Evolving Practices in Public Investment Management

Proceedings of the Eighth Public Investors Conference

Papers in this volume were prepared for a meeting of officials from central banks and official institutions held at the headquarters of the Bank of Canada on 27-28 October 2022. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the World Bank or the central banks represented at the meeting. Individual papers (or excerpts thereof) may be reproduced or translated with the authorisation of the authors concerned.

This publication is available on the Public Investors Conference website
(www.publicinvestorsconference.org).

ISSN 2708-0358 (online)
ISBN 978-92-9259-682-8 (online)

Contents

Preface

The aftermath of the Covid-19 pandemic, geopolitical tensions and the resurgence of inflation have been primary drivers of the economic environment and financial markets. In most countries, the extraordinary monetary policy measures that have shaped financial markets over the past 15 years are being phased out. The normalisation of monetary policy has brought new challenges, as certain parts of the financial sector and the real economy appear vulnerable to the rapid increases in yields and the decompression of risk premia. Adding to these unprecedented challenges are financial innovations, technological developments – such as artificial intelligence – and the urgent need for public investors to support the transition to a sustainable economy.

Given this environment, public investors are faced with several "known unknowns", compelling them to re-evaluate their governance and risk management processes.

To foster the debate about these challenges, the Bank for International Settlements and the World Bank – jointly with the Bank of Canada and the Bank of Italy – organised the Eighth Public Investors Conference in late 2022. The conference, originally planned in 2020 and postponed due to the pandemic, was held at the headquarters of the Bank of Canada in Ottawa on 27–28 October 2022 and gathered participants from 35 public and private institutions belonging to 26 different countries. The conference aimed to promote the exchange of innovative ideas among practitioners and academics, knowledge-sharing and collaboration across organisations, and the development and dissemination of best practices in the management of financial resources by the public sector.

The proceedings published in this book cover most of the papers presented at the event, including some of the most current research and ideas on public asset management. This publication aims to provide valuable insights for asset owners, academics and researchers, regulatory and oversight bodies, investment consultants and, notably, central bankers.

The book is organised into three parts, each reflecting the major topics debated:

- Part 1 focuses on public investment management, taking into account a number of issues that currently shape the investment process of central banks and other public institutions. The first chapter investigates the impact of governance arrangements on investment policies for central bank foreign exchange reserve portfolios. The second chapter presents a balance sheet model aimed at providing central bankers with a comprehensive framework for asset allocation. The model allows for the comparison of financial risks with existing financial buffers, considering different scenarios, geopolitical risks and economic uncertainty. The third chapter introduces the potential trade-off between sustainability and financial returns by analysing the historical performance of green bonds.
- As climate risk and sustainability are becoming key factors in the investment process of public investors, Part 2 is dedicated entirely to such issues. The first chapter examines abnormal flows into green exchange-traded funds as a possible indicator of market-wide changes in investor appetite for environmental responsibility. The second chapter

investigates the existence of a carbon risk premium in the corporate bond market, driven by investor preferences for environmentally responsible firms and the perceived credit risks associated with more carbon-intensive firms. The third chapter expands on the Network for Greening the Financial System climate scenario framework to estimate expected returns under climate scenarios for fixed income portfolios. The fourth chapter incorporates sustainability considerations into sovereign and corporate bond investments.

- Finally, Part 3 addresses topics related to quantitative portfolio management. The first chapter constructs and analyses active systematic investment strategies for sovereign fixed income investors based on carry and term premia. The second chapter introduces a novel statistical procedure for backtesting well known risk measures such as value-at-risk and expected shortfall. The third chapter applies machine learning techniques to the active management of bond portfolios. The fourth chapter emphasises the importance of the information contained in the tails of return distributions for portfolio management purposes.

This book would not have been possible without the contributions, first and foremost, of the presenters at the Eighth Public Investors Conference. The editors are grateful for their permission to publish their original work in this volume. We also wish to acknowledge the hospitality of the Bank of Canada and the funding provided by all sponsoring institutions to make the conference possible. We extend our thanks to the numerous participants, reviewers and staff from multiple institutions whose insightful comments and efforts contributed greatly to the preparation of this book, including Nicola Faessler and Margarita Sanchez.

Ottawa, Canada	Nicolas Audet
Washington, DC	Eric Bouyé
Basel, Switzerland	Ingo Fender
Rome, Italy	Patrizia Ferrauto
Washington, DC	Carmen Herrero Montes
Ottawa, Canada	Etienne Lessard
Rome, Italy	Simone Letta
Basel, Switzerland	Mike McMorrow
Rome, Italy	Tommaso Perez
Washington, DC	Daniel Vela Barón
Basel, Switzerland	Omar Zulaica

Part 1

Public investment management

Central bank governance and reserve portfolio investment policies: an empirical analysis[1]

Daniela Klingebiel,[2] Carmen M Herrero Montes,[3] Marco Ruiz[4] and James Seward[5]

Abstract

This paper uses a unique survey data set of 105 central banks to investigate whether investment policies for central bank foreign reserve portfolios are linked to the governance arrangements for reserve management. The paper evaluates whether a central bank's investment decision-making structure impacts how much risk institutions take in their reserve management operations and the level of diversity in their reserve portfolios. Additionally, it explores the implications of the broader governance environment on reserve management. The analysis yields four key findings. First, internal governance arrangements matter for foreign reserve portfolio investment policy; the empirical results indicate that reserve portfolios are more diversified in central banks where the middle office reports directly to the board. Second, controlling for the level of reserves, the macro environment and the broader governance environment, reserve portfolios are more diversified in central banks where the back, middle and front offices are separated. Third, the regression analysis also reveals that central banks in countries where the ministry of finance has an obligation to cover negative equity have fewer eligible currencies and are, therefore, less diversified. Fourth, central banks where boards actively exercise portfolio oversight usually have portfolios with more risk and diversification. Portfolios with longer investment horizons, more currencies and a broader set of asset classes have performed better historically while limiting downside risk. Given that the analysis controls for the broader governance environment, the data indicate that any central bank can improve its internal governance regardless of the external governance environment. This paper contributes to the literature on central bank foreign reserve management and on understanding the importance of governance arrangements in investment policy.

JEL classification(s): E5, G11, D81.

1 The findings, interpretations and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organisations or those of the Executive Directors of the World Bank or the governments they represent. We thank Mikaela Ballon Carneiro for her excellent support in the literature review and data analysis sections. An initial draft of this paper benefited from written comments by Eric Bouye, Pierre Cardon, Stijn Claessens, Erik Feyen, Bjoern Geir, Renee Licorish, Mike McMorrow, Steen Byskov and Shinya Tamada. An earlier version of the paper also benefited from comments and suggestions from colleagues of the World Bank Treasury's Asset Management and Advisory Department.

2 Former Reserve Advisory and Management Partnership (RAMP) Director, Asset Management and Advisory, World Bank Treasury, dmklingebiel@gmail.com.

3 Financial Officer, Asset Management and Advisory, World Bank Treasury, cherreromontes@worldbank.org.

4 Senior Financial Officer, Asset Management and Advisory, World Bank Treasury, mruizgil@worldbank.org.

5 Senior Financial Officer, Capital Markets and Investments, World Bank Treasury, jseward@worldbank.org.

1. Introduction

Governance is crucial for central banks. According to Bossu and Rossi (2019), the concept of central bank governance is multifaceted. They contend that a clearly defined central bank mandate and decision-making structure are essential to support accountability and legitimacy. Such a mandate should elaborate on objectives, functions and powers, specifying the legal tools available for a central bank to implement its functions. Clear decision-making structures are also necessary to ensure adequate implementation of the mandate. Given the complexity of central banks and their multiple responsibilities, it is crucial to define what to decide, who must decide, and how decisions must be made (Bossu and Rossi (2019)).

Our empirical analysis focuses on the governance of reserve management, a crucial part of central bank governance overall. As foreign currency reserves have significantly increased and represent an essential part of central bank assets, the process for making investment decisions and overseeing results has become more relevant. We aim to contribute to the reserve management governance discussion with data-driven analysis using a unique survey data set on central bank reserve management practices. Most previous publications on the topic are prescriptive and qualitative. Notably, our research shows that governance arrangements do matter for the foreign reserve portfolio investment policy.

As the first step in this analysis, we briefly review the literature on central bank governance, focusing on reserve management operations. We then describe the data that we use for empirical research. We use a unique data set collected by the World Bank's Reserve Advisory and Management Partnership (RAMP) using two surveys conducted across 99 central banks in 2018 and 105 central banks in 2019.[6] The surveys collected data on central banks' reserve management activities, including their governance structures, components of their investment policies and asset allocations, and their accounting and profit-sharing methodologies. We also deploy data that describe the broader institutional and macro environment in which these central banks operate and proxies for reserve adequacies. We use this data set to address novel research questions to find links between specific governance arrangements and reserve management investment policies. Finally, we conclude and discuss the possible implications of our key findings for the management of reserves.

2. Literature review

Central banks have various roles and are responsible for a broad list of functions, which have grown significantly over time. While central banks vary substantially in structure and purpose globally, most institutions have critical responsibility for monetary policy, financial system stability and safeguarding the financial infrastructure's core elements. Managing foreign reserves is typically another essential central bank function (Anasashvili et al (2020)). This section briefly summarises the literature on central bank legal frameworks, functions and governance arrangements, focusing on reserve management.

6 We considered the set of respondents that participated in both surveys, a total of 93 central banks. The set of respondents common to both surveys varies depending on the variable.

2.1 Legal framework

The legal frameworks of central banks differ widely from country to country. For example, most common-law countries do not have constitutional provisions for the central bank. Countries with these provisions often differ in the rules and authorities concerning central bank functions and mandates that are built into their constitutions (Ortiz (2009)). Most central bank laws, however, explicitly define independence, prescribe the central bank's policy goals, and provide discretion and autonomy to achieve those goals through policies and operations (Khan (2017)).[7] For many central banks, price stability, monetary policy, financial supervision and reserve management are part of a legislative mandate (Appendix III briefly reviews the relationships between financial supervision functions and the broader governance environment and reserve management). In other cases, central banks' objectives are implicit in more general economic goals. For example, price stability is critical for achieving stable economic growth. Following the 2008–09 global financial crisis, the objectives and powers delegated to central banks increased significantly to allow for unconventional responses to halt a worldwide financial meltdown (Balls et al (2018)). Building on this expansion during the global financial crisis, central bank actions in response to the Covid-19 pandemic have been unprecedented in speed, scope and scale.[8] Observers of central banks have raised questions about the rapid expansion of new instruments and unconventional monetary policies and the implications of these for central bank independence (de Haan and Eijffinger (2016)).

2.2 Reserve management function

Holding and managing a country's official foreign reserves are among a central bank's core functions (Bossu and Rossi (2019)). Central banks' reserve management operations have always been part and parcel of their monetary policy function. Initially, most central banks had gold-backed currencies, and their reserves primarily consisted of gold bullion. With the abolition of the gold standard in 1971, central banks started diversifying their reserves into foreign asset pools.

Holding foreign reserves has various purposes. Most notable are self-insurance for balance of payments crises and support of exchange rate policies. In some countries, reserves also support financial stability by providing lender of last resort functions in foreign currency. As a result, reserves protect the country against disruptions and volatility in international capital markets. Over the past two decades, successive financial crises have solidified and validated this approach. Recent research indicates that emerging market central banks holding relatively higher foreign reserves have experienced less currency depreciation, smoother credit growth and more stable credit ratings. Countries with higher levels of reserves also had better access to external funding during the global financial crisis of 2008 (Arslan and Cantú (2019)). Furthermore, empirical research identifies a relationship between political instability and weak institutions in emerging markets. This results in central banks accumulating international reserves as a risk mitigation method to deal with broader economic uncertainties (Ortega de la Rosa (2015)).

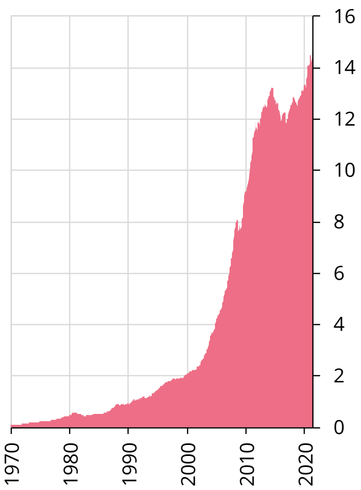
[7] For a compendium of such legislation globally, see Khan (2017), which provides an analysis.

[8] See IMF (2020) for a global accounting of the monetary and fiscal policy actions taken by governments worldwide.

Foreign reserves have grown significantly over the past four decades and are now at record highs (Figure 1a). Emerging market economies, notably China, have led this trend. Economic development and higher commodity prices have encouraged capital inflows and higher exports in emerging market economies. As a result, central banks have intervened in foreign exchange markets and accumulated foreign reserves to curb the appreciation of their exchange rates. As Figure 1 shows, this build-up has been a worldwide phenomenon. On average, emerging market central banks[9] have accumulated foreign reserve levels at or above 23% of GDP (Figure 1b).[10] Many countries hold reserves far above traditional benchmarks, such as three months of import cover, 20% of broad money, or 100% cover of short-term external debt repayments. In addition, according to the Assessing Reserve Adequacy (ARA) metric[11] of the International Monetary Fund (IMF), approximately half of the reported countries have reserves above the adequate level.

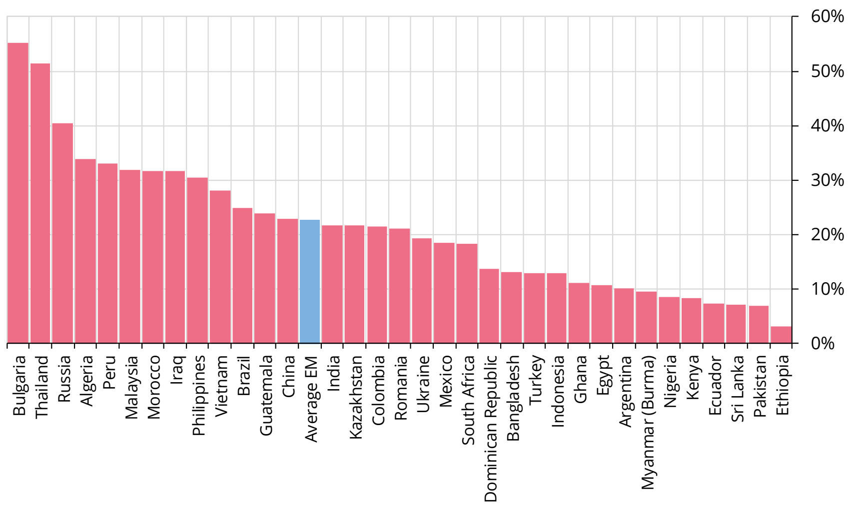
Figure 1

1a Volume of world reserves
In USD trillions



Last observation as of June 2021.
Source: IMF, *International Financial Statistics*.

1b Reserves relative to GDP – largest emerging markets
As a percentage of GDP



Annual data as of 2020.
Sources: IMF, *International Financial Statistics*, World Bank, *World Development Indicators*.

The build-up of reserves has transformed the way central banks manage these funds. Although they remain invested primarily in government bonds and other conservative instruments, adopting non-traditional asset classes, such as mortgage-backed securities, corporate bonds and equities, is on the rise (Anasashvili et al (2020)).

9 This conclusion draws on the World Bank current country classification. Emerging markets are countries classified as being those with low, lower-middle and upper-middle incomes.

10 This figure is based on available data for emerging markets for 2020 and refers to the GDP-weighted average of official reserve assets.

11 The IMF publishes the ARA metric for 65 countries. This metric combines some traditional reserve adequacy variables, like broad money, short-term debt and imports, to provide a more robust estimate of reserve adequacy. The data and methodology are available at www.imf.org/external/datamapper/Reserves_ARA@ARA/CHN/IND/BRA/RUS/ZAFtt.

2.3 Reserve management governance

Reserve management governance refers to the institutional arrangements and processes for policy development and investment of foreign exchange assets. According to de Abreau Faria and Streit (2016), an effective governance framework ensures clear delegation and separation of responsibilities. Governance arrangements typically establish the policymaking structure; in other words, they determine who is responsible for each type of decision. In addition, these arrangements include reporting lines and oversight mechanisms. Figure 2 depicts a stylised structure of governance arrangements.

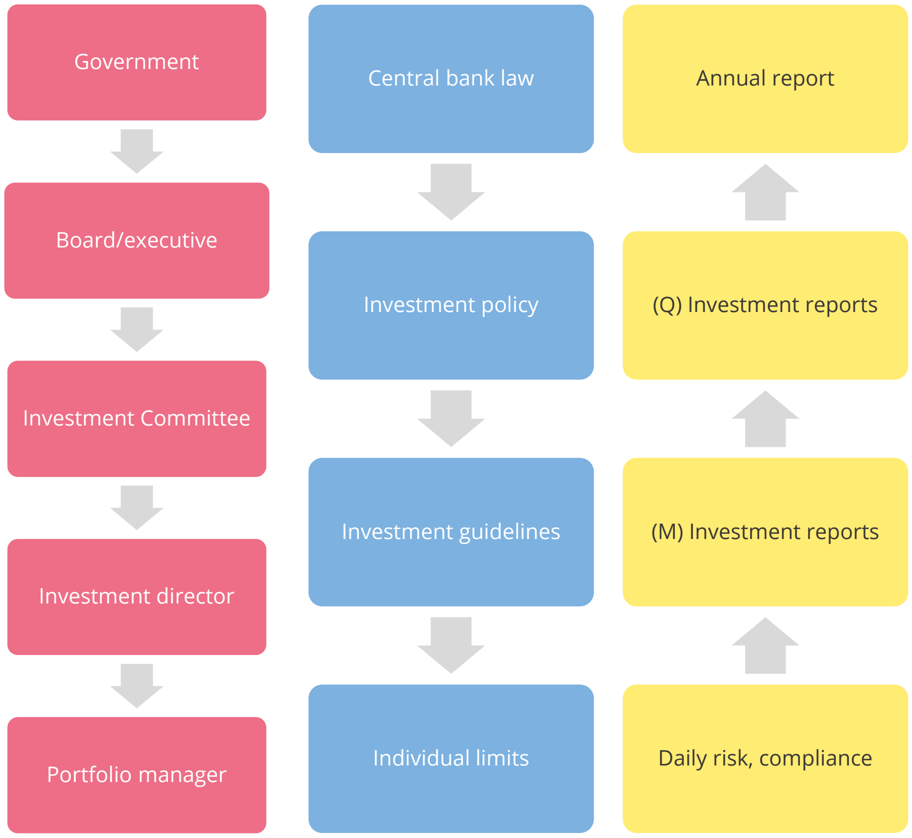
Reserve management governance reflects country-specific institutional, social and regulatory considerations like other central bank arrangements. Central bank laws are generally not specific and prescriptive about the reserve management function. This allows flexibility to define and adjust governance as markets evolve, asset classes change and portfolio practices modernise over time.

Despite the importance of country-specific factors, there is agreement on the most relevant principles for reserve management. The IMF Guidelines for Foreign Exchange Reserve Management are an excellent example of generally agreed reserve management principles. The IMF developed this document in consultation with various central banks and international organisations, including the World Bank. According to the Guidelines, the "internal governance structure of the reserve management entity should be guided by and reflect the principles of clear allocation and separation of responsibilities. Sound management of internal operations and risks requires appropriately qualified and well-trained staff, following sound business practices" ((Al-Hassan et al (2014)). The guidelines further suggest that central bank boards should make decisions at a strategic level and delegate decisions concerning strategy implementation to the investment committees. The operational units implement the decisions made by the board and investment committee. Such a division of responsibilities results in a three-tier decision-making structure for reserve management operations.

