strategy to a passive global government bond portfolio increases returns substantially with little impact on risk. The multifactor bond portfolio focuses on country and curve allocation and hence does not take a stance on the direction of the bond market, but only on the relative attractiveness of the different bond markets or maturities. Of course, there are also studies showing that bond market timing can be successful as well; see, for

[4] Asness, Moskowitz, and Pedersen (2013) related value to the yield level to inflation (*real yield*), and also used the term spread on a bond curve, which is very similar to the definition of carry covered by Koijen et al. (2018). When we refer to value, we mean both the real yield and carry.

[5] Frazzini and Pedersen (2014) referred to this phenomenon as *Betting-Against-Beta* (BAB).

EXHIBIT 2

Return and Risk for the Bond Market Portfolio and Bond Factor Premiums



NOTES: The exhibit shows the average annualized returns and standard deviations for the global bond market portfolio (Market Portfolio), the global multifactor bonds portfolio (Multifactor Bonds)—which combines the momentum, value, and low-risk factors—and a combination of the bond market portfolio and 50% multifactor bonds. The sample period is 1800–2020.

example, Baltussen, Martens, and Penninga (2021), who provided an out-of-sample test for timing variables introduced by Ilmanen (1995) and others. We find that one of those timing variables that we can test from 1800 onward, time-series momentum, can further add value to the combination of long-only and multifactor bonds.

Our work is closely related to that of Baltussen, Swinkels, and Van Vliet (2021), who examined global factor premiums across all major asset classes. Compared to their study, we focus on the most important factors documented in government bond markets: value and momentum across bond markets (i.e., country allocation) and low risk on the bond curve (i.e., maturity selection). Furthermore, we examine government bond factor premiums in depth, allowing us to share more insights specific to the bond market, and take an investor perspective, considering both single and multifactor portfolios. In addition, we show how an investor can benefit when moving from passive investing in the bond market to factor-based investing in bonds, from both a long–short and a long-only investor constraint perspective.

BOND FACTORS AND DATA

Factor Definitions

In this study, we focus on three key government bond factors, value, momentum, and low risk. Asness, Moskowitz, and Pedersen (2013) covered cross-sectional momentum and value for multiple asset classes. For momentum, we use their definition, which is to compute the past 12-month returns minus the last month. For value, the paper covers three definitions: five-year mean reversion, term spread, and the real yield, which is defined as the 10-year bond yield over the past one-year inflation. Their results for combining the three value measures are much more promising than the stand-alone results for five-year mean reversion. We therefore focus on the real yield (value 1) and term spread (value 2) as our value measures. The term spread, the bond yield minus the risk-free rate, is closely related to carry. The term spread omits the roll-down from Koijen et al. (2018), which is not calculable for the long historical dataset we have. For selecting countries, however, they showed that the term spread and carry are 94% correlated. The method to construct long–short factor portfolios for both value measures and momentum is described in Appendix B.

Finally, we follow Frazzini and Pedersen (2014) for low risk on the US curve. We test this factor by estimating the betas over a 36-month period (requiring at least 12 months of data) for the different maturities. We then go long the low beta maturities and short the high beta maturities.[6] The position sizes of each short and long leg are chosen such that the ex ante betas of both legs are equal, such that the factor has as little overall market effect as possible.

Data

Our dataset is described in detail in Appendix A, including the 16 developed government bond markets we cover. It also describes the data sources, with Global Financial Data and Macrohistory.net providing the deep history and Datastream and Bloomberg more recent data. In terms of data quality checks and cleaning data, we follow Baltussen, Swinkels, and Van Vliet (2021), who extensively analyzed data quality and applied multiple data filters to ensure a dataset of good quality. Most importantly, they applied a number of conservative screens on the data series and removed data points that did not pass these screens. These screens are (1) a zero-return screen, which leaves out data series with more than one zero or missing spot return observation in the past 12 months; (2) a return interpolation screen, which leaves out identical returns from one month to the next month; and (3) a stale return screen, which leaves out observations that do not have nine or more differentiating returns over the past 12 months. Note that some months, mostly in the 19th century, are missing because of these data filters. Some specific examples for bonds are provided in Appendix A. Momentum and the term spread start in 1800; the real yield, making use of inflation data, starts in 1872. For the low-risk strategy on the US curve, Global Financial Data provides bond returns for maturities of 1, 2, 3, 5, 10, and 30 years, along with a three-month short rate series. The shortest three maturities start in 1941, and the longer maturities start in 1919, allowing us to start the low-risk factor returns from 1922 onward (because we require a 36-month beta estimation window).

THE DEEPEST SAMPLE EVIDENCE ON BOND FACTOR PREMIUMS: 1800–2020

Recent Sample Results: Existing Evidence and Replication

We start our analyses by studying recent sample evidence for the individual government bond factors: value, momentum, and low risk. In the left-hand side of Exhibit 3, we show the results existing studies documented over their sample periods (i.e., Asness, Moskowitz, and Pedersen 2013; Frazzini and Pedersen 2014; Koijen et al. 2018). We show the start and end date of their sample period, the variable definition they used, and the Sharpe ratio they reported. To that, we add a *t*-statistic based on multiplying the Sharpe ratio by the square root of the number of years in the

[6] For a curve strategy, this is equivalent to going long the lowest maturity bonds and short the highest maturity bonds. Note that there is not a long history on durations; as such, it is more convenient to use betas.