A clear definition of roles and policies is considered critical. The investment policy statement for reserve management is the most suitable place to define eligible asset classes, investment instruments and transactions (Johnson-Calari and Strauss-Khan (2020)). The board's involvement in the investment policy definition is deemed the most effective structure to ensure robust oversight and informed decision-making. Furthermore, the role of every person who participates in the investment process must be defined clearly, along with well defined documentation of processes, to allow continuous decision-making (Ruiz (2020)).

The RAMP survey (see Anasashvili et al (2020)) collects data on how central banks organise their reserve management operations. The survey findings indicate that reserve governance arrangements vary across central banks. Most of the 105 respondents to the 2019 RAMP survey follow a three-tier governance structure. Ninety-two per cent of central banks reported that their respective boards approved the reserve management policy, including high-level decisions such as reserve management objectives, risk tolerance, investment horizon and strategic asset allocation. Moreover, the survey also showed that central banks' boards frequently approve the investment management guidelines; that is, the specific investment rules for managing the portfolio, indicating that many central bank investment committees have limited decision-making power. Many boards are also responsible for hiring

Governance arrangements Figure 2



Source: Johnson-Calari and Strauss-Khan (2020).

external managers. Finally, the survey showed that middle office reporting on risk and return information to the board varies across central banks.

Apart from the specific governance arrangements for reserve management operations, the RAMP survey investigates how central banks organise their operational units. Typically, central banks separate the reserve management function into three different operational units. The first is the front office, which plans and executes trades. The second is the middle office, responsible for measuring risks and producing reports. The third oversees the trade settlement and accounting of reserve operations. Even though these operational units have differentiated roles and responsibilities, the location of these individual units in the organisational structure varies, as the RAMP survey shows. Approximately a third of central banks have the front, middle and back office in the same department, while another third of institutions place them in completely separate departments. The remaining central banks opt for a hybrid approach (Anasashvili et al (2020)).

3. Approach and objectives

Little quantitative research is available on the link between overall central bank governance and reserve management, despite the critical role that sound reserve management plays in helping, supporting and maintaining confidence in monetary management (Al-Hassan et al (2014)). As shown above, most publications on reserve management governance are prescriptive and qualitative. Therefore, we contribute to the reserve management governance discussion with data-driven analysis using RAMP's unique survey data on governance and organisational arrangements of central banks and asset allocation and risk measures. Specifically, we investigate whether governance arrangements impact investment policies and central bank risk-taking and, if so, precisely which arrangements matter. We also analyse whether organisational arrangements impact central banks' investment policies and whether reporting structures influence their investment policies and risk-taking in their reserve management operations. This paper empirically analyses the relationships between a central bank's governance structure for its reserve management operations and investment policies and risk-taking.

4. Methodology and data

We use correlations and regression analysis to find links between specific governance arrangements and variables related to central banks' reserve investment policies. We start by analysing correlations between the above-mentioned variables and testing for their statistical significance. We then use regression analysis to explore whether some correlation results hold when adjusting for reserve adequacy and indicators that describe the macro environment.

We use data groups that capture the governance arrangement, investment policies and measures of risk-taking of individual central banks to test whether governance affects investment policies. First, we use data on governance arrangements for central banks' reserve management operations as an independent variable collected by combining the first (2018) and the second (2019) RAMP surveys. We use data for the central banks participating in both surveys for the relevant variables. Second, we utilise the most recent data from the second RAMP survey on the composition and risk of reserve portfolios as dependent variables describing a central bank's investment policy. We then deploy three types of control variables to empirically isolate the impact of the governance structure and investment policies.

4.1 Governance and macroeconomic variables

We compiled governance variables from multiple sources, collected at the national level, to assess the broader governance environment as a control variable to isolate the effect of the governance arrangements at the central bank level. We also use macroeconomic variables and data on reserve adequacy to isolate the effects further (see Table 1 for a summary).

Governance environment, macroeconomic and reserve adequacy variables					Table 1
Variable	Obs	Mean	Std dev	Min	Max
Governance environment variables					
Government effectiveness	204	0.0	1.0	–2.4	2.2
Central bank independence	144	2.8	1.2	0	4
Corruption Perceptions Index	179	43.3	18.9	9	87
Governance Pillar Score	143	5.7	2.1	1.1	9.2
Macroeconomic variables					
Country risk score	121	46.6	16.3	13	89
Number of currency crises (1971–2017)	164	1.5	1.5	0	7
Current account balance (% of GDP)	212	–3.0	9.0	–34	34.8
Reserve adequacy variables					
Reserves to GDP (%)	162	0.2	0.2	0.6	2.0
Reserves to M2 (%)	126	0.4	0.3	0.0	1.9

Sources: IMF, *World Economic Indicators*, *Official Foreign Exchange Reserves* (COFER); World Bank; French Treasury; Transparency International; Bloomberg; Economist Intelligence Unit (EIU); MSCI.

1. **Governance environment variables**: We use various indices to measure the broader governance environment in which a central bank operates. First, we use the government effectiveness index, compiled by the World Bank. This index measures the "perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies".[12] The scale runs from –2.5 (lowest) to 2.5 (highest) and covers 204 countries. The index is developed by compiling and summarising information from over 30 existing data sources that report the views and experiences of citizens, entrepreneurs and experts in the public, private, and non-governmental sectors. The rationale for using this variable is that countries with strong governance and effective public sectors usually have well implemented laws for their central banks. We also use data on central bank independence from the French Treasury. This variable measures the degree of central bank independence in each country. Central banks with no independence are assigned a value of zero, while those with substantial autonomy are assigned a value of 4.[13] The Directorate General of the French Treasury surveys this and other institutional characteristics for 144 countries. The survey is based on perception data. The study's results are captured in the Institutional Profiles Database and were last updated in 2016. In addition, clear accountability frameworks and low corruption increase public confidence in the central bank. To measure these elements, we use the Corruption Perceptions Index, which scores 180 countries based on experts' and business executives' perceptions of corruption in a country's public sector. It is a composite index combining 13 surveys and assessments of corruption collected by various institutions. The index has a scale of 0–100, where zero represents the highest level of perceived

[12] See https://info.worldbank.org/governance/wgi/Home/Documents.

[13] See www.cepii.fr/institutions/EN/ipd.asp.

corruption, and 100 the lowest level.[14] Finally, we use the Governance Pillar Score (MSCI). This index covers 198 countries to "assess the extent to which a country's long-term competitiveness is affected by its institutional capacity to support long-term stability and functioning of its financial, judicial, and political systems, and capacity to address the environmental and social risks".[15] The Governance Pillar Score is included in the MSCI assessment of sovereign countries' environment, social and governance (ESG) risk. The best score is zero; the worst is 10. The index is derived from third-party sources and focuses on political rights and civil liberties, stability and peace, control of corruption and public financial management.

2. **Macroeconomic variables**: We then use the country risk score from the Economist Intelligence Unit's Country Risk Model to control for the macro environment. This score rates country risk from zero (no risk) to 100 (maximum risk), taking a simple average of a country's sovereign, currency and banking sector risk scores.[16] We included this variable in the analysis to assess the macro environment in which central banks operate. Alternatively, we use the number of currency crises and the current account balance as a percentage of GDP to describe the macro environment in which the central bank operates.

3. **Reserve adequacy variables**: Finally, a central bank's risk tolerance and investment policy (number of eligible assets and currencies, portfolio duration, allocation to non-traditional asset classes) are linked to reserve adequacy. All else being equal, a central bank with higher reserve adequacy typically has a higher risk tolerance and can deploy a broader range of asset classes and currencies and longer duration. Therefore, we use reserve adequacy measures to control for differences in reserve adequacy across central banks. We include two reserve adequacy metrics as control variables: reserves to GDP and M2. Separately, we also control for GDP per capita and other macroeconomic variables, such as short-term external debt stocks as a percentage of total reserves.

4.2 RAMP survey variables

RAMP's global surveys on reserve management practices are the source of information on central banks' governance arrangements, investment policies and risk-taking measures. RAMP conducted two surveys in 2018 and 2019 with responses from 99 and 105 central banks, respectively.[17] The RAMP survey data are unique, systematically collecting data on governance arrangements of reserve management investment operations across central banks globally for the first time. In addition to questions on governance arrangements and investment policies, the surveys included questions on asset allocation, portfolio management, risk management, performance and risk reporting, and transparency. Table 2 lists the survey variables we used to test our questions empirically.

14 See www.transparency.org/en/cpi and https://images.transparencycdn.org/images/2019_CPI_SourceDescription_EN-converted-merged.pdf.

15 See www.msci.com/documents/10199/e092c439-34e1-4055-8491-86fb0799c38f.

16 See www.eiu.com/n/solutions/country-risk-model/.

17 Anasashvili et al (2020); see https://openknowledge.worldbank.org/bitstream/handle/10986/33657/K880541.pdf?sequence=1&isAllowed=y.

Summary statistics of RAMP survey variables Table 2

Variable	Obs	Mean	Std dev	Min	Max
Governance – discrete variables					
Independent investment committee	105	0.2	0.4	0	1
Middle office reports to the board	105	0.9	0.3	0	1
Middle office reports to the investment committee	105	0.8	0.4	0	1
Back office, middle office and front office are in the same department	105	0.3	0.5	0	1
Back office, middle office and front office are in separate departments	105	0.3	0.5	0	1
Obligation to cover negative equity	93	0.6	0.5	0	1
Governance – continuous variables					
Transparency of reserve management policies	95	0.4	0.3	0	1
Composition of reserve portfolio – continuous variables					
Number of eligible currencies	100	7.9	4.3	1	20
Number of eligible assets	97	6.8	2.8	1	14
Allocation to non-traditional assets (%)	71	10.1	17.0	0	72
Estimated risk of the portfolio	71	1.0	0.3	0.8	2.9
Investment horizon – total portfolio (months)	57	35.3	30.7	1	126
Duration – total tranched portfolio (months)	52	22.4	18.9	1	84
Duration – liquidity tranche (months)	64	7.8	12.6	1	76
Duration – investment tranche (months)	62	32.1	31.5	3	180

The survey specifically asked whether the middle office reports to the board on performance and risk metrics and compliance with the investment management guidelines.

Source: RAMP Survey on the Reserve Management Practices of Central Banks (World Bank Treasury).

Concerning governance arrangements, we consider as independent variables those indicated in the RAMP survey results as differing across central banks, which could therefore influence central bank investment policies and risk-taking. We use dummy variables for the following governance arrangements:

1. We consider an investment committee independent if it approves the investment management guidelines; that is, the specific investment rules for managing the portfolio.
2. We explore whether the middle office reports directly to the board.
3. We also consider whether the middle office reports performance and risk metrics to the investment committee.
4. We test whether a central bank organises its reserve management operations in one department.
5. We further assess whether the reserve management operations are arranged in separate departments.
6. Finally, we probe the impact of the ministry of finance's obligation to cover the central bank's negative equity.

Relatedly, we further probe transparency on the risks and returns of reserve management operations to the board as a continuous variable. The variable represents the central banks' degree of transparency, ranging from zero to one, based on the public disclosure of their currency composition, investment guidelines and benchmark selection.

We use RAMP survey data collected as continuous data on the number of eligible currencies and asset classes and the allocation to non-traditional asset classes as dependent variables to reflect a central bank's investment policies. We also deploy RAMP data to compute the measures of risk as reflected in a central bank's current asset allocation: (1) estimated risk of the portfolio;[18] (2) investment horizon (measured at the total portfolio level for both tranched and untranched portfolios); and (3) duration (measured at the total portfolio level and the liquidity or investment tranche level).

5. Descriptive statistical results

We run simple correlations for the various independent governance variables to identify patterns in the data. The correlations are run to develop context and show empirical relationships between critical elements of reserve management. The full results of the correlation analysis are found in Appendix I.

1. **Independent investment committee and investment policies**. We test whether a positive statistical correlation exists between the investment committee's independence and the investment guidelines' characteristics. As explained above, we define an investment committee as independent if it is responsible for approving the investment guidelines. Table 3 indicates that an independent investment committee negatively correlates with the number of eligible currencies and asset classes. Similarly, duration is negatively correlated with an independent investment committee.

Independence of the investment committee and investment policies — Table 3

Variables	Estimated risk of the portfolio	Non-traditional assets (%)	Number of eligible currencies	Number of eligible assets	Investment horizon – total portfolio (months)	Duration – total portfolio (months)	Duration – liquidity tranche (months)	Duration – investment tranche (months)
Independent investment committee	–0.179	–0.233*	–0.102	–0.148	–0.036	–0.282**	–0.108	–0.14

The significance level of the pairwise correlations is displayed with asterisks, where *** stands for a p-value below 0.01 (ie extremely significant), ** stands for a p-value below 0.05 (ie very significant) and * stands for a p-value below 0.1 (ie significant).

Source: Authors' calculations.

2. **Middle office reporting directly to the board and investment policies**. Additionally, we find a statistically positive correlation between the middle office's reporting and a central bank's risk-taking and investment policies (Table 4). When

[18] We estimate the implied annual volatility of the portfolio (ie absolute risk measure) based on the respondents' actual asset allocation using data from the benchmark's performance, assuming they follow a passive investment strategy and replicate the underlying benchmark.

the middle office reports directly to the board, we observe a higher allocation to non-traditional assets and more diversified reserve portfolios as the number of currencies and asset classes increases (see Figure 3).

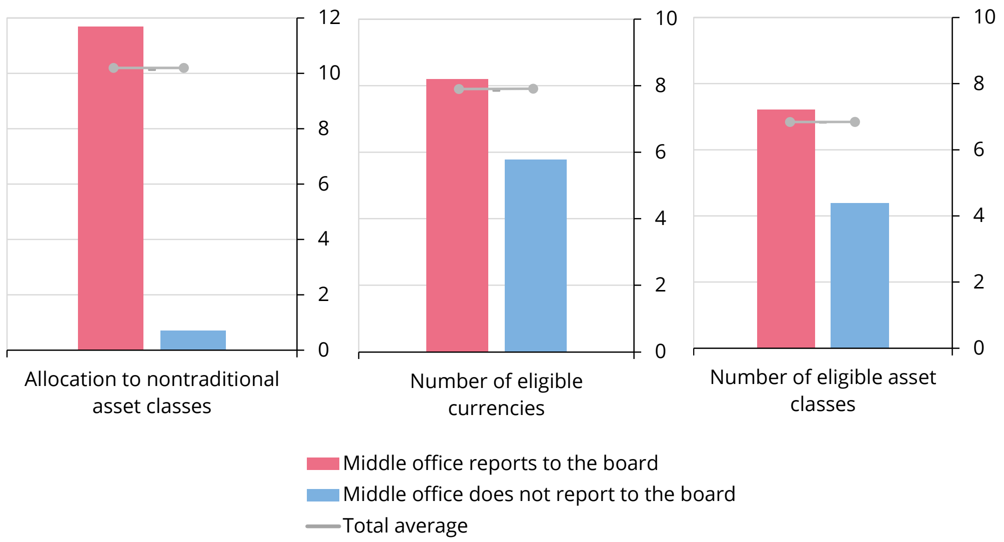
Reporting to the board and investment policies — Table 4

Variables	Estimated risk of the portfolio	Non-traditional assets (%)	Number of eligible currencies	Number of eligible assets	Investment horizon – total portfolio (months)	Duration – total portfolio (months)	Duration – liquidity tranche (months)	Duration – investment tranche (months)
Middle office reports to the board	0.123	0.227*	0.193*	0.345***	–0.1	0.114	0.153	–0.006

The significance level of the pairwise correlations is displayed with asterisks, where *** stands for a p-value below 0.01 (ie extremely significant), ** stands for a p-value below 0.05 (ie very significant) and * stands for a p-value below 0.1 (ie significant).

Source: Authors' calculations.

Reporting to the board and investment policies — Figure 3



Number of respondents: 71, 100 and 97, respectively.

Source: RAMP Survey on the Reserve Management Practices of Central Banks (World Bank Treasury).

3. **Organisation of operational units and investment policies**. We find a statistically significant correlation between the organisation of the operational units and the diversification of reserve management portfolios. Central banks with the back, middle and front offices in the same department have less diversified portfolios and shorter investment horizons (see Table 5). As discussed earlier, the RAMP surveys show that one third of central banks have their front, middle and back office functions together.

Organisation of operational units and investment policies — Table 5

Variables	Estimated risk of the portfolio	Non-traditional assets (%)	Number of eligible currencies	Number of eligible assets	Investment horizon – total portfolio (months)	Duration – total portfolio (months)	Duration – liquidity tranche (months)	Duration – investment tranche (months)
Back office, middle office and front office are in the same department	–0.178	–0.234**	–0.073	–0.204**	–0.264**	–0.088	–0.045	0.114

The significance level of the pairwise correlations is displayed with asterisks, where *** stands for a p-value below 0.01 (ie extremely significant), ** stands for a p-value below 0.05 (ie very significant) and * stands for a p-value below 0.1 (ie significant).

Source: Authors' calculations.

4. **Obligation to cover negative equity and investment policies**. Our empirical analysis shows that central banks that do not receive financial support from the government tend to have a more diversified strategic asset allocation with lower overall portfolio risk, even when controlling for distribution policies (see Table 6). Countries in which finance ministries have an obligation to cover central banks' negative equity also have riskier macro environments. This observation is consistent with the result that the obligation to cover negative equity is more common in countries with weaker governance. It also may indicate that the necessity of government support for the central bank may be higher in such environments as the economy is less stable and prone to unforeseen shocks that the central bank is not well prepared to manage with existing resources.

Obligation to cover negative equity and investment policies — Table 6

Variables	Estimated risk of the portfolio	Non-traditional assets (%)	Number of eligible currencies	Number of eligible assets	Investment horizon – total portfolio (months)	Duration – total portfolio (months)	Duration – liquidity tranche (months)	Duration – investment tranche (months)
Obligation to cover negative equity	–0.264**	–0.200*	–0.201*	–0.251**	–0.322**	–0.172	–0.309**	–0.067
No obligation to cover negative equity **and** transfer of realised income	0.251**	0.221*	0.063	0.211**	0.246*	–0.004	0.193	0.06
Obligation to cover negative equity **and** transfer of realised income	–0.244**	–0.201*	–0.217**	–0.230**	–0.287**	–0.153	–0.202	0.011

The significance level of the pairwise correlations is displayed with asterisks, where *** stands for a p-value below 0.01 (ie extremely significant), ** stands for a p-value below 0.05 (ie very significant) and * stands for a p-value below 0.1 (ie significant).

Source: Authors' calculations.

5. **The broader institutional environments and investment policies**. Our analysis finds that the broader governance environment in which central banks operate correlates with certain investment policy types (see Table 7). As the table indicates, the government effectiveness index has a strong, statistically significant positive correlation with duration risk and portfolio diversification, measured as the number of eligible asset classes and the allocation to non-traditional asset classes (see Figure 4). Because of these findings, we will use indices that describe the broader institutional environment as control variables in our regressions.

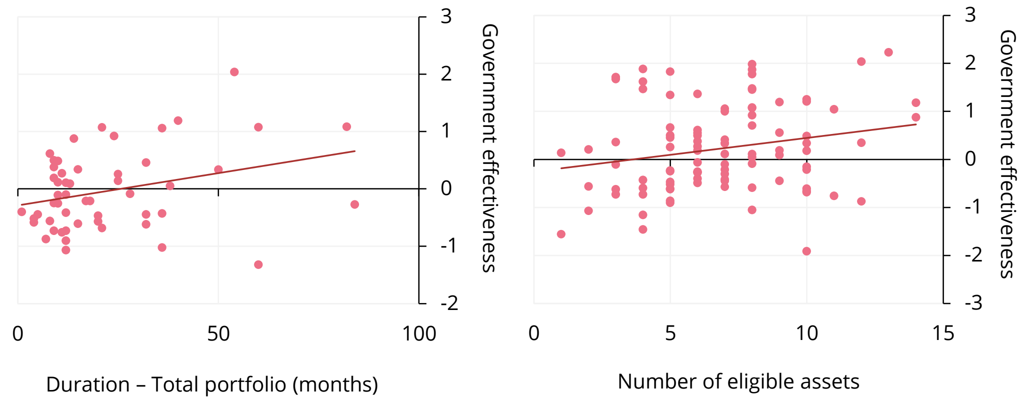
Government effectiveness and investment policies — Table 7

Variables	Independence of central bank	Corruption Perceptions Index	Governance Pillar Score	Government effectiveness
Estimated risk of the portfolio	0.275**	0.268**	0.253*	0.265**
Non-traditional assets (%)	0.162	0.114	0.177	0.239**
Number of eligible currencies	0.06	0.076	0.041	0.146
Number of eligible assets	−0.032	0.091	0.204*	0.217**
Investment horizon – total portfolio (months)	0.347**	0.348***	0.182	0.237*
Duration – total portfolio (months)	0.219	0.346**	0.087	0.308**
Duration – liquidity tranche (months)	0.349**	0.451***	0.225*	0.433***
Duration – investment tranche (months)	0.250*	0.124	0.185	−0.008

The significance level of the pairwise correlations is displayed with asterisks, where *** stands for a p-value below 0.01 (ie extremely significant), ** stands for a p-value below 0.05 (ie very significant) and * stands for a p-value below 0.1 (ie significant).

Source: Authors' calculations.

Investment policies and government effectiveness — Figure 4



Number of respondents: 52 and 97, respectively.

Sources: RAMP Survey (World Bank Treasury); Transparency International.

6. Multivariate regression results

We use regression analysis to delve into the correlation analyses on central bank governance and risk. We test the extent to which governance arrangements matter for risk-taking and diversification in foreign reserve portfolios while controlling for the broader governance environment in which central banks operate, the level of reserve adequacy, the macro environment, or country risk.[19] Additional regression results are found in Appendix II.

1. **Independent investment committee**. We examine if approving the investment guidelines at the board or investment committee level makes any difference for reserve management operations. In most central banks, either can have this role. The most significant advantage of leaving this decision to the board is that it has more authority. However, central bank board members may not be financial experts and may have less time to focus on reserve management policy. By contrast, the investment committee has less authority, but it can meet more often, and the members usually understand operational nuances and have financial expertise. We test whether the independence of an investment committee influences central bank investment policies, as suggested in our correlation exercise. As explained, we control for reserve adequacy, the general macro environment and government effectiveness, which may influence a central bank's risk-taking as expressed in its investment policy. We do not find with any statistical significance that having the investment committee approve guidelines matters, even when different control variables are used to describe the macro environment (see Table 8).

Measuring the impact of an independent investment committee on investment policies — Table 8

Independent variable	(1) Number of eligible assets	(2) Number of eligible currencies	(3) Estimated risk of the portfolio	(4) Investment horizon – total portfolio (months)
Independent investment committee	–0.544	–1.668	–0.270	4.038
	(0.810)	(1.297)	(0.239)	(11.305)
Government effectiveness	0.380	0.971	0.211	14.590**
	(0.435)	(0.728)	(0.148)	(7.062)
Reserves to GDP (%)	4.129**	5.223*	–0.789	–5.334
	(1.669)	(2.808)	(0.760)	(24.260)
Short-term external debt stocks, % of reserves	0.000	–0.000	–0.000	0.019**
	(0.000)	(0.000)	(0.000)	(0.009)
Current account balance (% of GDP)	0.064	–0.069	–0.008	–0.533
	(0.044)	(0.072)	(0.014)	(0.615)
Number of currency crises (1971–2017)	0.141	0.653	–0.062	0.621
	(0.234)	(0.394)	(0.074)	(3.841)
Constant	5.932***	5.877***	2.007***	22.020**
	(0.709)	(1.194)	(0.268)	(10.539)
Observations	83	85	63	48
R-squared	0.172	0.112	0.102	0.305

Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Each column in this table refers to a different regression with the column name as the dependent variable.

Source: Authors' calculations.

[19] This choice of variables was influenced by a general-to-specific regression (GETS) run as a way of selecting the most relevant variables out of a relatively large sample of variables when fitting a regression model.

2. **Middle office reporting directly to the board**. We find evidence that reporting lines impact investment policies. Our regression analysis confirms the importance of direct communication between the board and the middle office. We observe that reserve portfolios are more diversified in terms of eligible assets and currencies in central banks where the middle office reports to the board. This finding is significant, considering that we control for reserve adequacy, a country's macro environment and the broader governance environment (see Table 9). These results are robust when using different specifications for macro risks (see Appendix II), and the results also hold when subsampling the data by income level and level of reserves.[20] However, there does not seem to be a significant difference between the groups.

Measuring the impact of the middle office directly reporting to the board on investment policies — Table 9

Independent variable	(1) Number of eligible assets	(2) Number of eligible currencies	(3) Estimated risk of the portfolio	(4) Investment horizon – total portfolio (months)
Middle office reports to the board	2.854***	3.098**	–0.045	1.304
	(0.797)	(1.373)	(0.262)	(12.801)
Government effectiveness	0.421	1.043	0.229	14.347**
	(0.403)	(0.710)	(0.150)	(7.044)
Reserves to GDP (%)	4.065**	5.072*	–0.765	–5.020
	(1.547)	(2.748)	(0.772)	(24.300)
Short-term external debt stocks, % of reserves	0.000	–0.000	–0.000	0.019*
	(0.000)	(0.000)	(0.000)	(0.010)
Current account balance (% of GDP)	0.059	–0.055	–0.003	–0.616
	(0.039)	(0.068)	(0.013)	(0.585)
Number of currency crises (1971–2017)	0.255	0.774*	–0.065	0.564
	(0.219)	(0.390)	(0.076)	(3.843)
Constant	3.182***	2.745	2.005***	21.414
	(0.995)	(1.732)	(0.384)	(16.198)
Observations	83	85	63	48
R-squared	0.287	0.149	0.082	0.303

Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Each column in this table refers to a different regression with the column name as the dependent variable.

Source: Authors' calculations.

3. **Organisation of reserve management operations and investment policy**. We find support for the hypothesis that how institutions organise their reserve management operations matters for the investment policy of central banks. Central banks in which the front, middle and back offices are in the same department have, on average, significantly shorter investment horizons. As macro volatility, levels of reserves and the broader governance environment may influence the investment horizon of central banks' reserve

[20] The data were divided into high-income and low-income countries, on the one hand, and by level of reserves (above and below US$ 15 billion) on the other.

operations, we include these in the regression as control variables (see Table 10).

Measuring the impact of the organisational structure of reserve management operations on investment policies — Table 10

Independent variable	(1) Number of eligible assets	(2) Number of eligible currencies	(3) Estimated risk of the portfolio	(4) Investment horizon – total portfolio (months)
Back office, middle office and front office are in the same department	–0.842	–1.532	–0.243	–15.575*
	(0.603)	(1.010)	(0.194)	(8.952)
Government effectiveness	0.388	1.019	0.220	17.076**
	(0.431)	(0.723)	(0.147)	(6.974)
Reserves to GDP (%)	4.110**	5.188*	–0.702	–6.832
	(1.651)	(2.796)	(0.758)	(23.460)
Short-term external debt stocks (% of reserves)	0.000	–0.000	–0.000	0.016*
	(0.000)	(0.000)	(0.000)	(0.009)
Current account balance (% of GDP)	0.066	–0.054	–0.006	–0.885
	(0.041)	(0.069)	(0.013)	(0.581)
Number of currency crises (1971–2017)	0.153	0.661*	–0.048	1.374
	(0.232)	(0.392)	(0.074)	(3.738)
Constant	6.122***	6.111***	2.012***	25.570**
	(0.719)	(1.215)	(0.267)	(10.165)
Observations	83	85	63	48
R-squared	0.188	0.120	0.106	0.350

Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Each column in this table refers to a different regression with the column name as the dependent variable.

Source: Authors’ calculations.

4. **Obligation of the ministry of finance to cover central bank negative equity**. The distribution of profits and losses within the government is a sensitive issue for central banks. Although central banks commonly share their earnings with ministries of finance, there are questions regarding the source of those profits (ie foreign exchange revaluation or changes in market prices) and the timing (ie realised versus unrealised). Despite significant variation in their practices, central banks tend to distribute realised profits partially. Our regression analysis confirms that the obligation of the ministry of finance to cover central banks’ negative equity influences investment policy. Reserve portfolios in countries where finance ministries must cover negative equity are less diversified in terms of eligible currencies, even when controlling for reserve adequacy and the macro and broader governance environment. On average, central banks operating in countries where the ministry of finance had an obligation to cover negative equity held 2.4 to 2.7 fewer eligible currencies across the various regression specifications (see Table 11). It appears that this obligation narrowly impacts investment policies, as we cannot find robust results for asset diversification, level of risks, or the investment horizon.

Measuring the impact of the obligation of the ministry of finance to cover central bank negative equity on investment policies

Table 11

Independent variable	(1) Number of eligible assets	(2) Number of eligible currencies	(3) Estimated risk of the portfolio	(4) Investment horizon – total portfolio (months)
Obligation to cover negative equity	–0.414	–2.728**	–0.063	–7.680
	(0.756)	(1.280)	(0.235)	(10.882)
Government effectiveness	0.171	0.260	0.170	8.933
	(0.492)	(0.841)	(0.168)	(8.211)
Reserves to GDP (%)	4.516**	5.292*	–0.298	9.021
	(1.709)	(2.926)	(0.791)	(26.845)
Short-term external debt stocks (% of reserves)	0.000	–0.000	–0.000	0.039**
	(0.000)	(0.000)	(0.000)	(0.018)
Current account balance (% of GDP)	0.065	–0.084	–0.004	–0.840
	(0.043)	(0.072)	(0.013)	(0.606)
Number of currency crises (1971–2017)	0.134	0.511	–0.032	0.324
	(0.235)	(0.404)	(0.073)	(4.122)
Constant	5.930***	7.401***	1.797***	21.772
	(0.834)	(1.432)	(0.307)	(14.276)
Observations	77	78	58	43
R-squared	0.178	0.141	0.055	0.321

Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Each column in this table refers to a different regression with the column name as the dependent variable.

Source: Authors' calculations.

7. Conclusion and policy implications

Effective governance is essential for central banks and their reserve management function. Multiple authors have analysed central bank governance arrangements around the world. These publications conclude that governance practices vary from country to country, but some principles seem broadly relevant for achieving positive outcomes. Empirical evidence points to the importance of central bank independence and transparency for monetary policy. Different publications agree that a proper governance structure is the cornerstone of a successful investment operation in reserve management. However, most publications on reserve management governance are prescriptive and qualitative.

We contribute to the discussion on reserve management governance with data-driven analysis. Our novel empirical study indicates that specific governance arrangements impact investment policy and risk-taking in central banks' reserve management operations, controlling for the macroeconomic environment, reserve levels and the broader governance environment.

We find that three types of governance factors relate to diversification and risk in foreign reserve portfolios. First, direct communication between the board and the middle office often coincides with more diversified reserve operations. Notably, the result holds when controlling for reserve adequacy, and country risk indicates that anchoring risk-taking at the board level allows reserve managers to have more diversified portfolios in terms of eligible assets and eligible currencies, everything else equal. Second, we find that the organisational structure for reserve operations may impact investment policy. Having all three units responsible for managing reserve operations in the same department, with the same reporting line, affects central banks' ability to take risks, as reflected in the investment horizon. Controlling for the macro and governance environment and reserve adequacy, central banks with the same reporting lines for the front, middle and back offices, on average, have a significantly shorter investment horizon. We also find that this result is robust across different specifications. Third, countries where the ministry of finance is obligated to cover negative equity have, on average, investment policies with fewer eligible currencies than countries where the ministry of finance does not have such obligations.

The most important policy implication of our analysis is the board's critical role in reserve management. Central banks where boards actively exercise portfolio oversight (ie the middle office reports directly to the board) usually have portfolios with more risk and diversification. While ability and tolerance for risk-taking vary across central banks, portfolios with longer investment horizons, more currencies and more asset classes have performed better historically while limiting downside risk. Given that we control for the broader governance environment, our data indicate that any central bank can improve its internal governance regardless of the external governance environment. Several central banks in our database have implemented robust reserve management practices even without optimal external governance environments.

8. Appendix I: Correlation results

Central bank independence and the broader governance environment — Table 8.1

Variables	Independence of central bank	Corruption Perceptions Index	Governance Pillar Score	Government effectiveness
Independence of central bank	1			
Corruption Perceptions Index	0.472***	1		
Governance Pillar Score	0.346***	0.690***	1	
Government effectiveness	0.440***	0.928***	0.720***	1
GDP per capita	0.383***	0.803***	0.609***	0.784***

The significance level of the pairwise correlations is displayed with asterisks, where *** stands for a p-value below 0.01 (ie extremely significant), ** stands for a p-value below 0.05 (ie very significant) and * stands for a p-value below 0.1 (ie significant).

Source: Authors' calculations.

Country risk and the broader governance environment — Table 8.2

Variables	Independence of central bank	Corruption Perceptions Index	Governance Pillar Score	Government effectiveness
Country risk score	–0.484***	–0.828***	–0.638***	–0.880***

The significance level of the pairwise correlations is displayed with asterisks, where *** stands for a p-value below 0.01 (ie extremely significant), ** stands for a p-value below 0.05 (ie very significant) and * stands for a p-value below 0.1 (ie significant).

Source: Authors' calculations.

Transparency of reserve management policies and the broader governance environment — Table 8.3

Variables	Independence of central bank	Corruption Perceptions Index	Governance Pillar Score	Government effectiveness
Transparency of reserve management policies	0.298***	0.117	0.082	0.084

The significance level of the pairwise correlations is displayed with asterisks, where *** stands for a p-value below 0.01 (ie extremely significant), ** stands for a p-value below 0.05 (ie very significant) and * stands for a p-value below 0.1 (ie significant).

Source: Authors' calculations.

Transparency of reserve management policies and country risk

Table 8.4

Variables	Country risk score
Transparency of reserve management policies	–0.222*

The significance level of the pairwise correlations is displayed with asterisks, where *** stands for a p-value below 0.01 (ie extremely significant), ** stands for a p-value below 0.05 (ie very significant) and * stands for a p-value below 0.1 (ie significant).

Source: Authors' calculations.

Investment policies and the broader governance environment

Table 8.5

Variables	Independence of central bank	Corruption Perceptions Index	Governance Pillar Score	Government effectiveness
Estimated risk of the portfolio	0.275**	0.268**	0.253*	0.265**
Non-traditional assets (%)	0.162	0.114	0.177	0.239**
Number of eligible currencies	0.06	0.076	0.041	0.146
Number of eligible assets	–0.032	0.091	0.204*	0.217**
Investment horizon – total portfolio (months)	0.347**	0.348***	0.182	0.237*
Duration – total portfolio (months)	0.219	0.346**	0.087	0.308**
Duration – liquidity tranche (months)	0.349**	0.451***	0.225*	0.433***
Duration – investment tranche (months)	0.250*	0.124	0.185	–0.008

The significance level of the pairwise correlations is displayed with asterisks, where *** stands for a p-value below 0.01 (ie extremely significant), ** stands for a p-value below 0.05 (ie very significant) and * stands for a p-value below 0.1 (ie significant).

Source: Authors' calculations.

Country risk and investment policies

Table 8.6

Variables	Country risk score
Estimated risk of the portfolio	–0.408***
Non-traditional assets (%)	–0.423***
Number of eligible currencies	–0.041
Number of eligible assets	–0.157
Investment horizon – total tranched and untranched portfolio (months)	–0.404***
Duration – total portfolio (months)	–0.384**
Duration – liquidity tranche (months)	–0.521***
Duration – investment tranche (months)	–0.296**

The significance level of the pairwise correlations is displayed with asterisks, where *** stands for a p-value below 0.01 (ie extremely significant), ** stands for a p-value below 0.05 (ie very significant) and * stands for a p-value below 0.1 (ie significant).

Source: Authors' calculations.

Independence of the investment committee and the broader governance environment

Table 8.7

Variables	Independence of central bank	Corruption Perceptions Index	Governance Pillar Score	Government effectiveness
Independent investment committee	–0.018	–0.186*	–0.180*	–0.204**

The significance level of the pairwise correlations is displayed with asterisks, where *** stands for a p-value below 0.01 (ie extremely significant), ** stands for a p-value below 0.05 (ie very significant) and * stands for a p-value below 0.1 (ie significant).

Source: Authors' calculations.

Independence of the investment committee and investment policies

Table 8.8

Variables	Estimated risk of the portfolio	Non-traditional assets (%)	Number of eligible currencies	Number of eligible assets	Investment horizon – total portfolio (months)	Duration – total portfolio (months)	Duration – liquidity tranche (months)	Duration – investment tranche (months)
Independent investment committee	–0.179	–0.233*	–0.102	–0.148	–0.036	–0.282**	–0.108	–0.14

The significance level of the pairwise correlations is displayed with asterisks, where *** stands for a p-value below 0.01 (ie extremely significant), ** stands for a p-value below 0.05 (ie very significant) and * stands for a p-value below 0.1 (ie significant).

Source: Authors' calculations.

Reporting to the board and investment policies

Table 8.9

Variables	Estimated risk of the portfolio	Non-traditional assets (%)	Number of eligible currencies	Number of eligible assets	Investment horizon – total portfolio (months)	Duration – total portfolio (months)	Duration – liquidity tranche (months)	Duration – investment tranche (months)
Middle office reports to the board	0.123	0.227*	0.193*	0.345***	–0.1	0.114	0.153	–0.006

The significance level of the pairwise correlations is displayed with asterisks, where *** stands for a p-value below 0.01 (ie extremely significant), ** stands for a p-value below 0.05 (ie very significant) and * stands for a p-value below 0.1 (ie significant).

Source: Authors' calculations.

Obligation to cover negative equity and the broader governance environment

Table 8.10

Variables	Independence of central bank	Corruption Perceptions Index	Governance Pillar Score	Government effectiveness
Obligation to cover negative equity	–0.194*	–0.479***	–0.357***	–0.483***

The significance level of the pairwise correlations is displayed with asterisks, where *** stands for a p-value below 0.01 (ie extremely significant), ** stands for a p-value below 0.05 (ie very significant) and * stands for a p-value below 0.1 (ie significant).

Source: Authors' calculations.

Obligation to cover negative equity and investment policies

Table 8.11

Variables	Estimated risk of the portfolio	Non-traditional assets (%)	Number of eligible currencies	Number of eligible assets	Investment horizon – total portfolio (months)	Duration – total portfolio (months)	Duration – liquidity tranche (months)	Duration – investment tranche (months)
Obligation to cover negative equity	–0.264**	–0.200*	–0.201*	–0.251**	–0.322**	–0.172	–0.309**	–0.067
No obligation to cover negative equity **and** transfer of realised income	0.251**	0.221*	0.063	0.211**	0.246*	–0.004	0.193	0.06
Obligation to cover negative equity **and** transfer of realised income	–0.244**	–0.201*	–0.217**	–0.230**	–0.287**	–0.153	–0.202	0.011

The significance level of the pairwise correlations is displayed with asterisks, where *** stands for a p-value below 0.01 (ie extremely significant), ** stands for a p-value below 0.05 (ie very significant) and * stands for a p-value below 0.1 (ie significant).