EXHIBIT 3

Replication of Existing Studies for Government Bond Factor Premiums

			From Existing Studies			Our Replication		
	Start Date	End Date	Definition	Sharpe	*t*-Stat	Definition	Sharpe	*t*-Stat
Momentum	January 1982	July 2011	12M-1M	*0.06*	0.35	12M-1M	0.15	0.82
Value 1	January 1982	July 2011	5Y Reversal	*0.18*	0.97	Real Yield	0.29	1.59
Value 2	November 1983	September 2012	Slope + Roll	*0.52*	2.79	Term Spread	0.47	2.52
Low-Risk	February 1953	March 2012	Beta	*0.81*	*6.26*	Beta	0.41	3.12

NOTES: The exhibit contains sample start and end dates, factor definitions (Definition), Sharpe ratios (Sharpe), and *t*-statistics (*t*-Stat) for each of the original studies on global factor premiums: Asness, Moskowitz, and Pedersen (2013) for cross-sectional momentum (Momentum) and value (Value 1 and Value 2) and Frazzini and Pedersen (2014) for Low-Risk. Asness, Moskowitz, and Pedersen used five-year reversal for their main results and showed results for the combination of 5Y Reversal, Real Yield, and Term Spread (but not separately). We therefore base Value 2 on the work of Koijen et al. (2018). Sharpe ratios in italics are taken from these studies based on their results for government bonds or bond futures. The last three columns are based on our own calculations based on our own data; see Appendix A for details.

sample (see Sharpe 1994). In the right-hand side of Exhibit 3, we show the definitions we use and the Sharpe ratios and *t*-statistics we find based on our dataset.

The mentioned studies document Sharpe ratios of 0.06, 0.18, 0.52, and 0.81 for momentum, value 1, value 2, and BAB. The value 2 and BAB Sharpe ratios are statistically significant with *t*-statistics of 2.79 and 6.26. It is good to put the size of the Sharpe ratios into perspective. Dimson, Marsh, and Staunton (2021) reported Sharpe ratios of around 0.25 and 0.10 over the period 1900–2020 for world equity and bond markets, respectively. Note that we would need 61 years to get a *t*-statistic of 1.96 for a Sharpe ratio of 0.25.

We perform a replication exercise on these studies based on our sample. As mentioned before, we use the same definitions for momentum and low risk and use real yield and term spread as value measures. There are also some important differences between the existing studies and our replication exercise. Most notably, we cover 16 countries compared to 10 countries covered by Asness, Moskowitz, and Pedersen (2013). The six additional countries are Belgium, France, Italy, the Netherlands, Spain, and New Zealand. In the right-hand side of Exhibit 3, we show the results.

We find a higher Sharpe ratio for momentum (0.15 versus 0.06). For value 1 (i.e., real yield), we find a Sharpe ratio of 0.29 compared to the reported 0.18 for five-year mean reversion, whereas we find a Sharpe ratio of 0.47 for value 2 (i.e., term spread), similar to the 0.52 from Koijen et al. (2018). For low-risk, we get a lower Sharpe ratio of 0.41, compared to 0.81 from Frazzini and Pedersen (2014), albeit both highly significant. Frazzini and Pedersen used the CRSP Fama bond maturity portfolios covering one- to five-year maturities, whereas we also include 10- and 30-year maturities. This means that we look at long 1–2–3-year versus short 5–10–30-year bonds, whereas Frazzini and Pedersen looked at 1–2-year versus 4–5-year bonds. Hence, large differences are expected. Owing to 10- and 30-year bonds, we can cover a longer history, and covering the full range of maturities also makes sense from a practitioner's perspective because bond indexes also cover the full range of maturities.

Deep-Sample Sample Evidence

Our dataset allows for a very large out-of-sample test for the numbers on the right-hand side in Exhibit 3. We look at both data from 1800 until the start date of the global factor premium studies as provided in column 3 of Exhibit 3 and at newer

EXHIBIT 4
Bond Factor Premiums: 1800–2020

		Replication		Deep Sample		Full Sample	
Factor	**Measure**	**Sharpe**	***t*-Stat**	**Sharpe**	***t*-Stat**	**Sharpe**	***t*-Stat**
Momentum	12M-1	0.15	0.82	0.25	3.38	0.24	3.46
Value 1	Real Yield	0.29	1.59	0.18	1.98	0.20	2.44
Value 2	Term Spread	0.47	2.52	0.56	7.64	0.54	8.02
Low-Risk	Beta	0.41	3.12	0.38	2.29	0.40	3.85

NOTES: For 12M-1M momentum (Momentum), the real yield (Value 1), the term spread (Value 2), and beta (Low-Risk), we look at Sharpe ratios (Sharpe) and *t*-statistics (*t*-Stat) in three samples. The Replication sample period covers the replication results over the sample period covered by the global factor premiums studies, Asness, Moskowitz, and Pedersen (2013), Koijen et al. (2018), and Frazzini and Pedersen (2014) (see Exhibit 3). The Deep Sample period covers earlier data starting in January 1800 (Momentum and Value 2), 1872 (Value 1), or 1922 (Low Risk) and newer data ending in December 2020, but leaving out the replication sample data. The Full Sample period covers all data.

data from the end dates in Exhibit 3 until the end of 2020. The results are provided in Exhibit 4.

The deep-sample test generally shows strong results. For momentum, we see a highly significant Sharpe ratio of 0.25, higher than the 0.15 for the 1982–July 2011 replication sample period. For value based on real yields, the Sharpe ratio is 0.18, lower than the 0.29 for the replication sample period. It is, however, still significant at the 5% significance level. For value based on the term spread, the Sharpe ratio is a highly significant 0.56 for the deep-sample, compared to 0.47 for November 1983–September 2012. Finally, the Sharpe ratio for low risk is 0.38 (*t*-statistic 2.29), quite similar to the 0.41 replication-sample Sharpe ratio. The final two columns in Exhibit 4 show the full sample results. For all four factors, we see significant Sharpe ratios ranging from 0.20 for value 1 to 0.54 for value 2, with generally high *t*-statistics (from 2.44 to 8.02).

Multifactor Bonds

Next, we examine the combination of the previously mentioned individual factors in a multifactor portfolio. To this end, we combine the four bond factor returns using an equal weighting scheme:

$$R_{Multifactor,t} = 0.25R_{Momentum,t} + 0.25R_{Value1,t} + 0.25R_{Value2,t} + 0.25R_{Low-Risk,t} \tag{1}$$

Before the start of low risk in 1922, we use an equal-weighted combination of the other factors. A remark is in order about the weight of value because one could argue that value gets a higher weight than momentum and low risk. Our motivation is simplicity combined with having two value measures, which are quite distinct because their correlation equals –0.18. Overall, correlations between factor returns are limited. The highest correlation is 0.34 between momentum and value 2, which is related to the fact that momentum is based on total returns, and these returns depend on both the term spread (our value 2 measure) and yield changes. The lowest correlation is –0.39 between momentum and value 1. The correlations between low risk and the other three components in Equation 1 are very close to zero. The results that follow are robust to using different weighting schemes to the individual factors, being qualitatively similar when we first build one value basket and then give one-third weight to each factor, or when using mean–variance optimal weights. Exhibit 5 shows the

EXHIBIT 5

Multifactor Bonds

Factor	Momentum	Value 1	Value 2	Low-Risk	Multifactor
Panel A: 1800–2020					
Return p.a.	2.58%	2.39%	5.84%	–	3.48%
Stdev p.a.	10.93%	11.94%	10.73%	–	6.25%
Sharpe Ratio	0.24	0.20	0.54	–	0.56
t-Statistic	3.46	2.44	8.02	–	8.22
Panel B: 1922–2020					
Return p.a.	1.81%	3.38%	6.71%	4.46%	3.98%
Stdev p.a.	11.42%	11.18%	11.42%	11.27%	5.53%
Sharpe Ratio	0.16	0.30	0.59	0.40	0.72
t-Statistic	1.58	3.01	5.83	3.85	7.16

NOTES: In this exhibit, we examine the results for 12M-1M Momentum (Momentum), real yield (Value 1), term spread (Value 2), and beta (Low-Risk) for the 1800–2020 and the 1922–2020 sample periods. In addition, we look at the equally weighted average multifactor bonds combination (Multifactor). We show average returns, standard deviations, Sharpe ratios, and *t*-statistics per factor and their combination for two sample periods. Momentum and Value 2 start in 1800, Value 1 in 1872, and Low Risk in 1922.

results of the multifactor bond combination over our full sample period (1800–2020) and over the subsample when low risk is available (1922–2020).

Multifactor bonds have a highly significant Sharpe ratio of 0.56 for 1800–2020 and of 0.72 for 1922–2020 when low risk also is available. Owing to the aforementioned low correlations among the four individual factors, we see that the multifactor bonds combination has a substantially lower annualized standard deviation of 6.25%, compared to the average of 11.22% for the individual factors. As mentioned in Appendix B, we always ensure an ex ante volatility of 10% for each individual factor at the start of each month. Ex post, we see that volatilities for the individual factors are just above 10%.

Persistence Over Time

Next, we examine the robustness of bond factor premiums over time. To this end, we first examine the performance over rolling 10-year subperiods. Exhibit 6 summarizes the results in terms of success ratio (i.e., number of 10-year periods with positive performance). The success ratio of all four factors is at least 72%; that is, the factors have a positive performance in at least 72% of the rolling 10-year subperiods. Multifactor bonds has a success ratio of 89%.

Second, we look at the cumulative performance over time in Exhibit 7. We see that, perhaps apart from a modest start in the first decades of the 19th century, when the universe consists of only 7 out of the in total 16 countries, performance is generally stable over time.

MARKET RISK AND FACTOR RETURNS ACROSS GOOD AND BAD STATES

In this section, we look in more detail at the return and risks of factor premiums relative to the bond market and across market and macroeconomic states. To this end, we use our full sample of 221 years of data because this gives us a substantial number of observations across good and bad states compared to a sample of, for example, the most recent 30 years. We first regress the returns of multifactor bonds on the global government bond market portfolio, which we proxy by the equal-weighted