Source: Authors' calculations.

9. Appendix II: Additional regression results

Measuring the impact of organisational structure on investment policies — Table 9.1

Independent variable	(1) Number of eligible assets	(2) Number of eligible currencies	(3) Estimated risk of the portfolio	(4) Investment horizon – total portfolio (months)
Middle office reports to the board	2.544***	2.262*	0.0365	–8.061
	(0.757)	(1.263)	(0.251)	(11.46)
Reserves to GDP (%)	2.375	3.755	–0.330	–18.82
	(1.445)	(2.479)	(0.668)	(19.93)
Government effectiveness	0.625**	0.322	0.254**	14.07***
	(0.306)	(0.512)	(0.117)	(4.892)
Constant	3.865***	5.050***	1.736***	41.58***
	(0.756)	(1.254)	(0.267)	(11.00)
Observations	91	94	66	53
R-squared	0.201	0.073	0.076	0.156

Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Each column in this table refers to a different regression with the column name as the dependent variable.

Source: Authors' calculations.

Measuring the impact of organisational structure on investment policies — Table 9.2

Independent variable	(1) Number of eligible assets	(2) Number of eligible currencies	(3) Estimated risk of the portfolio	(4) Investment horizon – total portfolio (months)
Back office, middle office and front office are in the same department	–0.834	–0.869	–0.247	–14.83*
	(0.584)	(0.954)	(0.189)	(8.311)
Government effectiveness	0.701**	0.401	0.222*	14.53***
	(0.321)	(0.516)	(0.118)	(4.768)
Reserves to GDP (%)	2.700*	4.102	–0.273	–19.35
	(1.518)	(2.508)	(0.659)	(19.20)
Constant	6.231***	7.169***	1.848***	38.74***
	(0.492)	(0.801)	(0.181)	(6.111)
Observations	91	94	66	53
R-squared	0.118	0.049	0.101	0.200

Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Each column in this table refers to a different regression with the column name as the dependent variable.

Source: Authors' calculations.

Measuring the impact of requiring the ministry of finance to cover negative equity — Table 9.3

Independent variable	(1) Number of eligible assets	(2) Number of eligible currencies	(3) Estimated risk of the portfolio	(4) Investment horizon – total portfolio (months)
Obligation to cover negative equity	–0.977	–2.403**	0.0702	–7.063
	(0.703)	(1.179)	(0.223)	(10.02)
Reserves to GDP (%)	3.562**	4.414	–0.151	–19.22
	(1.637)	(2.755)	(0.684)	(22.62)
Government effectiveness	0.345	–0.317	0.258*	11.14*
	(0.403)	(0.666)	(0.139)	(6.363)
Constant	6.327***	8.437***	1.655***	39.26***
	(0.622)	(1.047)	(0.203)	(8.733)
Observations	82	84	61	46
R-squared	0.122	0.079	0.067	0.143

Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Each column in this table refers to a different regression with the column name as the dependent variable.

Source: Authors' calculations.

10. Appendix III: Central bank governance and financial supervision

Central banks often have mandates regarding financial stability and supervision, but central bank legislation is often less specific about central banks' clear functions and responsibilities in this regard (Ortiz (2009)). Nonetheless, elements of this task, such as the lender of last resort function or oversight of the payment system, have long been central bank functions (Freixas et al (2000)).[21] Although the trend since the 1990s has been toward consolidation of financial supervisory functions, these changes have primarily focused on consolidating functions in an agency separate from the central bank. Despite those developments, according to a survey of 160 countries completed in 2019, approximately 68% of central banks still have banking supervisory responsibilities (Anginer et al (2019)).

Empirical evidence on the link between central bank independence and financial supervisory responsibilities remains limited. Some research has shown that central bank independence positively impacts bank soundness and that bank involvement in banking supervision mitigates the adverse effects of financial crises (Doumpos et al (2015)). However, this more extensive mandate can impact the central bank's autonomy, decision-making and transparency. Initial findings from the research explored in this report indicate that countries with better overall governance environments and lower country risk usually have an entity different from the central bank for financial supervision, as seen in Table 10.1. The variables used here are

[21] Henry Thornton (1802) and Walter Bagehot (1873) developed the classic doctrine of lender of last resort. According to Bagehot, in a panic situation monetary authorities should lend unsparingly but at a penalty rate to illiquid but solvent banks. See Freixas et al (2000) for further background.

described earlier in the paper. As a measure of whether central banks also have supervisory responsibilities, we use the World Bank's regulation and supervision survey.[22] This comprehensive survey covers multiple subjects on banking regulation and supervision practices in 160 jurisdictions. We use this survey to establish whether the central bank or a separate entity is responsible for financial supervision. This variable aims to analyse whether reserve management differs in central banks with more responsibilities and more complex governance structures. The survey's last edition was in 2019.

Financial supervision and the broader governance environment — Table 10.1

Variables	Independence of central bank	Corruption Perceptions Index	Governance Pillar Score	Government effectiveness
Central bank is financial supervisor	–0.086	–0.268***	–0.281***	–0.291***

The significance level of the pairwise correlations is displayed with asterisks, where *** stands for a p-value below 0.01 (ie extremely significant), ** stands for a p-value below 0.05 (ie very significant) and * stands for a p-value below 0.1 (ie significant). 2

Source: Authors' calculations.

Regarding the impact on reserve management, the correlations suggest that countries where the central bank is also a financial supervisor tend to have an independent investment committee to which the middle office reports (see Table 10.2). This arrangement may result in a lower risk appetite in managing reserves and imply that the central bank accounts for and manages the risks of a more volatile governance environment in which it operates. As other findings of this paper show, in deficient overall governance environments, central banks tend to have independent investment committees because they can help safeguard decisions made by a weak and/or politically influenced board. However, the literature and data are scant on whether a central bank with supervisory functions impacts reserve management and policy decision-making. This question may be another line of future inquiry and research.

Financial supervision and the governance of reserve management — Table 10.2

Variables	Independent investment committee	Middle office reports to the investment committee	Obligation to cover negative equity
Central bank is financial supervisor	0.255**	0.243**	0.277**

The significance level of the pairwise correlations is displayed with asterisks, where *** stands for a p-value below 0.01 (ie extremely significant), ** stands for a p-value below 0.05 (ie very significant) and * stands for a p-value below 0.1 (ie significant).

Source: Authors' calculations.

22 www.worldbank.org/en/research/brief/BRSS.

References

Adrian, T and A Khan (2019): *Central bank accountability, independence, and transparency*, IMFBlog: Insights and Analysis on Economics and Finance, International Monetary Fund, Washington, DC, USA, https://blogs.imf.org/2019/11/25/central-bank-accountability-independence-and-transparency/.

Alekasir, K, N Anasashvili, M Antonio, P Hong, D Klingebiel, B Murira-Njogu, G Pereira Alves, M Ruiz Gil, A Slater and D Takahashi (2019): *Inaugural RAMP Survey on the reserve management practices of central banks: results and observations*, RAMP (Reserve Advisory and Management Partnership), World Bank Group, Washington, DC.

Al-Hassan, A, P Farahmand and M Papaianou (2014): *Revised guidelines for foreign exchange reserve management*, International Monetary Fund, Washington, DC, USA.

Anasashvili, N, C Herrero Montes, P Hong, D Klingebiel and M Ruiz Gil (2020): *Central bank reserve management practices: insights into public asset management from the second RAMP Survey*, RAMP (Reserve Advisory and Management Partnership), World Bank Group, Washington, DC, USA.

Anginer, D, A Bertay, R Cull, A Demirguc-Kunt and D Mare (2019): "Bank regulation and supervision ten years after the global financial crisis", *Policy Research Working Papers*, no 9044, World Bank Group, Washington, DC, USA, https://openknowledge.worldbank.org/handle/10986/32589.

Archer, D and P Moser-Boehm (2013): "Central bank finances", *BIS Papers*, no 71, April, www.bis.org/publ/bppdf/bispap71.pdf.

Arslan, Y and C Cantú (2019): "The size of foreign exchange reserves", in "Reserve management and FX intervention" *BIS Papers*, no 104, October.

Balls, E, J Howat and A Stansbury (2018): "Central bank independence revisited: after the financial crisis, what should a model central bank look like?", *M-RCBG Associate Working Paper Series*, no 87, Harvard Kennedy School, Mossavar-Rahmani Center for Business and Government, Cambridge, USA.

Bossu, W and A Rossi (2019): "The role of board oversight in central bank governance: key legal design issues", *IMF Working Papers*, no 2019/293, December.

Cukierman, A (2008): "Central bank independence and monetary policymaking institutions – past, present and future", *European Journal of Political Economy,* vol 24, no 4, December, pp 722–36.

de Abreau Faria, F and R Streit (2016): "Governance in central banks: a comparative study of the governance practices used by the central banks of Brazil, Canada and England", *Revista de Administração Pública*, vol 50, no 5, September/October, pp 765–93, www.scielo.br/pdf/rap/v50n5/en_0034-7612-rap-50-05-00765.pdf.

de Haan, J and S Eijffinger (2016): "The politics of central bank independence", *DNB Working Papers*, no 539, December.

Dincer, N and B Eichengreen (2014): "Central bank transparency and independence: updates and new measures", *International Journal of Central Banking*, vol 10, no 1, March, pp 189–253, www.ijcb.org/journal/ijcb14q1a6.pdf.

Doumpos, M, C Gaganis and F Pasiouras (2015): "Central bank independence, financial supervision structure and bank soundness: an empirical analysis around the crisis", *Journal of Banking and Finance*, vol 61, supplement 1, December, pp 69–83, www.sciencedirect.com/science/article/pii/S0378426615001296.

Freixas, X, C Giannini, G Hoggarth and F Soussa (2000): "Lender of last resort: what have we learned since Bagehot?", *Journal of Financial Services Research,* vol 18, October, pp 63–84, https://doi.org/10.1023/A:1026527607455.

Furceri, D and A Zdzienicka (2013): "The euro area crisis: need for a supranational fiscal risk sharing mechanism?", *IMF Working Papers*, no 2013/198, September.

Haggard, S and L Tiede (2011): "The rule of law and economic growth: where are we?", *World Development*, vol 39, no 5, May, pp 673–85, https://doi.org/10.1016/j.worlddev.2010.10.007.

International Monetary Fund (2019): "Safeguarding financial stability: the role of macroprudential policy", in K Beaton, R Garcia-Saltos and L Figliuoli (eds), *Paving the way to sustained growth and prosperity in Central America, Panama, and the Dominican Republic*, September, www.elibrary.imf.org/display/book/9781484353844/9781484353844.xml.

International Monetary Fund (2020): "A global accounting of the monetary and fiscal policy actions taken by governments worldwide: International Monetary Fund policy responses to COVID-19." *IMF Topics*, International Monetary Fund, Washington, DC.

Johnson-Calari, J, and I Strauss-Kahn (2020): "Good governance: principles, pitfalls, and best practice" Chapter 18, *Asset Management at Central Banks and Monetary Authorities: New Practices in Managing International Foreign Exchange Reserves*, edited by Jacob Bjorheim, Springer.

Keefer, P and S Knack (2002): "Polarization, politics and property rights: links between inequality and growth", *Public Choice*, vol 111, no 1–2, pp 127–54, https://doi.org/10.1023/A:1013816423302.

Khan, A (2017): "Central bank legal frameworks in the aftermath of the global financial crisis", *IMF Working Papers*, no 2017/101, May.

Kontopoulos, Y and R Perotti (1999): "Government fragmentation and fiscal policy outcomes: evidence from OECD countries", in J Poterba and J von Hagen (eds), *Fiscal institutions and fiscal performance*, pp 81–102, University of Chicago Press.

Levy Yeyati, E, F Sturzenegger and I Reggio (2010): "On the endogeneity of exchange rate regimes", *European Economic Review*, vol 54, no 5, July, pp 659–77, https://doi.org/10.1016/j.euroecorev.2009.12.004.

Mauro, P (1995): "Corruption and growth", *The Quarterly Journal of Economics*, vol 110, no 3, August, pp 681–712, https://doi.org/10.2307/2946696.

Nellor, D (2011): "Central bank independence, accountability, and transparency: a global perspective", in D Nellor and C Leong (eds), *Building strong banks through surveillance and resolution*, pp 65–90, Singapore: World Scientific Publishing.

Neuenkirch, M and B Hayo (2015): "The influence of media use on laymen's monetary policy knowledge in Germany", *MAGKS Papers on Economics*, no 201511, Philipps-Universität Marburg, Faculty of Business Administration and Economics, Department of Economics (Volkswirtschaftliche Abteilung).

Ortega de la Rosa, J (2015): "International reserves and governance: a comparative study for Colombia", *Dimensión Empresarial,* vol 13, no 2, July–December, pp 11–31, www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1692-85632015000200002.

Ortiz, G (2009): "Issues in the governance of central banks", Report from the central bank governance group, BIS.

Rose, A and M Spiegel (2009): "Cross-country causes and consequences of the 2008 crisis: early warning", *NBER Working Papers*, no 15357, September.

Ruiz, M (2020): "Governance, risk management, reporting, and control at the central bank of Colombia" Chapter 20, *Asset Management at Central Banks and Monetary Authorities: New Practices in Managing International Foreign Exchange Reserves*, edited by Jacob Bjorheim, Springer.

——— (2010): "Cross-country causes and consequences of the 2008 crisis: international linkages and American exposure", in R Baldwin and L Evenett (eds), *The collapse of global trade, murky protectionism, and the crisis: recommendations for the G20*, pp 169–94, Centre for Economic Policy Research, London, United Kingdom.

Talvi, E and C Vegh (2005): "Tax base variability and procyclical fiscal policy in developing countries", *Journal of Development Economics*, vol 78, no 1, October, pp 156–90, https://doi.org/10.1016/j.jdeveco.2004.07.002.

Vegh, C and G Vuletin (2015): "How is tax policy conducted over the business cycle?", *American Economic Journal: Economic Policy*, vol 7, no 3, pp 327–70.

World Bank Group (2019): *Doing business 2020: comparing business regulation in 190 economies*.

Effective financial risk management in unconventional times

Thomas Brophy, David Doran and Steve Kilkenny[1]

Abstract

In recent years, in particular after the Great Financial Crisis and the pandemic, the balance sheets of central banks have become more complex. This has been due not only to the provision of large amounts of liquidity into the financial system through traditional monetary policy lending, but also to implementation of unprecedented quantitative easing (QE) policies. These developments have enlarged central bank balance sheets while changing the composition of the asset base, creating the conditions for an interest rate mismatch in central bank balance sheets. Up until recently, this has coincided with a low and negative interest rate environment across many developed economies, where risk premia were compressed due in part to the above policies. However, in the post-pandemic period, a sustained inflationary environment has forced central banks to raise interest rates and tighten financial conditions, with knock-on effects on the yields of many sovereign markets. This had led to a higher cost of liabilities for many public investors relative to the income earned on assets. This paper presents examples of two approaches that public investors can adopt to identify and respond to the risks of this scenario for their investment assets in a manner that recognises the contemporaneous financial environment relative to the cycle. The first method entails adopting an "all-balance sheet" approach to assessing investment risk, in the light of material non-investment risks arising from QE. The second approach proposes a means of defining a risk appetite for a public investor, with the aim of balancing investment risk tolerance with changing risk environments over time, in the light of risks elsewhere on the balance sheet.

JEL classification: G32.

[1] The authors are Head of Financial Risk Management and Head of Organisational Risk Division and Head of Investment Policy & Risk Monitoring, respectively, in the Central Bank of Ireland. The views expressed in this article are solely the views of the authors and are not necessarily those held by the Central Bank of Ireland or the European System of Central Banks. The authors would like to thank Glenn Calverley and Eimear Keogh for helpful comments – any remaining errors or omissions are our own. Corresponding email: steve.kilkenny@centralbank.ie.

1. Introduction

Since the Great Financial Crisis (GFC) and up until the Covid-19 pandemic, the traditional investment universe of European public investors such as central banks, (principally reserve currency sovereign fixed income markets), was characterised by persistently low, and even negative nominal interest rates. This was largely the result of the non-standard asset purchase programmes of many central banks aimed at supporting the monetary policy stance of low interest rates and higher liquidity, which significantly increased the amount of very low or negative yielding securities on central bank balance sheets.

These developments posed a number of challenges in terms of how best to maintain effective investment risk management through the cycle, particularly as interest rates rise and the market pricing of risk premia begins to normalise. In the aftermath of the Covid-19 pandemic, as well as due to wider geopolitical developments, there has been a sustained increase in inflation rates across many global economies. A primary response of central banks to this inflationary environment, has been to tighten monetary policy through increasing policy rates and winding down QE policies. For many public investors, this is leading to a materialisation of losses due to a mismatch of interest rates on both the longer-term assets and the shorter-dated liabiliies of the balance sheet, arising from such non-standard monetary policy measures. This underscores the importance of using an all-balance sheet model when assessing the financial risk impact of exposures, including investment exposures, as well as employing scenario analysis incorporating the potential for higher interest rates. A balance sheet model is presented, that can be used for this purpose.

The post-GFC period of low and negative rates which had prevailed until end-2021 led many central banks and other public investors to diversify outside their normal investment universe, in an effort to maintain investment income and contribute towards covering operating costs and profitability. This was at a time when the effects of the record levels of systemic liquidity, coupled with record low interest rates and the impacts of asset purchase programmes, affected the extent to which market prices accurately reflected the range of financial risks (Albertazzi et al (2021)). This circumstance can lead to an unintendedly larger increase in the risk of investment exposures of a central bank as it seeks to generate better returns, but which could increase further as financial market conditions change.

The cyclical shift from a period characterised by financial market search for yield, against a backdrop of large amounts of liquidity and zero bound interest rates, conceivably created an environment conducive to a “Minsky moment” that public sector investment frameworks must guard against in a way that reflects the asset and liability dynamics of the whole balance sheet. Since 2022, there has been an increase in market volatility across most asset classes, as higher yields led to a materialisation of asset value adjustments across both fixed income and equity markets as well as other asset classes. Moreover, the recent policy rate tightening and a rise in nominal bond yields in advanced and emerging markets has altered the balance sheet asset and liability dynamics as well as the relative contribution of risk premia to asset prices, and any further tightening in monetary policy and financial conditions is likely to continue this trend.

Therefore, we present a framework for assessing financial risk that uses the traditional risk metrics such as expected shortfall and value-at-risk, as compared with financial buffers, but complements it with a novel environmental score that seeks to account for changes in the financial risk environment, including the impact of geopolitical risks on the central bank's investment universe. The parameters of this score can be calibrated to reflect variables such as a public investor's risk appetite, as well as other parts of the balance sheet where there is potential for financial risk. This facilitates the application of scenario analysis, to allow the public authority to assess the potential riskiness of its portfolio or balance sheet, relative to its risk appetite, in prescribed future states.

The paper is structured as follows:

Section 2 sets out the recent period of financial market risk pricing uncertainty, set against the prevailing interest rate environment. Section 3 introduces a balance sheet model, and how it can be used as an investment risk tool in light of interest rate mismatch risk. Section 4 outlines a novel approach to assessing investment risk that operates as a through-the-cycle indicator of risk levels and helps to guide asset allocation decisions. Section 5 presents conclusions.