EXHIBIT 6

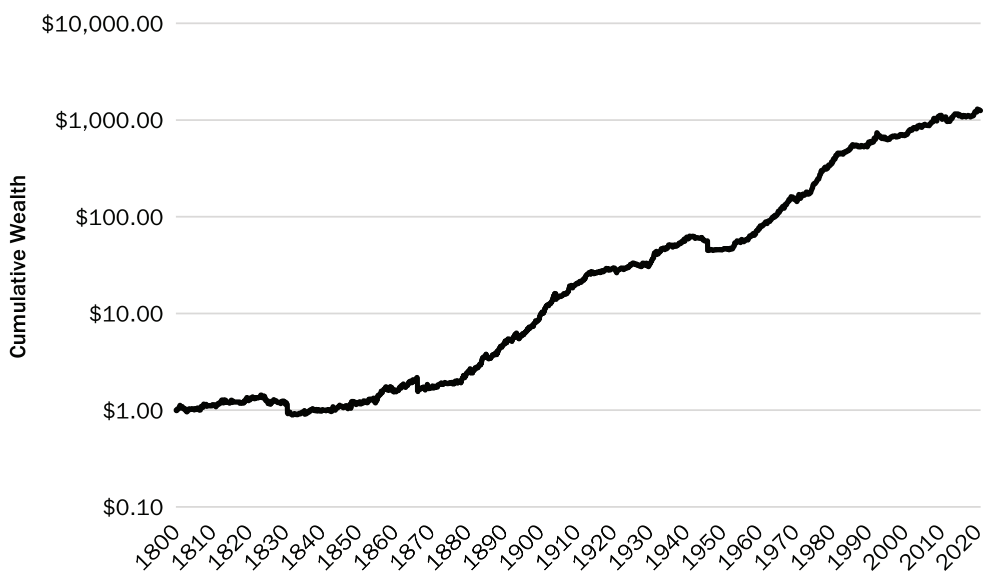
Success Ratios in 10-Year Rolling Periods

Factor	Momentum	Value 1	Value 2	Low-Risk	Multifactor	Market
Success Ratio	72.1%	74.3%	87.7%	92.3%	89.0%	91.9%
No. of Obs.	2,533	1,786	2,533	1,188	2,533	2,533

NOTES: In this exhibit, we report the Success Ratio and number of rolling 10-year subperiods (No. of obs.) of the individual bond factors, their equally weighted Multifactor combination, and the bond market (Market). The success ratio is computed by taking the performance over rolling 10-year (120-month) subperiods and counting how often these are positive, which is subsequently divided by the total number of observations.

EXHIBIT 7

Cumulative Performance Multifactor Bonds



NOTE: The exhibit shows the cumulative wealth on an initial investment of $1 at the end of 1799 in the multifactor bond portfolio that combines momentum, value, and low risk government bond factors.

return over the bond markets included in our sample. Even though we are looking at cross-sectional strategies, they may still benefit from a bottom-up structural bond beta. More specifically, we run the following regression:

$$R_{Multifactor,t} = \alpha + \beta \cdot R_{Global\ Market,t} + \varepsilon_t \tag{2}$$

Exhibit 8 summarizes the results by means of the beta, the appraisal ratio (i.e., the estimated alpha divided by the standard deviation of the residuals of Equation 2), and its *t*-statistic. We present the multifactor bond results and the results for the individual factors by replacing the multifactor returns in Equation 2 with the individual factor returns. In general, we see betas close to zero, indicating there is no structural positive beta that could explain the performance of government bond factor strategies. Because the betas are small, the appraisal ratios are close to the Sharpe ratios in Exhibit 5. A negative beta leads to an appraisal ratio that is slightly higher than the Sharpe ratio, and vice versa. Because multifactor has a small negative beta, the

EXHIBIT 8

Appraisal Ratios

Factor	Momentum	Value 1	Value 2	Low-Risk	Multifactor
Panel A: 1800–2020					
Beta	–0.02	0.06	–0.07	–	–0.04
Appraisal Ratio	0.24	0.18	0.57	–	0.59
t-Statistic	3.58	2.25	8.39	–	8.63
Panel B: 1922–2020					
Beta	0.01	0.04	–0.07	–0.15	–0.03
Appraisal Ratio	0.15	0.29	0.61	0.45	0.75
t-Statistic	0.15	2.86	6.07	4.41	7.41

NOTES: We regress the factor returns on the global market, the average of the individual bond market returns. The exhibit shows the beta of the regression, the appraisal ratio (i.e., the alpha of the regression divided by the standard deviation of the residuals), and its *t*-statistic for the 1800–2020 and the 1922–2020 sample periods.

appraisal ratios of 0.59 for 1800–2020 and 0.75 for 1922–2020 are slightly higher than the Sharpe ratios of 0.56 and 0.72, respectively.

Next, we examine whether perhaps the multifactor bond returns are poor in bad states as a possible explanation for the good performance in the long run. For this, we assign each of the calendar years to a bad or a good state based on:

- Recessions and expansions
- Crisis and non-crisis
- Negative and positive equity performance
- Yields rising or declining
- Inflation above or below the median

Appendix A provides detail on the source of the classifications. Exhibit 9 shows the results.

Multifactor bonds provides, on average, positive returns in both bad and good states. The average return is 1.9% in recessions, 4.1% in crisis periods, 3.9% when yields rise, 4.1% when inflation is high, and 3.1% when equity markets have negative returns. Hence, based on this evidence, the returns of multifactor bonds cannot be explained by poor performance in bad states. From this, we can conclude that a risk-based explanation for the factor premiums seems unlikely.

Perhaps most importantly for investors, the performance of multifactor bonds is stable across the different scenarios, including, for example, falling or rising equity markets (thereby providing evidence of added value over equity markets) and high and low inflation episodes. Interestingly, performance is good regardless of yields rising or falling. Our deep historical sample makes it possible to make such a statement because it also contains multi-year periods with also rising yields, unlike a sample that covers only post-1980 data. As mentioned before, the key academic studies cover a period from 1982 onward, with mainly declining yields. To provide more color on performance during rising and declining yield periods, we show the cumulative performance of multifactor bonds conditional on both scenarios in Exhibit 10. Performance is consistent over time across both scenarios. Moreover, results not shown in the exhibit reveal that all factors contribute positively in both rising and declining yield calendar years; the second value factor (term spread) and low risk in particular perform strongly in periods with rising yields.

EXHIBIT 9

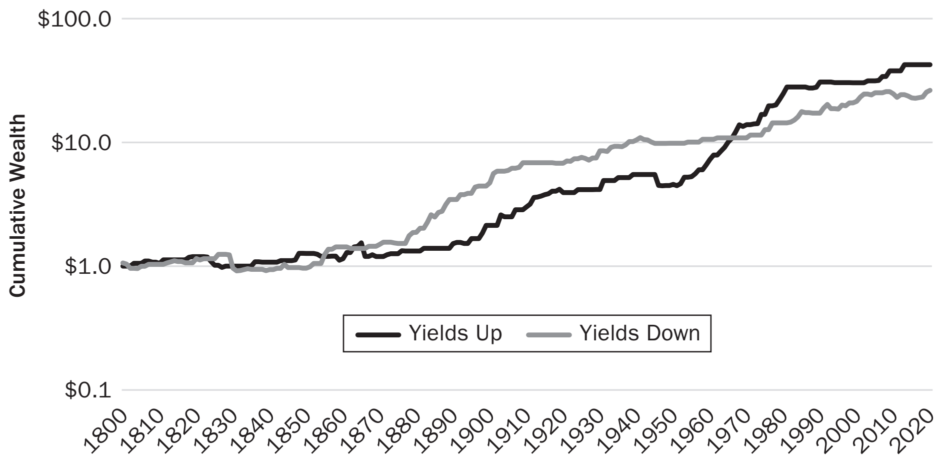
Performance Multifactor Bonds in Good and Bad States

	Growth		Crisis		Yields		Inflation		Equity	
All	Recession	Expansion	Yes	No	Up	Down	High	Low	Negative	Positive
221	55	166	75	146	103	118	74	78	44	177
3.5%	1.9%	3.9%	4.1%	3.1%	3.9%	3.0%	4.1%	4.9%	3.1%	3.5%

NOTES: We divide all calendar years into two groups per indicator: recession or expansion; crisis or no crisis; yield up or down; inflation high or low; and negative or positive equity market performance. The first row shows the number of calendar years in each category. The second row shows the performance of the multifactor bonds combination. The sample period is 1800–2020. Inflation data only start in 1869 and hence cover 152 years instead of 221 years.

EXHIBIT 10

Performance of Multifactor Bonds in Rising and Declining Yield Periods



NOTES: The exhibit shows the cumulative wealth on an initial investment of $1 at the end of 1799 in the multifactor bond portfolio that combines momentum, value, and low risk government bond factors, conditional on yield up or yield down states. Each of the 221 calendar years is labeled either yields up when the average yield of all the markets increased or yields down when the average yield declined.

THE ADDED VALUE OF BOND FACTOR PREMIUMS TO A BOND MARKET PORTFOLIO

So far, we have presented the results of long–short (zero-investment) portfolios, as is common in academic studies. Next, we examine what happens when we add the long–short multifactor bond portfolio to a long-only passive bond market portfolio that invests equally weighted in all government bond markets. Exhibit 11 shows the results. The return–risk ratio of the bond portfolio can be improved from 0.52 to 0.69 (adding 50% multifactor bonds) or 0.77 (adding 100% multifactor bonds).

The bottom two rows of Exhibit 11 splits these results for years of rising or declining yields. The bond market portfolio on average returns 0.42% in years in which yields increase, compared to 7.27% when yields decline. Adding multifactor makes calendar years with rising yields substantially more attractive for investors. The same holds for years with declining yields. Hence, a bond market investor would clearly benefit from using multifactor bonds. Exhibit 12 shows the cumulative performance of the four cases in Exhibit 11. Finally, we would like to note that we have verified that the the added value of factor premiums to a passive bond portfolios is robust to different

EXHIBIT 11

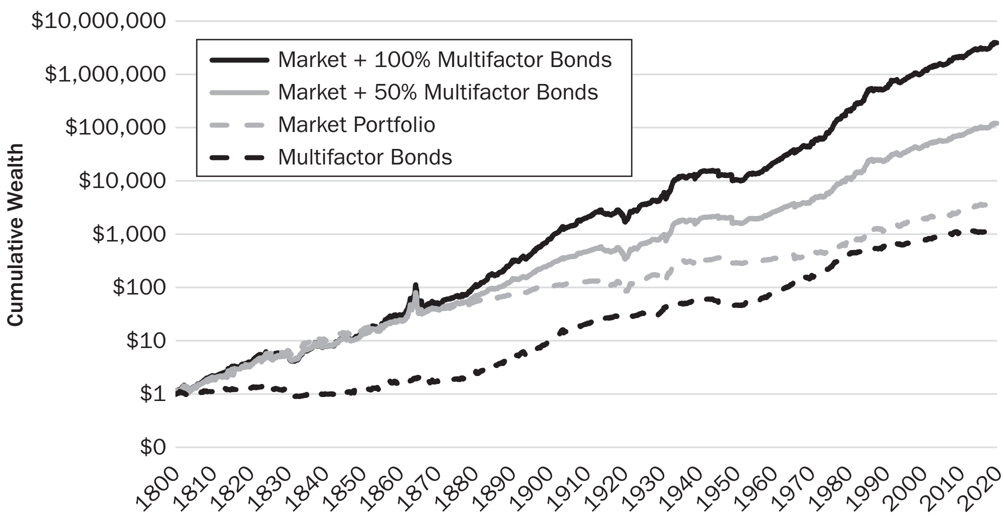
Combining a Passive Bond Market Portfolio with Factor Premiums