2. Interest rate cycles and pricing of financial risk

2.1 Low interest rate environment

Traditionally, the investable universe of public investors, in particular central banks, has focused on highly liquid and highly rated fixed income assets. This is reflective of the low risk appetite characteristic of public investors, as well as the role that such assets play in the delivery of mandates related to exchange rate and macroeconomic stability, ie these assets need to be sellable within a reasonable timeframe and close to current market price (Doran et al (2020)). Other roles for investment assets include being able to help cover the operating costs of a public investor, as part of contributing to central bank policy independence.

In the midst of the exceptionally long period of low and often negative interest rates for the assets until end-2021, as mentioned above, public investors' investment frameworks faced acute challenges as returns became increasingly difficult to generate within their risk appetite. In addition to historically low interest rates, asset purchase programmes transmitted accommodative monetary policy across the yield curve as well as reducing asset availability due to the scale of asset purchase programmes.

In particular, the advent of negative interest rates across much of public investor's traditional universe created something of a paradox, challenging the ability to adhere to conservative investment policy principles. Also, given the scale and extended time frame of the asset purchases undertaken during this period, the availability of suitable investments became increasingly constrained. In the context of a policy principle of avoiding losses, this became incompatible with investing in highly liquid and highly rated fixed income assets. This was particularly the case for public investors, who use their investment assets to help cover their operating costs.

This forced public investors to question their risk appetite and to consider purchasing new or additional asset classes, often with lower credit quality and liquidity than they would otherwise have done, in order to generate return and keep investment losses to a minimum.

2.2 Pricing of financial risk

One of the drivers of the low and negative interest rate environment that public investors have faced was the QE policies implemented by several global central banks (ECB (2015)). These policies involved large-scale purchases of government bonds or other financial assets, in order to stimulate economic activity or to achieve a price stability mandate. The policies were intended to operate through a number of channels, one of which was the portfolio rebalancing channel. Through this channel, as central banks purchased eligible assets and expanded the excess liquidity in the financial system and lower yields, investors were encouraged to rebalance towards riskier assets as demand for "safer" assets increases (Albertazzi et al (2021), Bua and Dunne (2017)). This could then have the effect of increasing the demand and prices for riskier assets and, all else being equal, reduce the risk premia being offered to investors.

This compression of risk premia had the effect of de-anchoring asset prices from fundamentals in some cases. In these areas of the financial markets, risk premia were forgone in lieu of seeking assets that offered some level of nominal return. An example is the S&P 500 index, which from peak to trough rose almost sevenfold between 2009 and end-2021. Another example is the corporate bond market, where yields on many issuers fell, with some turning negative in recent years.[2]

2.3 Effect on public investors' universe

The portfolio rebalancing channel described above also affected the investable universe of public investors, and any relevant risk-return trade-offs. Until recently, returns have been compressed across many of the traditional assets that public investors use to construct their portfolios.

By way of example, in the Central Bank of Ireland's (CBoI) case, its discretionary investment assets were wholly denominated in euros prior to the beginning of QE and the period of sustained low interest rates. Partly in reflection of the low and negative interest rate environment, the CBoI made efforts to progressively diversify its investment assets to ensure a robust variety of income sources and to help cover its operating costs and contribute towards financial independence (CBoI (2022)). This involved the creation of a number of foreign currency (FX) portfolios, an equity portfolio as well as increasing holdings of physical gold. These actions were intended to enhance the resilience of income over the longer term, and help to safeguard financial independence, notwithstanding some potential for variability in returns over the short term.

As with other euro area central banks, the purpose of asset diversification was to build the resilience of the balance sheet through the economic cycle, reducing concentration and gaining exposure to different interest rate cycles. Additionally, the recent sustained increase in inflation across many of the traditional public investor markets, with a related increase in central bank policy rates and nominal rates, has meant that some of the markets, such as in the euro area, have become a more compelling investment case.

[2] The ICE BofA 1-10 Year AAA-A Euro Corporate Index was negative-yielding on a number of occasions over 2019–21.

3. Balance sheet perspective on investment risk

3.1 Interest rate mismatch risk

The effect of monetary policy asset purchase programmes has been that central banks (including the CBoI) have purchased large amounts of medium- to longer-dated sovereign bonds. In the case of the Eurosystem, this has been taking place since 2015, and these holdings are accounted for at amortised cost. Typically, due to the prevailing interest rate environment, these bonds were purchased at low or negative rates, and will likely remain on public investors' balance sheets for some time, given the long-dated nature of the holdings.

These transactions correspond with a sizeable increase in the liabilities of central banks' balance sheets. These liabilities are remunerated at variable interest rates linked to short term policy rates. Given this balance sheet structure, central banks such as the CBoI are exposed to an interest rate mismatch between income earned on assets, and the cost of funding attached to liabilities. This mismatch can have a material impact on income should the cost of liabilities increase significantly (Donnery et al (2015)) while income from assets is slower to adjust. The recent sustained increase in inflation has led many central banks to respond by increasing their policy rates, in order to deliver on their price stability mandate. As a result, the anticipated interest rate mismatch risk has materialised, with many central banks expecting to experience a period of very low profitability or losses.

3.2 Central Bank of Ireland balance sheet model

Given the interest rate mismatch on the CBoI balance sheet, a quantitative analysis is undertaken on a regular basis to assess the required level of financial provisions to be held against this risk, as well as other financial risks. In order to complete these analyses, the CBoI uses an in-built model that incorporates both the current and expected structure of the balance sheet over a 10-year horizon. This is combined with a scenario-based approach to key financial market and economic variables, which allows the CBoI to compute a breakdown of profitability outcomes across various potential future states.

The quantification is also supported by an assessment made according to the prevailing financial environment and the applicable accounting rules. Supporting this assessment, the model identifies a range of risk sensitivities, using financial market and economic scenario data generated from both a historical and market-implied perspective. It is important to note here that it is a statistically driven quantification exercise, and does not involve taking a specific market view on scenarios.

Using this balance sheet information as well as the scenario data, the CBoI can project forward both the profitability and balance sheet positions for 10 years. A scenario can then be chosen at the relevant percentile for the assessment of provisions (eg the 99th percentile). This is completed by aggregating losses across the 10-year horizon, to help quantify the provision that is required. This is described as a gross cumulative loss estimate. As required, this analysis can be supplemented by internally developed scenarios to further validate assumptions and judgments in respect of the anticipated evolution of idiosyncratic financial risks.

3.3 Investment risk in the context of balance sheet

While the balance sheet model described above was developed primarily for provisioning exercises and estimating interest rate mismatch risk, it can also be used to assess investment risk for both current and prospective investments. It can help analysts and decision-makers to better understand the interaction between investment risk, and the other risks present on the balance sheet, particularly interest rate mismatch risk.

In order to illustrate this in a hypothetical way, we can consider the baseline balance sheet position against a number of changes in investment strategy:

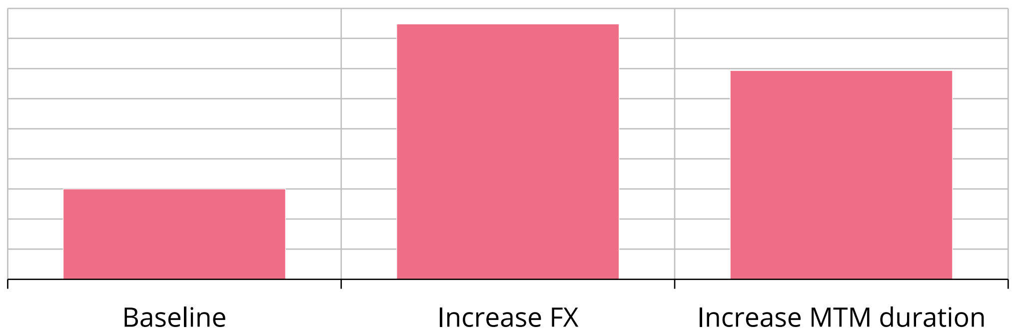
1. *Increase duration of portfolios*: In this example, the duration, or sensitivity to changes in the interest rates, is increased for all relevant marked-to-market portfolios.
2. *Increase the size of FX portfolios:* In this example, the size of the overall foreign denominated currencies, is increased, funded by a reduction in euro-denominated assets, but the composition and other parameters remain unchanged.

The analysis from these different options can be assessed on a standalone basis, but also in conjunction with the overall balance sheet risk.

3.4 Interaction of investment risk and interest rate mismatch risk

It can be seen from Graph 1 below that changes in investment strategy, as described in the previous section, can impact on the gross cumulative loss estimates that are produced over a 10-year horizon, although interest rate mismatch risk remains the predominant risk on the balance sheet. In the hypothetical examples presented, the increase in FX holdings and duration would increase the cumulative loss estimate relative to the baseline (ie status quo) scenario.

Cumulative balance sheet loss estimates under different strategies Graph 1



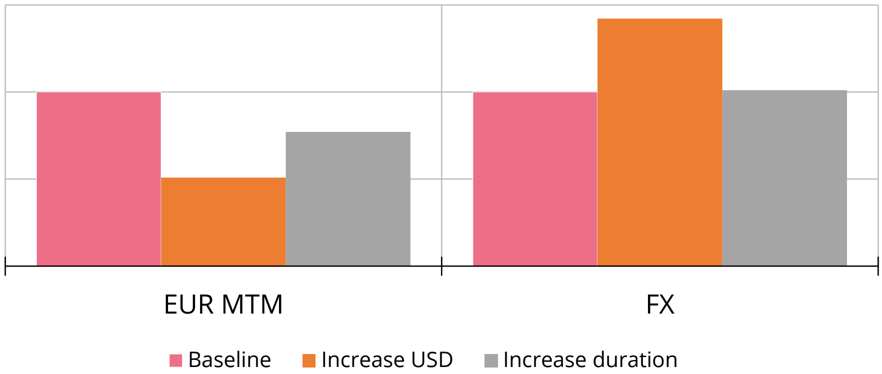
Source: authors' elaboration.

For the increase in FX, over the 10-year period overall income is expected to be higher across the scenarios on average. However, when focusing on the gross cumulative loss estimate (which includes only years where losses materialise at an overall balance sheet level), the scenarios where losses are expected due to interest overall balance sheet level), the scenarios where losses are expected due to interest rate mismatch risk, these risks can be amplified by the potential for adverse moves in foreign exchange rates. This incremental increase in risk is, however, less than the increase in standalone risk estimates shown below, illustrating the benefits of diversification.

For the increase in duration, the negative impact on the marked-to-market portfolios, through capital losses on fixed income holdings, coincides with the scenarios where an increase in interest rates leads to a materialisation of interest rate mismatch risk. It is important to note here that this analysis assumes an increase in duration over the full 10 years, with no adjustments made in response to market conditions at any one point in time. This illustrates the importance of fully considering the balance sheet perspective in respect of investment strategy, and the interaction of investment and interest rate mismatch.

As can be seen below in Graph 2, such hypothetical changes in investment strategy can increase the standalone risk estimates for a type of risk. This chart shows, for each of the two hypothetical scenarios as well as the baseline, the impact on the standalone cumulative loss estimates, for the euro marked-to-market portfolio and FX portfolios respectively. For example, the increase in FX can reduce the EUR MTM risk, but can lead to a greater exposure to adverse FX rate movements. This underscores the importance of being aware of the potential for higher volatility and potential for valuation losses, over shorter time horizons, in specific asset classes.

Standalone loss estimates under different strategies, for EUR MTM and FX portfolios respectively — Graph 2



Source: authors' elaboration.

4. Defining an investment risk appetite for a public investor

4.1 Interest rates and pricing of risk: The Minsky hypothesis

As outlined in Section 2, up until end-2021, public investors operated in an environment of historically low interest rates and record low levels of risk pricing. A combined effect of these factors created an environment for potentially hidden investment risks.

The Minsky hypothesis considers that, when markets are overly benign, risks may be underrepresented or underpriced In Danielsson et al (2018), the authors' key conclusion is that low volatility is a strong predictor of financial crises. In such a financial crisis, volatility can be expected to increase significantly, and can do so suddenly. In Bhattacharya et al (2015), the authors construct a model to demonstrate that optimism among investors and associated risk-taking can be accompanied by lower risk premia, facilitating the build-up of excessive risk, which can materialise at a later point in the cycle. In Fostel and Geanakoplos (2014), the authors describe the concept of a "leverage cycle", whereby low volatility for an extended period of time can lead to an increase in leverage, which can then increase the vulnerability of economies and markets in a downturn.

For public investors, this can mean that, if they change their asset allocations in a more benign risk environment, future risk levels could increase substantially due to materialisation of a "Minsky moment". Thus, the potential for such a development should inform risk appetite today.

In the period between the GFC and Covid-19, risk assets experienced a benign risk environment, as outlined in Section 2. Arguably, features of this Minsky hypothesis are currently materialising, as central banks increase policy rates and tighten monetary policy in response to a sustained inflationary environment and against a backdrop of elevated geopolitical risks. Since the GFC, this has led to a fall in the value of riskier assets, an overall increase in market volatility, as well as the increased potential for greater movements in asset values.

The investment risk frameworks of public investors should therefore be calibrated to look through these changes in risk conditions and capture all relevant environmental factors, as well as anticipating the effects of price normalisation on risk appetite. In this way, the risk appetite of the public investor is protected, to the extent possible, from changes in financial market regimes, and associated repercussions for risk levels, and seeks to protect against the paradigmatic changes in risk dynamics which characterise a Minsky moment.

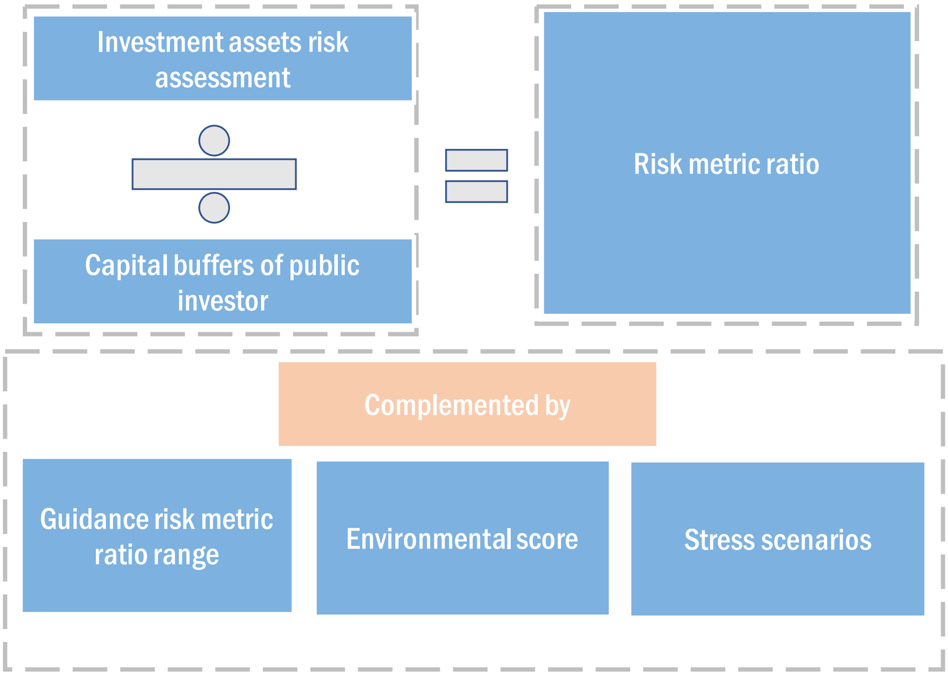
The implication of this from a risk appetite or risk-budgeting perspective is that there should not be a full utilisation of risk appetite in benign market conditions. This is to avoid the risk exposure increasing substantially above risk appetite, should the risk environment shift in an adverse manner. Conversely, when markets are under severe stress, it could be expected that the full risk appetite is utilised (thereby avoiding a harmful fire sale of assets or inefficient use of capital), as this would be considered an appropriate risk level in the context of the prevailing environment.

The CBoI has developed a framework ("risk metric framework") that seeks to incorporate measures to represent the risk-taking capacity of the investment assets. The framework takes into account risks present elsewhere on the balance sheet and the wider risk environment while also considering an element of future directional dynamics. The features of this framework may also be relevant to other public investors, with a number of parameters customisable to the specifics of the investor.

4.2 Risk metric approach – key elements

The risk metric framework consists of a number of key elements, which are described below (see also Table 1). These elements can be tailored for any public investor, according to their mandate and risk appetite.

Key elements of the risk metric approach Table 1



Source: authors' elaboration.

- **Investment assets risk assessment:** This risk assessment is based on the assessment and quantification of both market and credit risk of the applicable assets/portfolios, and the preferred CBoI risk estimate is expected shortfall at the 99th percentile, over a one-year horizon.[3] This quantification can be completed on an economic or accounting basis (eg if there are unrealised revaluation gains applicable to some assets). Under the *accounting basis*, the potential losses are estimated based on the worst simulated economic scenarios impacting the central bank's P&L over a one-year horizon *after* subtracting the applicable revaluation accounts. The economic basis does not subtract revaluation accounts from risk estimates, but includes them as part of capital buffers. The economic approach therefore reflects unrealised results on security portfolios that are revalued and on foreign currencies that would not necessarily be reflected in the P&L account.
- **Capital buffers of public investor:** This is used as a measure of the capacity of a public investor to bear risk, as these buffers could then be used to offset losses should they occur. This can reflect all financial

[3] This horizon is constructed through the "square root of time", using more frequent returns observations.

buffers (reserves and provisions) or just those have been apportioned to investment assets, on an ex ante basis.

- **Risk metric ratio:** This is computed as the first item above (investment risk assessment), divided by the second. This gives the overall risk of the investment assets, proportionate to the risk-taking capacity of the public investor.

The three elements above are also complemented by the following elements:

- **Guidance risk metric ratio range:** This would outline the range within which the risk metric ratio should sit, and can be interpreted as lower and upper bounds for a public investor's risk appetite for their investment assets. Another way of saying this is that the range captures the risk appetite in the most benign risk conditions up to the most severe. This range would also be heavily informed by other risks that are on the balance sheet ie the appropriate range would need to be assessed relative to other significant risks, such as interest rate mismatch risk. Conversely, where a public investor has large holdings of foreign reserves for exchange rate management, a higher guidance range may be necessary. As a public investor's balance sheet size and composition changes over time, this range should be re-examined to ensure it remains appropriate, and does not interfere with primary mandates and objectives (eg price stability, financial stability).
- **Environmental score:** This can be an internally constructed measure, which reflects the riskiness of the public investor investable universe, in a manner independent of actual holdings or strategy. It can also reflect the presence of geopolitical risk, through inclusion of asset classes that would not normally form part of a public investor's investable universe, but which are more sensitive to the geopolitical backdrop (eg commodities).
- **Stress scenarios:** The inclusion of prescribed market and credit shock scenarios can also help to illustrate the key risk sensitivities of the investment assets. In order to calibrate the market stress scenario, the most volatile exposures are subjected to a market shock at a certain VaR percentile level based on historical data, which obviates the need for assumptions on correlations. For the credit risk scenario, downgrades are modelled for the lowest rated issuers, as such events have the highest impact on the CBoI risk profile. These scenarios can then be approved by the governance bodies and computed on a periodic basis, for additional information to aid decision-making.

Bringing all the elements described above together, the risk metric ratio can be compared with the environmental score, at a point in time, to provide guidance to internal stakeholders as to the level of risk on the investment assets, relative to the risk environment prevailing, as well as the public investor's risk appetite. It is also possible to map associated colour coding onto directional guidance that can be provided to governance bodies, as an aid to investment asset decision-making.