	Market Portfolio	Multifactor Bonds	Market + 50% Multifactor Bonds	Market + 100% Multifactor Bonds
Return p.a.	4.08%	3.48%	5.75%	7.49%
Stdev p.a.	7.88%	6.25%	8.32%	9.79%
Return-Risk Ratio	0.52	0.56	0.69	0.77
Yields Up	0.42%	3.94%	2.34%	4.31%
Yields Down	7.27%	2.98%	8.55%	10.04%

NOTES: We combine the buy-and-hold bond market portfolio that invests equally weighted in all bond markets (Market Portfolio) with the (long–short) multifactor bonds (Multifactor Bonds). Shown are the average returns, standard deviations, and return–risk ratios of the market portfolio; multifactor bonds; the bond market portfolio plus 50% multifactor bonds; and the bond market portfolio plus 100% multifactor bonds. In the last two rows, we divide the calendar years into years in which yields went up and years in which yields declined, as defined in Exhibit 9, and show average returns per scenario. The sample period is 1800–2020.

EXHIBIT 12

Cumulative Performance Long Bonds with Multifactor Overlay



NOTES: The exhibit shows how the cumulative wealth on an initial investment of $1 in various bond portfolios at the end of 1799. We combine the buy-and-hold bond market portfolio, which invests equally weighted in all government bond markets (Market Portfolio), with the (long–short) multifactor bonds (Multifactor Bonds). Shown are the cumulative performances of the bond market portfolio, multifactor bonds, the bond market portfolio plus 50% multifactor bonds, and the bond market portfolio plus 100% multifactor bonds.

portfolio construction approaches for multifactor bonds, such as tercile portfolios, two portfolios, or lower weights (up to 10%) to multifactor bonds, approaches that actively prevent short positions and enforce a long-only portfolio.

Above we have analyzed bond factors that do country allocation or curve allocation, so-called *cross-sectional strategies*. Such strategies by nature do not take a stance on the direction of the bond market; they only assess the relative attractiveness of the different bond markets or maturities. Of course, studies also have showed that bond market timing can be successful as well; see, for example, Baltussen, Martens, and Penninga (2021), who provided an out-of-sample test for timing variables introduced by Ilmanen (1995) and others. Therefore, we next examine whether cross-sectional multifactor bonds can add value to bond market timing. To this end, we take as an example one timing variable that is available from 1800, time-series momentum.

EXHIBIT 13

Combining a Passive Bond Market Portfolio with Factor Premiums and Time-Series Momentum

	Market Portfolio	Multifactor Bonds	TS Momentum	Combination
Return p.a.	4.08%	3.48%	6.48%	7.39%
Stdev p.a.	7.88%	6.25%	11.68%	9.00%
Return-Risk Ratio	0.52	0.56	0.55	0.82

NOTES: We combine (Combination) the buy-and-hold bond market portfolio, which invests equally weighted in all bond markets (Market Portfolio), with 50% of multifactor bonds (Multifactor Bonds) and 25% time-series momentum (TS Momentum). Shown are the average returns, standard deviations, and return–risk ratios of the market portfolio, multifactor bonds, time-series momentum, and their combination. The sample period is 1800–2020.

Both cross-sectional and time-series momentum are based on 12-month momentum, skipping the most recent month. Instead of building a long–short portfolio, however, time-series momentum takes a long or short position in each market based on the sign of the return, following Moskowitz, Ooi, and Pedersen (2012). Hence, potentially it is long in all bond markets or short in all bond markets.

Exhibit 13 shows the results for the bond market portfolio, multifactor bonds, and time-series momentum, as well as combing these strategies by adding 50% multifactor bonds and 25% time-series momentum to the passive bond market portfolio. We apply a lower weight for time-series momentum to account for its volatility, approximately a factor of two higher than the diversified multifactor bond portfolio (we have verified that more sophisticated techniques such as mean–variance optimizations lead to qualitatively similar conclusions). The correlation between multifactor and time-series momentum is relatively low (0.19). Hence, it is no surprise the risk–return ratio increases from 0.69—when combining the bond market portfolio with 50% multifactor bonds—to 0.82 when adding 25% time-series momentum in government bonds. Hence, both multifactor bonds and time-series momentum add value on top of each other and relative to a passive bond market portfolio.

CONCLUDING REMARKS

We extensively study factor premiums in global government bond markets over a deep sample of 221 years of data between 1800 and 2020. Existing bond factor studies typically cover the post-1980 period, which is characterized by strongly declining yields and is subject to potential p-hacking concerns. Our findings reveal that bond factors (value, momentum, and low risk) offer attractive premiums that do not decay across samples, are persistent over time, and are consistent across various market or macroeconomic scenarios. As such, we provide both deep-sample evidence and the important insight that bond factor performance is also strong in periods of rising yields. A multifactor bonds strategy that combines value, momentum, and low risk provides the strongest risk-adjusted returns and has a stable performance over time that is strong regardless of being in good or bad states, as characterized by expansions and recessions, non-crises or crises, positive or negative equity returns, or low or high inflation. Furthermore, the factor premiums diversify to each other as well as to bond or equity market risks and consistently add value over a bond market portfolio. Overall, a multifactor bond portfolio is interesting for bond investors as it offers over a passive government bond portfolio.

APPENDIX A

HISTORICAL DATABASE CONSTRUCTION

We have compiled our data from several sources to obtain a reliable and historically extensive dataset. Our sample covers 221 years of data from December 31, 1799 through December 31, 2020.

Bond Data

We source bond futures price and return data from Bloomberg and splice these with bond index-level data from Datastream, backfilled before inception with Global Financial Data (GFD). From the same sources, we obtain yields and inflation data, the latter extended where possible with data from Macrohistory.net. We apply a two-month lag to inflation numbers to mimic their real-time availability. The markets we consider are the major developed bond markets around the globe. Exhibit A1 summarizes the start dates of the data series in our sample.

Economics

We construct our global recession data from splicing the OECD G7 recession indicator from the OECD website (1960–2020), the NBER US recession indicator from the NBER website (1864–1959), and the contraction of real GDP from GFD (1800–1863). We obtain the historical data on crisis periods from Carmen Reinhart and Kenneth Rogoff, using their Banks, Currency, Default, Inflation (BCDI) index, which starts in 1800.[7]

Data Quality

The deep historical data tend to be of lesser quality compared to the more recent data because digital archives and the use of indexes with strong requirements on data processes did not exist in the past. Instead, data were maintained typically by exchanges, statistical agencies, newspapers, and investor annuals, often in manual writing. For specific issues concerning historical data for government bonds, we refer to Baltussen, Swinkels, and Van Vliet (2021) for more detail.

To construct a high-quality dataset, we build on the work of Baltussen, Swinkels, and Van Vliet (2021). They took the following three steps. First, they checked and corrected each data series for potential data errors (see "Data Cleaning Procedure"). Second, they verified the data sources, when possible, against other sources and found that average returns and volatilities are generally of comparable magnitude across databases. Third, they applied a number of conservative screens on the data series and removed data points when they did not pass these screens. These screens are (1) a zero-return screen, which leaves out data series with more than one zero or missing spot return observation in the past 12 months; (2) a return interpolation screen, which leaves out identical returns from one month to the next month; and (3) a stale return screen, which leaves out observations that do not have nine or more differentiating returns over the past 12 months. The first screen filters for data historically available at a non-monthly frequency and for reduced liquidity. The second screen filters the unlikely return pattern of exactly identical consecutive monthly returns, which indicates return interpolation. The third screen filters returns not updated at a monthly frequency. To this end, they removed an asset when, over the past 12 months, fewer than nine unique monthly returns were rounded to 5 bps. They found that such a pattern is unlikely under a normal distribution or in the replication-sample return distribution for the government bond markets in our universe. Furthermore, following

[7] http://www.reinhartandrogoff.com/data/.

EXHIBIT A1
Bond Market Data Sample

Bond Market	GFD	BB Futures	DS
Australia	1857–1986	1987–2020	1987–2020
Belgium	1831–2020		
Canada	1953–1984	1989–2020	1985–2020
Denmark	1800–2020		
Germany	1800–1979	1990–2020	1980–2020
France	1800–1984	2012–2020	1985–2020
Italy	1807–1990	2009–2020	1991–2020
Japan	1870–1981	1985–2020	1982–2020
Netherlands	1800–1979		1980–2020
Norway	1822–1991		1991–2020
Spain	1800–1989		1990–2020
Sweden	1853–1986		1987–2020
Switzerland	1900–1979		1980–2020
United Kingdom	1800–1979	1982–2020	1980–2020
United States	1800–1979	1982–2020	1980–2020
New Zealand	1861–1990		1991–2020

NOTES: The GFD and DS data are 10-year bond returns and bond yields (DS: 7–10-year maturity bucket). From BB, we get 10-year bond futures returns.

BB = Bloomberg; DS = Refinitiv Datastream; GFD = Global Financial Data.

Baltussen, Swinkels, and Van Vliet (2021), we skip a month between the momentum signals and investing, which removes possible spurious autocorrelation at the monthly frequency. We would like to stress that these screens mitigate data quality concerns but could bias factor premium estimates downward if they remove correct data points.

Data Cleaning Procedure

We have taken the following steps to check the quality of each data series and clean for obvious measurement errors. First, we run a studentized residuals outlier test on each series to capture potential data errors and visually check each series for jumps and outliers. Potential outliers are manually verified (by comparison to other data sources where possible and by searching for reasons for large price moves) and, when due to a data error, are corrected. The corrections in pre-sample data points include the following:

- Dutch government bond returns have a misprint on November 30, 1964, which we replaced with the approximation based on changes in yields times duration.
- Several systematic corrections were applied for financing rates and bond yields because some yields switch between decimal and percentage notations.