Graph 3 below demonstrates all these features brought together. A risk matrix is constructed as a visualisation of the factors considered and is intended to help support discussion around the risk metric at a point in time given the prevailing risk environment and to inform investment decisions. In this example, two different portfolio compositions are compared at two different points in time as different risk environmental scores apply. Using this hypothetical example, the following conclusions can be drawn:

- Portfolio Example B is riskier than Portfolio Example A – this is evidenced by the ES99 to capital ratio (the risk metric ratio) being higher.
- However, the risk of Example B was computed in a higher risk environment, this is demonstrated through the use of the environment/risk capacity score matrix. Therefore, in a higher risk environment, it would be expected that more of the risk appetite (ie a higher risk metric ratio) is utilised, as this would be considered appropriate in the context of the prevailing environment. In this context, both Portfolio A and B correspond to the same amber colour coding on the chart, meaning there may be only limited room for tactical (as opposed to strategic) changes in risk of the portfolios. A straight horizontal range for the risk metric ratio has been applied for ease of reference for internal decision-making, however, alternative ranges can be set to exclude the possibility of reaching the red zone (as defined below) at the top of the range in a benign environment. In any event, the framework can be used flexibly with the chosen range depending on the overall balance sheet risk profile (taking into account the monetary policy related assets) and income required to meet expenses.
- The application of stress scenarios or sensitivity analysis increases the risk of these portfolios further. These measures provide further contextual and dynamic or forward-looking information on the potential for a deterioration in the credit quality or market risks of the investment exposures, and how these risks correspond to the public investor's current risk appetite. Such scenarios serve to highlight how likely the CBoI is to breach the range in the event of stressed but plausible scenarios materialising, and are particularly useful when considering alternative strategies or a variations in the investment asset composition.
- The dotted lines provide the guidance range applicable. In both these examples presented, the risk levels of the portfolios are currently below the guidance range. Once the more severe stress scenarios are applied, the portfolios would fall within the guidance range applied.
- The colour coding of both portfolios are broadly similar, meaning that the same guidance applies at this level. This is irrespective of the lower risk metric ratio in Example A and is consistent with the finding that, when market conditions are benign, risks may be underrepresented and underpriced (Danielsson and Zhou (2016), Danielsson et al (2018)). Broadly speaking, the further towards red that the risk metric ratio moves, the lower the capacity to take on furher investment risk. The mapping of colour coding onto worded guidance is not intended to act as a hard constraint (although it could be used in that way), but more as a useful aid to decision-makers when interpreting the matrix and applying it to investment decision-making. The colour coding broadly corresponds to the following guidance:

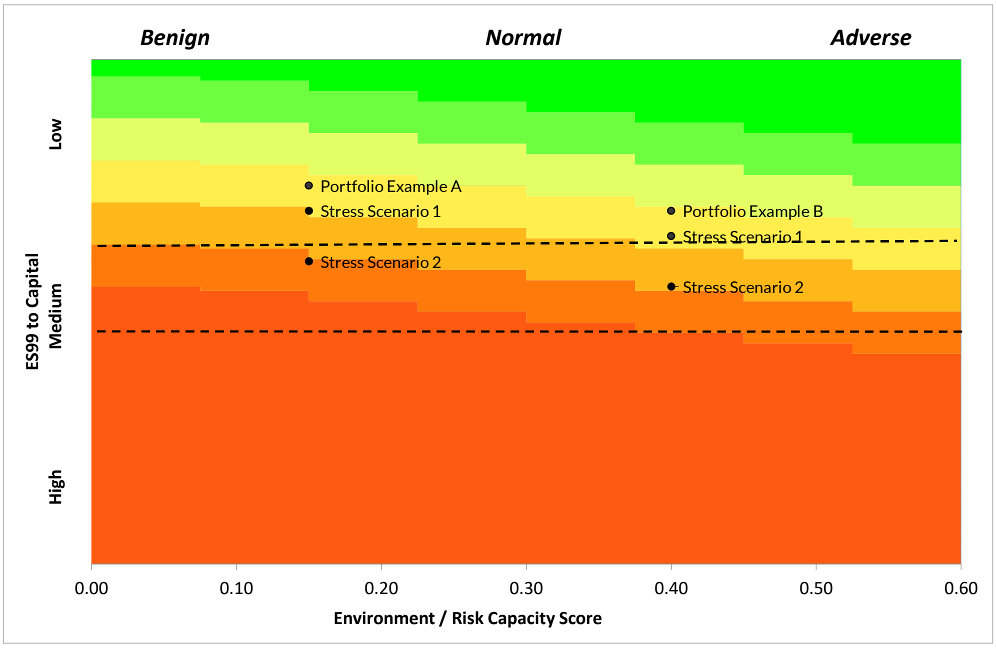
i) **Green**: additional strategic risk-taking may be considered for the investment assets. This could include additional exposure to foreign denominated fixed income holdings;

ii) **Yellow/amber**: there may be some scope for additional strategic risk-taking or adding tactical adjustments (such as increasing duration) to the investment strategy;

iii) **Red**: risk is at or above risk metric guidance, and the current levels of investment asset-related risk should be monitored closely or be formally reviewed.

If the risk metric ratio is persistently in the red zone, this could necessitate a discussion on the potential to either reduce the investment risk appetite, or to increase capital buffers, with either option potentially moving the metric out of this zone.

A risk metric assessment for two example portfolios — Graph 3



Source: authors' elaboration.

4.3 Reflections on framework in current environment

The importance of the preceding models and frameworks presented in the paper has been underpinned by the recent change in the economic and risk environment. In particular, the risk metric framework presented contributes to a more comprehensive through-the-cycle indicator of risk levels. This allows the public investor to factor in the current risk environment, which until the pandemic had been characterised by low volatility and supported by the accommodative monetary policies of many global central banks. However, the environmental score would have provided this context alongside investment decisions and led to a cautious approach in increasing the risk metric ratio in such benign conditions (thereby limiting the tendency to search for yield by lowering the minimum permissible credit threshold or diversifying outside the more traditional investment universe for public investors). Consistent with this approach, the environmental score in the more recent risk environment would tolerate a higher risk metric ratio and avoid the need to de-risk an investment strategy that had been calibrated during a period of low volatility and against a calmer geopolitical backdrop.

Recent changes in the environment raises specific challenges for central banks. Firstly, the increase in interest rates has led to the materialisation of interest rate mismatch risks for many central banks, with knock-on implications for income and a potential for financial losses. In this scenario, public investors will require the use of their financial buffers in order to absorb these losses, and this could mean that the available capital or buffers for their investments could be lower than previously anticipated. This would imply a reduction in investment risk appetite, all else being equal.

However, set against this, many public investors hold investment assets primarily to help cover operating costs (in order to pursue their respective mandates) and to produce a reasonable return. Therefore, in the coming years, there could be a natural tension between the traditional financial risks that have been outlined in this paper, with a "financial independence" risk if a public investor perceives a risk to covering its costs (Jones (2016), Doran et al (2018)). In the event that operating costs are not matched by income, due predominantly to the materialisation of interest rate mismatch risk arising from the price stability mandate, this could lead to successive years of losses that could progressively deplete financial buffers and lead to concerns about credibility and financial independence. In this case, consideration could be given to reducing the investment risk appetite or the guidance risk metric ratio range in response, as compared with other options such as rebalancing the portfolio towards safer, less volatile assets, selling selected investment assets in order to realise revaluation gains or reducing expenditure.

5 Conclusion

This paper presented two examples of approaches public investors can adopt to more effectively identify and respond to the changes in risks to their investment assets which explicitly takes into account the implications of the contemporaneous financial environment relative to the cycle. Firstly, an all-balance sheet approach to assessing investment risk was presented, in light of material non-investment risks arising from purchase programmes associated with QE. The second approach proposed a means of defining a risk appetite for a public investor, which aims to balance investment risk tolerance with changing risk environments over time, in the light of risks elsewhere on the balance sheet. It was demonstrated that the environmental score (and therefore the guidance on future risk-taking capacity) can be similar for investment assets with varying levels of risk exposure when taking into account the stage of the current market cycle, and that this can be a useful tool in avoiding procyclicality when it comes to risk-taking for public investors.

References

Albertazzi, U, B Becker and M Boucinha, (2021): "Portfolio rebalancing and the transmission of large-scale asset purchase programs: Evidence from the Euro area", *Journal of Financial Intermediation*, vol 48, no 100896.

Bhattacharya, S, C Goodhart, D Tsomocos and A Vardoulakis (2015): "A reconsideration of Minsky's financial instability hypothesis", *Journal of Money, Credit and Bankin*g, vol 47, no 5, pp 931–73.

Bua, G and P Dunne (2017): "The portfolio rebalancing effects of the ECB's asset purchase programme", SSRN 2993136.

Central Bank of Ireland (2022): *Annual Report 2021*.

Danielsson, J and C Zhou (2016): "Why risk is so hard to measure", *DNB Working Papers*, no 494.

Danielsson, J, M Valenzuela and I Zer (2018): "Learning from history: volatility and financial crises", *The Review of Financial Studies*, vol 31, no 7, pp 2774–805.

Donnery, S, D Doran, R Gleeson and K Carroll (2017): "Non-standard monetary policy measures and the balance sheets of Eurosystem central banks", *Quarterly Bulletin*, no 3, pp 87–102.

Doran, D, R Gleeson, S Kilkenny and Š Ramanauskas (2018): "Assessing the financial risks and buffers of the central bank", Central Bank of Ireland, *Quarterly Bulletin*, pp 58–71.

Doran, D, S Kilkenny, Š Ramanauskas and A Shablov (2020): "An alternative approach to measuring the liquidity risk of public investors' investment assets", *Evolving Practices in Public Investment Management*, Proceedings of the Seventh Public Investors Conference.

European Central Bank (2015): "Box: The Governing Councils expanded asset purchase programme", *Economic Bulletin*, pp 15–18.

Fostel, A and J Geanakoplos (2014): "Endogenous collateral constraints and the leverage cycle", *Annual Review of Economics*, vol 6, no 1, pp 771–99.

Jones, B (2016): "Institutionalizing countercyclical investment: a framework for long-term asset owners", *IMF Working Papers*, no WP/16/38.

Fact, fiction, and green bond investing – a central bank's perspective

Juliusz Jabłecki[1]

Abstract

Building on Narodowy Bank Polski's several years of experience in the green bond market, this article discusses a number of questions surrounding green bond investing. Trying to separate fact from fiction and concepts from misconceptions, the essay looks at the size and depth of the green bond universe, pricing patterns, risk profile and performance, as well as the potential for engineering real environmental impact.

JEL classification: G12, G14, G20.

[1] Financial Risk Management, Narodowy Bank Polski, juliusz.jablecki@nbp.pl. I wish to thank Mike McMorrow, from the BIS, for his careful reading of the chapter draft and his many insightful comments and suggestions. Any remaining errors are of course mine alone. Any statements made herein are those of the author and do not constitute the official position of NBP.

1. Introduction

Green bonds refer to any type of debt instrument where the use of proceeds is directed exclusively towards financing/refinancing environmentally sustainable projects. The asset class was born in 2007, when the European Investment Bank (EIB) issued its first "climate awareness bond", but began to gain ground in earnest several years later, when the 2015 Paris Agreement ushered in a renewed sense of urgency around transitioning to a low-emissions economic growth. Despite a certain divergence of views around the various technological aspects of achieving carbon neutrality, there appears to be a consensus that reaching net zero will in any case require massive investment, the estimated value of which ranges between $50 trillion almost $300 trillion.[2] And while a large part of the financing for all these sustainable projects will come in the form of equity (eg when a company issues new shares or retains earnings to finance the greening of its production processes), it is estimated that only about 40% of global emissions originate from listed companies (Generation IM (2021)), and therefore debt instruments do have a role to play as well – especially in view of their importance for financing government expenditures. Thus, green bonds, along with their slightly more general cousins "sustainable" bonds,[3] have become a much sought-after product among investors – both public and private – seeking to incorporate climate change objectives in their portfolios, either on their own or as part of a broader ESG strategy.[4]

Narodowy Bank Polski (NBP), Poland's central bank, has been an increasingly active player in the green bond market. Entrusted with managing the country's FX reserves, NBP actively manages a portfolio worth over $100 billion (of the total official reserve assets exceeding $160 billion), investing in predominantly fixed income instruments across eight different currencies, including USD, EUR, GBP, CAD, AUD, NOK and NZD. Beginning with small-scale purchases of labelled bonds in 2018, allocation to the asset class across all currency portfolios grew to about $600 million by 2021. By that time, the Management Board adopted a formal Green Bond Strategy, aiming to provide some structure to the previously somewhat idiosyncratic investment process, and providing allocation targets along with a time frame for achieving them. With close to $900 million invested in green bonds across a range of markets and issuers, NBP can claim to be a significant player in this burgeoning market.

However, the gradual expansion of our exposure has been a learning process, during which we have developed a better understanding of the market itself and

[2] For example, Morgan Stanley (2019) estimates that $50 trillion will need to be invested in new technologies over the next 30 years to reach net zero. A similar number is reported by Oliver Wyman and the World Economic Forum, while the consultancy McKinsey, building on hypothetical scenarios developed by the Network for Greening the Financial System, estimates that spending on physical assets would need to reach about $275 trillion by 2050, or $9.2 trillion per year on average, to achieve net zero.

[3] Whereas green bonds fund strictly designated environmental projects, sustainability bond may fund a mix of environmental and social projects, while sustainability-linked bonds do not fund particular projects but their coupon or principal step up if the issuer fails to meet the pre-agreed environmental or social targets (Kini et al (2021)).

[4] According to Bloomberg data, as of February 2022, there were 100 sustainable fixed income ETFs with total AUM exceeding $50 billion (14 of which were dedicated green bond ETFs with $2 billion in AUM). Owing to the inflow of funds into these vehicles, AUM has increased by almost 40% since the beginning of 2021.

gained a clearer picture of the rationale behind green bond investing and its desired (and sometimes undesired) outcomes. As it would be difficult to distil all the internal discussions and memos into a coherent "lessons learned" narrative, this article opts for a more modest, and perhaps also more entertaining approach – it attempts to retell our educational experience by reviewing the following common questions about green bond investing, framed here in a purposefully provocative and controversial way:

1. **Is the green bond market too small and underdeveloped?**

2. **Does establishing a green bond mandate mean sacrificing returns and should it be a concern?**

3. **Are green bonds "safer" than conventional bonds and do they outperform in risk-off episodes?**

4. **Is reputational risk an important issue in the green bond market?**

5. **Does green bond investing make a positive environmental impact?**

The questions are presented without specific attribution, as the intention is not to engage in polemics, but rather to present important considerations regarding the green bond investment process in an intuitive and entertaining way. That said, each question will be addressed and examined through the prism of relevant up-to-date practitioner and academic literature as well as empirical data and examples.

Readers will doubtless find that the above list is by no means exhaustive and the answers provided – most of which go along the lines of "it's complicated…" – can hardly claim originality. Still, the narrative will hopefully shed some light on important considerations in green bond investing and might help to inform the decision-making of those central banks and public investors who have yet to gain a foothold in the green bond market.

2. Perspective on market structure: is the green bond market too small and underdeveloped?

One of the first issues to arise in discussions around green bonds is often the size, liquidity and development of the market. A natural concern for a public investor with a mandate that implicitly or explicitly requires a large and liquid portfolio to be held might be that the green bond market is still too niche and immature to be tapped in meaningful size, and thus would be best left to "specialised" funds and other players.

Is the green bond market really too small to bother with? Straddling the continuum between "fact" and "fiction", an honest response to such a question should probably be: it's complicated. Before getting a little deeper, it might be useful to reiterate that green bonds are "use of proceeds" debt instruments which fund (but are not secured against!) strictly designated environmental projects. Typically, the designation is performed by the issuer in the bond prospectus, and clarifies how the earmarking of proceeds is going to work, which specific projects are to receive the funding, and potentially even what impact they are expected to generate (see Graph 2.1 for a snapshot example of a prospectus). A watershed moment in the evolution of the green bond market was the publication in 2017 of the ICMA Green Bond Principles,[5] which sought to provide more transparency for investors and clarify requirements for issuers, pertaining to eligible project types (eg renewable energy, energy efficiency or clean transportation) and best practices with respect to project evaluation and selection, management of proceeds, and reporting.

Sample use of proceeds section of the prospectus of a green bond issue by the energy company E.ON — Graph 2.1

Reasons for the offer and use of proceeds
Gründe für das Angebot und Verwendung der Erträge

An amount equivalent to the net proceeds of the Green Senior Unsecured Bonds ("Green Bond") will be used exclusively to finance Eligible Green Projects, including related partnerships and joint ventures, in the following eligible categories, together forming the Eligible Green Project Portfolio, as set out in the Issuer's Green Bond Framework dated April 2019 (https://www.eon.com/content/dam/eon/eon-com/investors/bond/E.ON_Green_Bond_Framework.pdf). Pending the full allocation to the Eligible Green Project Portfolio, E.ON will hold and / or invest the balance of net proceeds not yet allocated, at its own discretion, in its treasury liquidity portfolio in accordance with the provisions of the Framework.

Until the maturity of the Notes, in case of divestment or cancellation of an allocated Eligible Green Project, or if an allocated project no longer meets the eligibility criteria, the Issuer commits to reallocate the proceeds to other Eligible Green Projects depending on availability.

Eligible Green Projects include projects or assets in the following eligible categories:

- Renewable Energy: a) Investments and / or expenditures to directly connect renewable energy production and storage units to the grid (including powerlines and related infrastructure such as substations) and b) Investments in or expenditures for the acquisition, conception, construction, development and installation as well of re-powering of renewable energy production and storage units (including wind, solar (PV), biomass / biomethane and power-to-x)
- Energy Efficiency: Investments for energy efficient replacements in the grid including investments and / or expenditures to increase the flexibility and technical availability of the grid in the context of fluctuating feed-in of renewables incl. Energiewende / smart grid investments, investments to decrease / minimize grid losses and energy efficient street lighting with LED, Smart meters and Integrated on-site business and city energy solutions
- Clean Transportation: Investments in development and construction of electric vehicle charging stations and related infrastructure

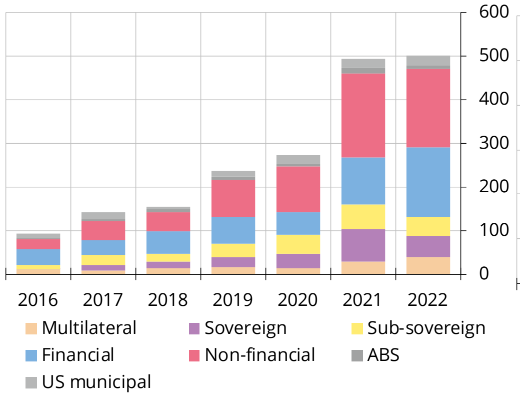
Source: Company website.

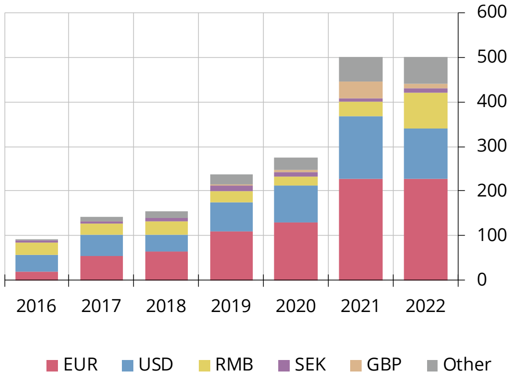
As of January 2023, the green bond market stands at over $1.8 trillion, ie roughly equivalent to the total stock of UK government gilts. And while this might correspond to just 15% or so of total global foreign exchange reserves, the market's recent growth has been spectacular, with issuance levels of about $500 billion in 2021–22, up from less than $50 billion in 2014 (Graph 2.2).

On the back of increasing volumes, the market has steadily gained in diversity along both the currency and issuer dimensions. Initially, the market was almost entirely the domain of supranational institutions such as EIB and the World Bank. Gradually, municipalities, local governments and government agencies joined in, and finally in 2013 – with the total market size still at roughly $10 billion – the first corporate green bonds appeared. Importantly for public investors, almost 80% of the outstanding stock originates in developed markets, and 75% is denominated in EUR and USD, ie the currencies preferred by central bank reserve managers.

[5] The 2021 version of the principles is available at: Green-Bond-Principles-June-2021-140621.pdf (icmagroup.org).

Green bond issuance by issuer type and currency (2016–21, USD bn)	Graph 2.2

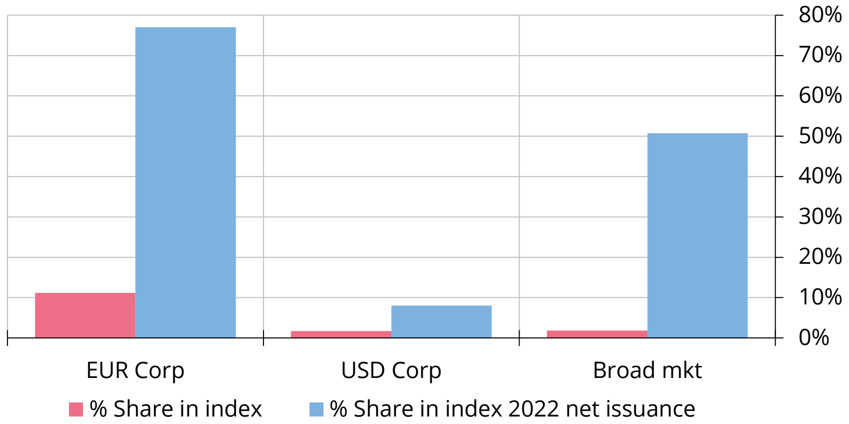




Source: HSBC, Bloomberg.

The growing importance of green bonds in the broader fixed income universe is well illustrated by their increasing share in the broadly followed benchmark indices – especially within the corporate sector. Thus, for example, of the 156 new issues added to the EUR investment grade corporate bond index in 2022, as many as 138 (or 77% in face-value terms) were green-labelled, which translates into a share of just above 11% in the index. The corresponding numbers for the – much larger – USD sleeve of the IG corp index are lower, accounting for 8% of the 12-month "flow" and about 2% of the "stock". Likewise, for the broad market benchmark, about 50% of new issuance came in the form of labelled bonds, which as of end-2022 make up 2% of the index (Graph 2.3).

Green bonds as a share of corporate and broad market bond indexes	Graph 2.3



Bond indexes are ICE/BAML benchmark corporate bond indexes for the United States and euro area: (i) US Corporate Index (C0A0), and (ii) Euro Corporate Index (ER00), and the global broad market benchmark (ICE BofA Global Broad Market Index, GBMI.

Source: ICE, Bloomberg.

The increasing depth and breadth of the green bond market – as manifested in index coverage – improve liquidity as bonds are turned over more frequently as part of portfolio maintenance, in response to index/benchmark rebalancing etc. For example, a recent HSBC study, based on an analysis of the TRAX reporting service, which covers about 50% of all European bond trades, finds that EUR labelled bonds have similar, reported weekly trading volume (as a percentage of nominal) than non-labelled peers (both numbers are around 2% on average throughout 2022; Kini et al (2023)). This seems to be at least partly borne out in trading conditions, as the median bid-ask spread-to-price ratio for the two universes is virtually the same at 48 bp.

Thanks to these structural improvements in market breadth it has become possible to replicate broad fixed income benchmarks using green bonds only. While this necessarily implies some sampling or optimisation, the results are quite encouraging. To appreciate this, consider a simple experiment, whereby each month, the available universe of green corporate bonds, say the almost 500 EUR issues and 375 USD bonds as of March 2023, is used to build portfolios in each currency such as to match the main characteristics of the broad investment grade corporate benchmark indexes, which for the sake of this exercise are the ICE/BAML Euro Corporate Index (ticker ER00) for the euro-denominated corporate bonds and the ICE/BAML US Corporate Master Index (ticker C0A0) for the USD-denominated corporate bonds. As of March 2023, the C0A0 comprised 9,831 securities with a market value of $7.7 trillion. At the same time, ER00 had 4,114 issues with a market cap of EUR 2.6 trillion.

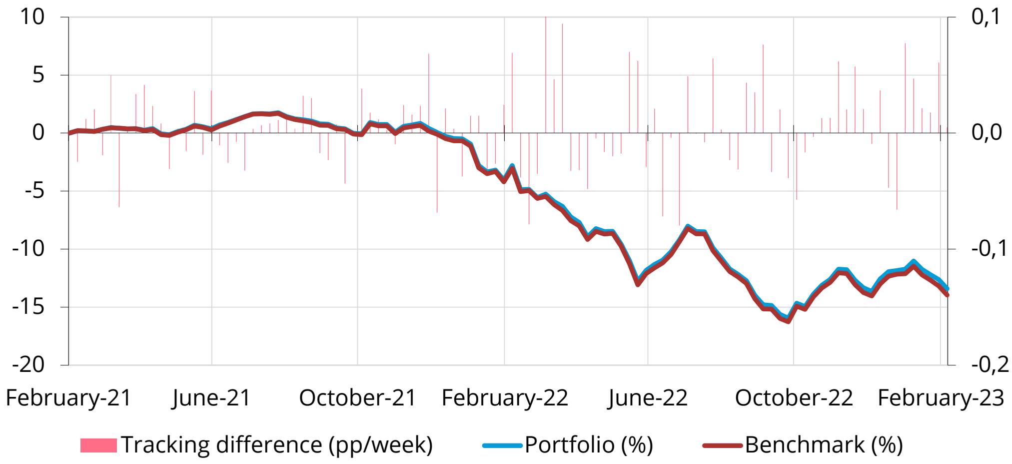
Since the universe of green bonds is evidently much narrower than either index, the replication process needs to be somewhat selective and in this case consists in matching benchmark key rate durations, as well as its overall spread duration, weights of the major rating buckets (AAA, AA, A and BBB), and duration times spread contribution of each of the three key sectors (Financials, Utilities, Industrials). Finally, while the bonds aren't screened for liquidity characteristics, the algorithm imposes a concentration limit of 2% on each issue. The sample used in this exercise runs from January 2021 through February 2023, a period marked by exceptionally high volatility in fixed income markets.[6]

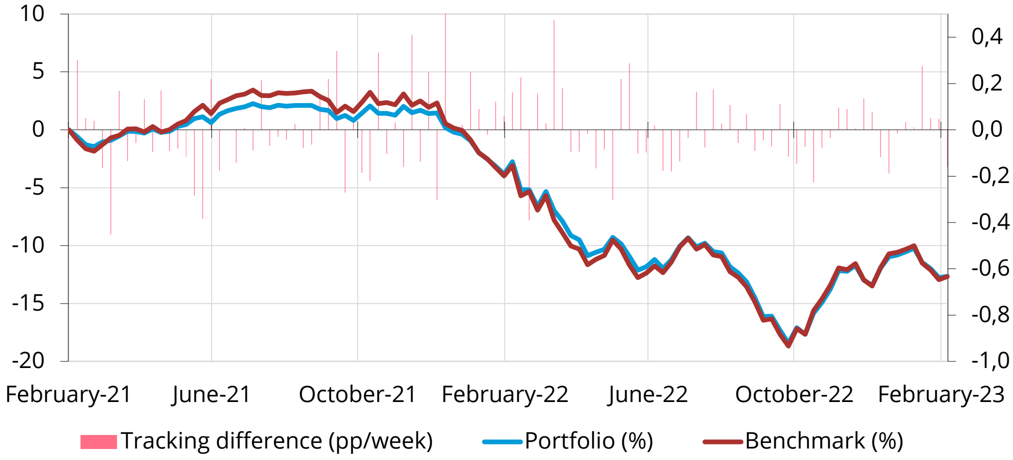
Graph 2.4 illustrates the cumulative performance of the replicating portfolios against their respective benchmarks. In both cases the fit is very good, and the cumulative performance virtually undistinguishable. The total return of the euro replicating portfolio – which held on average about 120 issues – was 13.41% against –13.96% on the ER00 benchmark, with a tracking error of just 4 bp/week. The corresponding figures for the US were –12.65% on the portfolio vs. –12.66% on the C0A0 benchmark index, with a tracking error of 17 bp/week, and an average number of 76 positions in the portfolio. The noticeably larger tracking error for the US portfolio partially stems from the lower breadth and depth of the green bond market in USD (note that there are about half as many USD green corporate issues as EUR

[6] Formally, the optimisation problem is cast as follows: Maximise $\sum_{i=1}^{N} w_i \, OAS_i$ subject to: (i) $\forall i \; 0 \leq w_i \leq 0.02$ (long-only portfolio with concentration limit set to 2%); (ii) $\sum_{i=1}^{N} |(w_i - b_i) \times OASD_i| < 0.2$ (portfolio spread duration constraint); (iii) $\sum_{i=1}^{N} \left|(w_i - b_i)KRD_{i,j}\right| < 0.1$, where $KRD_{i,j}$ is the key rate (partial) duration of i-th position with respect to j-the rate, j=1Y,2Y,3Y,5Y,7Y,10Y,20Y and 30Y (key rate duration constraint); (iv) $\sum_{i \in Rating} |w_i - b_i| < 0.02$, where Rating stands for rating bucket AAA,AA, A, and BBB (rating constraint); $\sum_{i \in Sector} |w_i - b_i| < 10$, where Sector stands for the following sector groups: Industrials, Financials, Utility (sector DTS constraint).

green bonds) and also greater market volatility on the back of the significant monetary policy tightening in 2021–22.

Cumulative performance of EUR (top panel) and USD (bottom panel) corporate green replicating portfolios — Graph 2.4





Note: Benchmark is the ICE BAML Euro and USD Corporate index respectively; Portfolio is a replicating portfolio of green corporate bonds optimised to have minimum ex ante tracking error while matching key rate durations, as well as overall benchmark spread duration, weights for the major rating buckets (AAA, AA, A, and BBB), and duration times spread contribution of each of the three key sectors (Financials, Utilities, Industrials). Replicating portfolio is rebalanced monthly.

Source: ICE, Bloomberg.

3. Perspective on the safety and return characteristics of green bonds

3.1 Does establishing a green bond mandate mean sacrificing returns and should it be a concern?

Unlike fund managers, public investors may not have clearly defined fiduciary duties to uphold, and for some of them – like central bank reserve managers – return maximisation is not even typically a first-order priority. Still, being accountable to the government and the broader society means that they are unlikely to be able to avoid risk-return considerations altogether in developing their investment process. And thus, one of the key questions involved in setting up a green bond portfolio or strategy is likely to be about the financial impact of such a step. Framing this a little more formally, the question is whether, and to what extent, green bonds trade at a premium – astutely called the greenium – relative to conventional counterparts with the same risk profile, and whether that premium then feeds through to returns.

As unsecured instruments, green issues have in principle indistinguishable risk characteristics from conventional bonds of the same issuers. After all, even though the proceeds from the issuance of green bonds are earmarked to particular environmentally friendly projects, their cashflows are ultimately serviced through income from the entirety of the issuer's operations (or tax base, in the case of sovereigns), not just the particular green project. Hence, a priori, there should be little systematic difference in pricing between green and non-green bonds with the same financial features, particularly so if both debt instruments are issued by the same entity. However, it could also be argued, that the green label does in fact impact pricing (up or down) via two distinct channels: (i) higher demand from environmentally conscious investors eager to demonstrate alignment with emerging best practice or regulations (eg Article 8 of the EU's Sustainable Finance Disclosure Regulation); and (ii) higher costs related to additional tracking, monitoring, and reporting processes, as well as up-front investment to define the bond's green criteria and sustainability objectives – all of which could be passed on to investors (eg EIB estimates that the additional cost of issuing its green bonds in terms of dedicated staff, IT systems etc to be at 0.02% of issue size; EIB (2021)). To the extent that this effect (i) outweighs (ii), a green bond may be expected to trade at a premium (greenium) vis-à-vis its non-green counterpart.

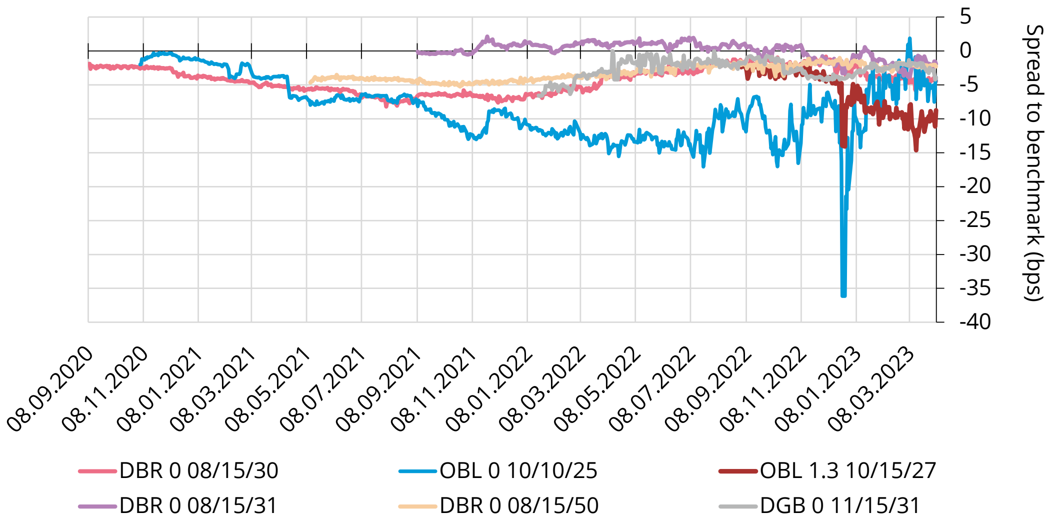
Although there is by now a considerable empirical literature studying the existence of the greenium, the evidence is generally inconclusive and consensus on its size and even sign is yet to emerge (see MacAskill et al (2021) for a comprehensive review of the literature on greenium). Somewhat surprisingly perhaps, there does not even appear to be a consensus on how best to measure the greenium, and the approaches in the academic and practitioner literature differ depending on context. The most straightforward approach involves comparing the spread of a given green bond relative to some benchmark curve vs its non-labelled counterpart, whereby a tighter spread of the green bond (ie a negative spread differential) means it trades richer (is more expensive) than the non-green peer, suggesting a potential greenium.[7] In practice, it is not always straightforward to define a "conventional counterpart" with

[7] We follow the convention that seems to have developed in the literature of speaking of a positive greenium when the spread between green and non-green is negative, and vice versa.

matching maturity – after all, the peer bond should in principle differ from the labelled issue in nothing but its use of proceeds.

One example where this type of approach might be applied is in the context of so called twin bonds issued originally in 2020 by the German federal government, and more recently also by Denmark. The idea consists in always issuing a green bond alongside a conventional one with the same maturity, coupon and even interest dates to facilitate comparison and facilitate price discovery. Although the twins still differ markedly in issuance amounts (eg the amount outstanding of the German green bond maturing in 2030 is EUR 9.5 billion, less than 25% of the amount of its non-labelled twin), this is perhaps as close as one can practically get to a like-for-like comparison. The overall pattern is that green bonds tend to trade slightly richer (with lower yields) than non-green counterparts and the resulting greenium averages about 4 bp for the German issues and about 2 bp for the Danish one (Graph 3.1). Still, the pattern isn't clear or consistent and seems to be at least partly influenced by technical factors, as the greenium is actually largest for the shortest-maturity German bond (OBL 10/10/25) and least pronounced for the longest-maturity one.[8]

Greenium estimates for German and Danish green government bonds Graph 3.1



Note: DBR and OBL denote the German issues while DGB stands for the Danish issue; greenium calculated as the mid-yield spread to the non-green twin benchmark.

Source: Bloomberg.

Even when finding a perfect twin bond is impossible, the greenium can still be calculated as the spread gap between an issuer's green bond and a maturity-matched non-green peer. Granted, some entities are not big or active enough in the market

[8] The technical factors might refer to investors' preferred habitats and market segmentation. As of this writing (May 2023), there are 23 sovereign bond issues in the euro market with an average maturity (weighted by amount outstanding) of about 14 years. This makes the German note maturing in 2025 the only viable investment on the short end of the curve, which may result in additional demand particularly from institutional/public investors constrained in the amount of duration risk they can take.

(in particular in Europe where corporations have traditionally relied to a greater extent on bank credit than on the bond market) as to have a fully developed spread curve that would facilitate such comparisons. In such cases one can construct a theoretical synthetic duration-matched counterpart either by forming a barbell portfolio of two bonds with durations above and below that of the green bond, or via an interpolation/regression scheme or even a fully fledged yield curve model to supplement the missing points on the curve. Finally, an even cruder approach consists in analysing bond spreads for entire groups of issuers belonging to the same category and/or risk bucket. Thus, for example, one could filter out all GBP issues of AA companies, select those labelled as green, and compare their median spread with the one calculated for the rest of the group.

Attempting to sidestep these methodological problems, Jabłecki (2022) looks for evidence of a greenium in EUR and USD investment grade corporate bond universes by running a series of cross-sectional regressions of bond spreads (option-adjusted, or OAS[9]) on a range of explanatory variables including a dummy for "greenness," to check whether it is possible to attribute any part of corporate bond's spread to the green label itself. The regression results are reproduced in Table 3.1. They point to a clear, statistically significant spread greenium in both markets of about 5–8 bp on average, which over the long run should marginally erode the carry on the portfolio (although in , short run, it might still have little impact on returns).[10]

Fama-MacBeth credit spread regressions — Table 3.1

	EUR IG		US IG	
	Coeff	t-stat	Coeff	t-stat
Green	-8.00	-10.86	-4.85	-4.46
Spread duration	6.66	24.76	6.58	20.42
Rating	17.22	26.2	21.20	13.48
Industrials	-59.24	-17.77	-81.78	-20.58
Financial	-27.77	-11.93	-67.40	-21.39
Utility	-70.81	-17.59	-83.44	-16.37
R2	0.89		0.87	

Note: EUR IG includes all euro-denominated investment grade corporate bonds within the ICE/BAML Euro Corporate Index (ticker ER00); US IG includes all USD-denominated investment grade corporate bonds within the ICE/BAML US Corporate Master Index (ticker C0A0). The EUR sample covers 60 end-of-month observations for 2017–21; the USD sample covers 36 end-of-month observations for 2019–21. The average number of bonds is 3,159 for the EUR index and 8,447 for USD. The dependent variable in each case is the bond option-adjusted spread (OAS). Rating is index rating according to ICE (2018).

Source: Jabłecki (2022).

Although perhaps unwelcome from a narrow investment perspective, the latter conclusion actually attests to the meaningfulness of green bonds as a vehicle of

[9] The option-adjusted spread (OAS) is the number of basis points that needs to be added to the government spot curve so that the present valued of the bond's discounted cashflows matches the traded market price (accounting for any embedded options).