APPENDIX B

PORTFOLIO CONSTRUCTION PROCEDURE

After obtaining the factor measures per the bond market, we construct factor investment portfolios at the end of every month in the following manner. We rank the markets based on the factor measure and take a position equal to the rank minus its cross-sectional average (requiring a minimum of two markets to be present). This procedure is similar to that used by Asness, Moskowitz, and Pedersen (2013), Frazzini and

Pedersen (2014), and Koijen et al. (2018). Consequently, positions for all factors add up to zero at each point in time:

$$w_t^i = z_t \cdot \left(Rank\,(S_t^i) - \frac{N_t + 1}{2} \right)$$

with w_t^i the weight of asset *i* at time *t*, S_t^i the factor signal, N_t the number of assets in the cross-section, and z_t a scaling factor to ensure that the portfolio weights sums to zero.

Next, we size positions in each market by their simple three-year rolling volatility estimate or beta estimate (for BAB only), in the same spirit as Asness, Moskowitz, and Pedersen (2013) and Frazzini and Pedersen (2014) but fitted to our sample frequency (i.e., monthly data). To prevent undue impact from extremely low-volatility estimates (and hence keep the factor strategy robust from an investor perspective), especially in the earlier part of our sample, we floor each volatility (beta) estimate at the maximum of the 10% quantile of volatility (beta) estimates or 2.5% (0.25), whichever is greater.

We subsequently sum the product of position, sizes, and market returns across markets for each date to generate the return on the factor strategy. We then adjust the position sizes of each factor strategy using a 10-year rolling window such that each factor strategy targets an ex ante volatility of 10% per annum. This approach takes an ex ante view of portfolio construction, as available in real time. However, our results are not materially different if we simply scale by the sample ex post volatility, as done by Koijen et al. (2018).

We rebalance the portfolios each month based on the signals and volatility estimates. This methodology results in balanced long–short portfolios.

ACKNOWLEDGMENTS

We would like to thank David Blitz, Patrick Houweling, Laurens Swinkels, Pim Van Vliet, and Casper Zomerdijk for valuable contributions and discussions.

REFERENCES

Asness, C., T. Moskowitz, and L. Pedersen. 2013. “Value and Momentum Everywhere.” *The Journal of Finance* 68 (3): 929–985.

Baltussen, G., M. Martens, and O. Penninga. 2021. “Predicting Bond Returns: 70 Years of International Evidence.” *Financial Analysts Journal* 77 (3): 133–155.

Baltussen, G., L. A. P. Swinkels, and P. Van Vliet. 2021. “Global Factor Premiums.” *Journal of Financial Economics* 142 (3): 1128–1154.

Begley, C. G., and L. M. Ellis. 2012. “Raise Standards for Preclinical Cancer Research.” *Nature* 483: 531–533.

Blitz, D. 2012. “Strategic Allocation to Premiums in the Equity Market.” *The Journal of Index Investing* 2 (4): 42–49.

Blitz, D., G. Baltussen, and P. Van Vliet. 2020. “When Equity Factors Drop Their Shorts.” *Financial Analysts Journal* 76 (4): 73–99.

Brooks, J., D. Palhares, and S. Richardson. 2018. “Style Investing in Fixed Income.” *The Journal of Portfolio Management* 44 (4): 127–139.

Camerer, C. F., A. Dreber, E. Forsell, T. H. Ho, J. Huber, M. Johannesson, and E. Heikensten. 2016. “Evaluating Replicability of Laboratory Experiments in Economics.” *Science* 351 (6280): 1433–1436.

Dimson, E., P. Marsh, and M. Staunton. "Global Investment Returns Yearbook 2021." Credit Suisse Research Institute. 2021.

Doeswijk, R., T. Lam, and L. A. P. Swinkels. 2020. "Historical Returns of the Market Portfolio." *Review of Asset Pricing Studies* 10: 521–567.

Fama, E. F., and K. R. French. 1992. "The Cross-Section of Expected Stock Returns." *The Journal of Finance* 47 (2): 427–465.

——. 2015. "A Five-Factor Asset Pricing Model." *Journal of Financial Economics* 116 (1): 1–22.

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

Harvey, C. 2017. "Presidential Address: The Scientific Outlook in Financial Economics." *The Journal of Finance* 72 (4): 1399–1440.

Harvey, C. R., Y. Liu, and H. Zhu. 2016. "… and the Cross-Section of Expected Returns." *The Review of Financial Studies* 29 (1): 5–68.

Houweling, P., and J. Van Zundert. 2017. "Factor Investing in the Corporate Bond Market." *Financial Analysts Journal* 73 (2): 100–115.

Ilmanen, A. 1995. "Time-Varying Expected Returns in International Bond Markets." *The Journal of Finance* 50 (2): 481–506.

Koijen, R., T. Moskowitz, L. Pedersen, and E. Vrugt. 2018. "Carry." *Journal of Financial Economics* 127 (2): 197–225.

Kothe, J., H. Lohre, and C. Rother. 2021. "Rates Factors and Global Asset Allocation." *The Journal of Fixed Income* 30 (3): 6–25.

Moskowitz, T. J., Y. H. Ooi, and L. H. Pedersen. 2012. "Time-Series Momentum." *Journal of Financial Economics* 103: 228–250.

Sharpe, W. F. 1994. "The Sharpe Ratio." *The Journal of Portfolio Management* 21 (1): 49–58.

Disclaimer

The views expressed in this article are not necessarily shared by Robeco Institutional Asset Management. We welcome comments, including references to related papers we have inadvertently overlooked.