[10] Interestingly, a slightly modified set of regressions for bond excess returns (over duration-matched Treasuries), suggests that the green label does not affect excess returns once typical measures of systematic credit risk are taken into account.

change – a topic addressed more directly in Section 4, and apparently gaining in importance on public investors' agenda. Indeed, survey results indicate that one of the main drivers behind implementing green bond portfolios in central banks is the desire to foster long-term sustainable economic growth (Fender et al (2020)). As desirable as this goal sounds, it should be accompanied by a realisation that there are only two channels of influence that can be ultimately traced to specific portfolio choices: (1) control through voting rights; and (2) affecting financing costs which are a key input into any company's, agency's and even government's strategic planning (Jones et al (2023)). Although most of the literature and discussion on impact investing has tended to focus on the former, the latter channel can be just as important – and clearly more relevant for green bond investors.

So how can bond investors – public or otherwise – try to promote sustainable economic growth? On a conceptual level, this would require a greater number of environmentally friendly projects to be promoted, by making them more affordable relative to the unsustainable, wasteful and polluting ones. In practical terms such a course of action entails according a higher price to the debt used to fund sustainable projects – either on the primary or secondary market – ie paying a greenium, or equivalently, accepting a lower spread on a labelled bond relative to non-labelled counterpart.

This need not be a mere signalling effect, though. After all, a greenium may work to the investor's detriment, but it is advantageous for the issuer who must compare the cost of funding with the internal rate of return (IRR) on the projects pursued. Absent a greenium, some projects may not be viable, as their discounted future cashflows would fall short of initial outlays. Of course, a greenium of 1–2 bp is unlikely to be economically significant enough to fundamentally change the budgeting picture. However, owing to the power of compounding, a 50 bp discount rate differential (corresponding to levels seen eg in segments of the HY corporate market) translates into a 5% NPV difference for a cashflow 10 years out, which can already become significant. And although a 1–2 bp greenium may have a more limited mechanical impact on the cost of financing and a project's NPV, it is certainly enough to signal investor preferences, and as such is likely to affect issuers' behaviour.

Thus, even if investors in green bonds sacrifice carry and fail to pocket a return premium, they should find at least some consolation in the fact that the underperformance of green bonds (relative to conventional counterparts) is a sign that the pursuit of environmentally friendly projects is less costly on a relative basis and thus an encouragement for issuers to undertake more of them. Given the enormous scope of investment necessary to reach net zero (see figures referenced in the introduction), such a price incentive may be a small, but important contribution to the greater cause. Virtue is its own reward, as the adage goes.

3.2 Are green bonds "safer" than normal bonds and do they outperform in risk-off episodes?

With safety of investments high on public investors' list of priorities, it is natural to ask how well green bonds fare on that count. As stressed above, since green bonds are serviced from the entirety of an issuer's operations – not merely the projects they were meant to fund – their credit risk characteristics are identical to those of the other, conventional (unsecured) bonds of that issuer. However, although green bonds *individually* may not offer a particularly attractive credit profile, they might still do so viewed *collectively*, as a subset of the broader fixed income universe.

Unfortunately, this does not appear to be the case. Consider first the broadly followed Bloomberg Global Aggregate Index, which includes government, government-related, corporate and securitised debt from a multitude of local currency markets, both developed and emerging, and by design aims to represent the global investable investment grade fixed income universe. As of March 2023, the outstanding amount of the roughly 30,000 issues covered by the index amounted to almost $66 trillion, 22% of which were rated AAA, which should come as no surprise given that virtually two thirds of the index comprises government and government-related bonds. Sifting through the broad index for green issues produces a subset of roughly 1,100 bonds with an outstanding of $1.1 trillion (ie 1.6% of the total). Yet owing to the dominant role of corporate bonds in the labelled subsample, the share of AAA-rated issues is just below 15%. What the green universe lacks in the highest-quality names, it more than makes up for in the broad AA bucket, which accounts for 11% of the outstanding, relative to just above 5% in the total. Still, the cumulative share of bonds rated AA– or above is a good percentage point higher in the broad Global Aggregate index and its green subset. The situation does not change materially when – instead of the broad market index – we consider its corporate subsets. Here, again, the broad market features a higher share of AAA-rated issues, in both the USD and EUR, than the labelled sleeve, while the share of securities rated AA– or above is roughly similar across the board.

Cumulative performance of the US green corporate bond index vs US Aggregate Corporate benchmark (2020–22) Graph 3.2

Source: Bloomberg.

Against such a background, it would be difficult to expect that green bonds *as an asset class* should outperform their conventional peers in periods of market stress. Confusingly, such outperformance is sometimes reported based on index-level statistics, which – while accurate – may mask significant compositional differences between the respective indices. Consider, for example, the subset of green bonds in the corporate sleeve of the US Aggregate index. A simple total return analysis reveals that they easily outperformed the broader benchmark both in 2020 and – even more significantly – in the volatile 2022, by a wide margin at 159 bp (Graph 3.2). However, this was largely the result of a markedly shorter duration of the labelled bonds (6 vs 7). More generally, differences in sector weights and ratings can also impact index-level performance. Once these were methodically controlled for in a regression, the role of greenness during the Covid sell-off turned out to be actually detrimental to corporate bond returns – although the results were only weakly statistically significant at best (see Table 3.2 borrowed from Jabłecki (2022), which reports the results of excess return regressions for EUR and USD corporate bonds controlling for greenness, ratings, duration and sectors).

Cross-sectional regressions of bond excess returns in March 2020 Table 3.2

	EUR IG		US IG		Global	
	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat
Green	-0.26	-1.74	-0.28	-0.52	1.40	1.36
OAS	-0.02	-72.86	-0.03	-110.86	-0.02	-74.17
Spread duration	-0.58	-62.12	-1.01	-100.82	-0.54	-56.83
Rating	0.18	8.62	-0.16	-5.50	0.18	9.80
Industrials	1.60	10.38	5.95	26.06		
Financial	1.81	12.13	6.00	27.23		
Utility	2.08	11.10	5.59	22.15		
Corporate					-1.98	-15.79
Sovereign					3.91	17.86
R2	0.92		0.91		0.78	

Note: EUR IG includes all euro-denominated investment grade corporate bonds within the ICE/BAML Euro Corporate Index (ticker ER00); US IG includes all USD-denominated investment grade corporate bonds within the ICE/BAML US Corporate Master Index (ticker C0A0); Global includes all FX-G10 investment grade (sovereign, corporate and quasi-government) issues within the ICE BofA Global Broad Market Index (ticker GBMI). The dependent variable in each set of regressions is excess return over a synthetic duration-matched treasury security.

Source: Jabłecki (2022).

The results reported above suggest that green bonds do not – in general – offer shelter from market volatility, at least when compared with non-labelled peers on a like-for-like basis. However, perhaps they can play some role in hedging the more fundamental, climate-related risk drivers which are not immediately priced into short-term market moves? The underlying reasoning would be that, if and when environmental risks do materialise, non-labelled bonds issued by "brown" companies may be subject to more significant adverse valuation changes. However, as pointed out by Ehlers and Packer (2017), this need not be the case.

First, and most importantly, just because a bond funds an environmentally friendly project that moves a company – or even an entire country – closer to carbon neutrality, this does not mean that the project itself is protected from climate-related risks. Most hydro plants in the world are subject to either flood or drought, wind farms are exposed to storms and other extreme weather events while solar panels can be damaged by hurricane hailing and flooding. All of these weather events are likely to become more severe and frequent as the climate changes, exposing companies, especially in the utilities sector, to losses and affecting the income stream on the green bonds funding the respective projects.

A second, somewhat more nuanced point is that, even though issuing green bonds might reflect a company's ambition in reducing carbon emissions,[11] carbon intensity may not go down fast enough in some sectors to shield the business from higher carbon costs (which could materialise through carbon taxes or due to reducing allowances in cap and trade schemes such as the EU's ETS). Investigating a number of hard-to-abate sectors, like cement, chemicals and airlines, White et al (2022) found no correlation between a good emissions reduction strategy and the prospective credit profile. This is particularly acute in the highly carbon-intensive cement sector, where the carbon cost will likely exceed a significant percentage of revenue by 2030, eating into margins and threatening the cash flow that supports interest and debt repayments.

4. Perspective on reputational risk, greenwashing and impact

4.1 Is reputational risk an important issue in the green bond market?

Structurally green bonds are identical to conventional bonds in terms of seniority and cashflows, and thus they neither meaningfully diversify nor expose holders to any additional sources of systematic credit risk. However, green bond investing may be associated with reputational risk in a way that conventional fixed income is not. The paradox lies in the fact that both investing and abstaining from investment may be – depending on context – harmful to a public investor's reputation.

The perils of investing in green bonds come primarily in the form of "greenwashing" concerns. According to the European Securities and Markets Authority (ESMA) – which made tackling greenwashing one of the priorities of its sustainable finance roadmap[12] – the term "refers to market practices, both intentional and unintentional, whereby the publicly disclosed sustainability profile of an issuer and the characteristics and/or objectives of a financial instrument or a financial product either by action or omission do not properly reflect the underlying sustainability risks and impacts associated to that issuer, financial instrument or financial product". Public investors who, owing to their unique charters, at times involving market oversight responsibilities, tend to view their fiduciary duty to stakeholders in the government and the broader society particularly seriously. That is why some are understandably wary of investing in a market still struggling with transparency and information asymmetry.

Just how topical the issue is can be illustrated by a simple search of the Financial Times archive, which yields 248 articles mentioning "greenwashing" in 2022. That's

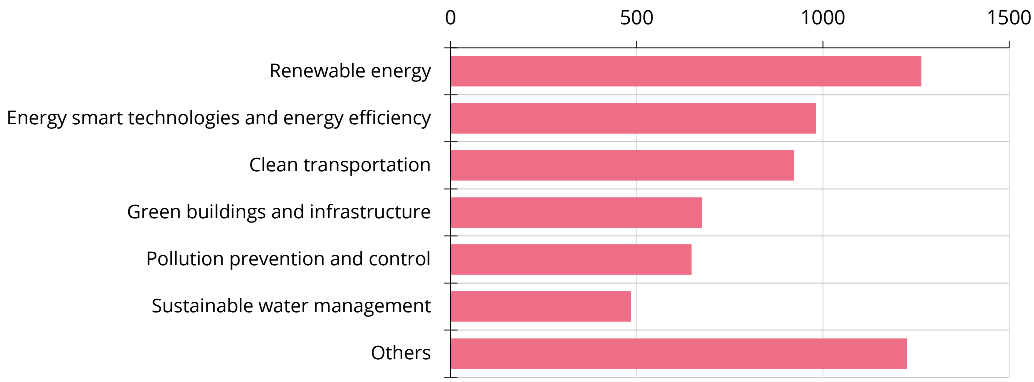
[11] Ehlers et al (2020) show this is actually debatable, as they are unable to find strong evidence that green bond issuance is associated with any reduction in carbon intensities at the firm level.

[12] www.esma.europa.eu/sites/default/files/library/esma30-379-1051_sustainable_finance_roadmap.pdf

roughly an article per business day, and if one were to include all other content, ie videos, podcasts and blog posts the number of references would increase fourfold.

In theory, green bonds are structured in a way that should ensure some form of accountability. Recall, that the "use of proceeds" section in the prospectus lays out specifically which projects a particular issue is going to fund – most of which focus on valid and important climate friendly areas such as renewable energy, clean transportation, pollution prevention and water management (Graph 4.1).

Green bond projects per use of proceeds (since inception; USD billions) Graph 4.1



Source: Bloomberg NEF.

At the same time new legislative initiatives and taxonomies – such as the original ICMA, CBI as well as national and regional standards – all strive to alleviate some of the risks inherent in the self-labelling process, bringing in much-needed clarity and uniformity to the process of selecting eligible projects, the management of proceeds (tracking spending), and post-issuance reporting. Moreover, specialised institutions – such as Sustainalytics, DNG-GL and Vigeo Eiris – offer third-party verification of the use-of-proceeds, which is becoming increasingly popular and reduces the risks of misrepresentation of the use of funds or sustainability of the financed projects.

However, there is also no escaping the fact that the green label is, by design, rather binary –a bond either is or isn't green – and not only is there no distinction between the validity and "greenness" of the underlying project but, more importantly, the green bond itself focuses on the specific projects referenced in the prospectus and not on the issuer's broader operations. In principle, this latter feature can be a virtue, as it allows predominantly "brown" issuers to launch their transition towards carbon neutrality. And indeed, if a company has a poor track record on climate but wishes to change its business profile and use green bond proceeds to do so, the there is no reason to penalise such behaviour. Unfortunately, however, raising funds under the guise of a green label could be a form of a box-ticking exercise, without any intention of making a meaningful change. There is a difference between an energy company which uses green bonds to diversify away from coal and towards renewable sources and one which uses the funding to install wind turbines on its oil platform or solar-powered flaring valves. In the latter case, issuing green bonds could simply be a way of acquiring relatively cheap funding to support current business, and not necessarily part of a broader, material and intentional transition strategy, thus deserving to be called out as greenwashing.

A practical example to that effect is provided by the case of green bond issuer State Bank of India, the country's biggest lender, which amid some controversy extended a credit facility to the Carmichael coal mine project in Queensland, Australia. The controversy lies in the fact that, although the loan looked similar in size to SBI's green funding, it was estimated to have a CO_2 footprint roughly 20 times higher than what was being saved through the green projects. Thus, although SBI's green bonds were financing (actually – refinancing) legitimate green projects in line with the bank's overall sustainability strategy, their positive impact was being negated (many times over) by the funding provided to the Australian mine. According to the Anthropocene Fixed Income Institute, which published estimates of the carbon intensity inherent in the Carmichael credit facility, such behaviour on the part of SBI was an instance of greenwashing in the sense that it demonstrated SBI's lack of true dedication to sustainable activities and climate transition in general (Erlandsson (2020)).

Given that the investor's stated goal behind investing in green bonds is likely to be to contribute to funding the transition to a net zero economy, it can be problematic to finance notionally green projects that do not change the overall carbon profile of the issuer. Hence, managing the reputational risk involved in investing in labelled bonds might require a more thorough assessment of the issuer's entire operations, not just verifying the alignment of the use-of-proceeds with a chosen set of principles. Mindful of this, some market participants and asset managers have proposed frameworks to screen, assess and compare green bonds so as to reduce the potential risk of greenwashing. Although approaches differ in their specific methodologies, they often rely on issuer-level ESG scores assigned by external providers. Thus, for example, the IFC/Amundi EGO bond fund focusing on green bonds in emerging markets explicitly states that selection process of green bonds is to ensure "that such investments contribute to a specific sustainable objective without significantly harming other objectives". Specifically, the fund follows an exclusion policy at the issuer level "based on Issuers' ESG score, taking into account portfolio exposure to high ESG risk and carbon-intensive sectors and to projects associated with potentially significant environmental and social risks and impacts, and/or sector-exclusion".[13]

Granted, relying on ESG scores to assess issuers can be problematic, since ESG ratings providers can significantly disagree on company ratings. For example, a recent study found that the correlation between the ESG scores of different ESG ratings providers was only 0.54, and even lower when looking at the individual E, S, and G pillars (Berg et al (2022); Boffo and Patalano (2020)).[14] All this is not say that managing reputational risk in green bond investing is impossible – rather, the point is that green bonds do carry their own specific risks and public investors should acknowledge them and establish rules and internal procedures to protect themselves against potential charges of naïveté or malpractice. This is particularly important given that abstaining from investment in green bonds may in some cases have an adverse impact on

13 ezjscamundibuzz::sfForwardFront::paramsList=service=ProxyGedApi&routeId=_dl_dacdf4fa-c323-46d9-8f4f-973f2a72c9e0_download See also "Green & Sustainable bonds: a label is not enough", Generali Investments, 2022, Microsoft Word - White_Paper#3_Green_Bonds_vfinal_(002).docx (generali-investments.com) and Barclays ESG Bond Handbook, 2020.

14 For comparison, the correlation between credit ratings assigned to issuers by Moody's and S&P stood at 0.99.

reputation as well – specifically, if stakeholders in the government or the society at large expect central banks and other public investors to contribute to funding the transition to net zero (Fender et al (2020)).

4.2 Does green bond investing make a positive environmental impact?

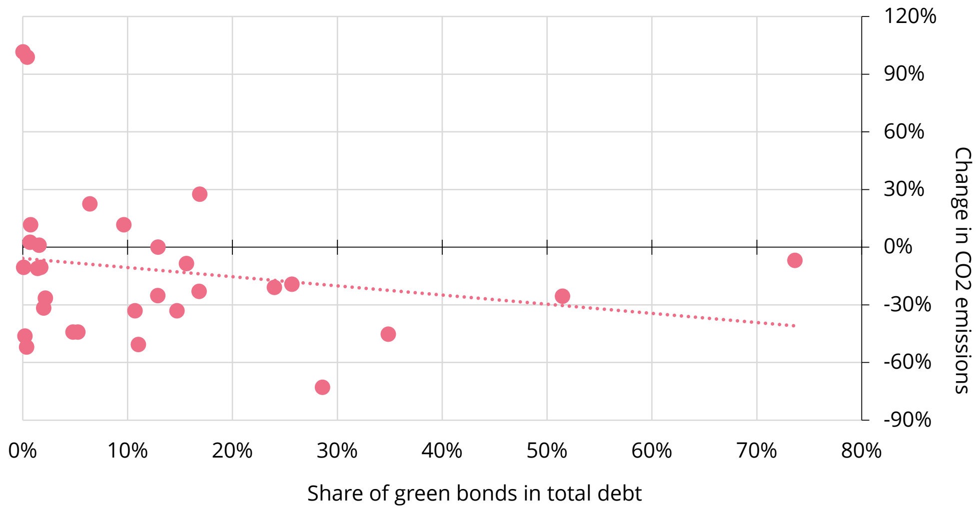
By definition, green bonds finance eligible environmental projects. However, investors might be asking themselves if, and to what extent, their purchases of green bonds are actually helping make a real difference in terms of facilitating the transition to a less carbon-intensive economy. Borrowing terminology from Busch et al (2021), we could rephrase the question and ask whether buying green bonds is an instance of true impact-generating investment or merely impact-aligned one. The difference is subtle, yet crucial, and goes back to the green/impact-washing concerns discussed above. Unfortunately, given limitations in data coverage and quality, providing a conclusive answer is not straightforward.

A recurring theme of the preceding pages was a palpable, strong momentum behind green finance. And certainly this could be taken as a first sign of impact – a "wall of money" flowing towards sustainable projects, rendering them more economically viable, as evidenced by the small greenium detectable at issue, particularly among some issuers. On the flip side, companies have also made big strides in trying to make it easier to assess their environmental credentials. An increasing number have formally adopted near- and long-term emissions reduction targets approved by the non-profit Science Based Targets (SBTi) initiative.

However, the relationship between progress on decarbonisation and green bond issuance is weak at best. One way to illustrate this is to screen the entire universe of close to 1,000 companies from around the world with approved decarbonisation targets, narrow down the sample to those who fund themselves even partially using green bonds, and compare their progress (relative to the chosen base year) against the share of green bonds in total debt outstanding. The result – plotted in Graph 4.2 – is rather underwhelming. The use of green bonds as a funding vehicle for decarbonisation activities appears to explain just about 4% of the variation in the reported change of generated emissions, and the association, while directionally consistent with intuition, is practically insignificant (t-stat of 1.12). A similar conclusion is derived more formally in a recent paper by Ehlers et al (2020), who confirm that there is no strong evidence that green bond issuance is associated with any reduction in carbon intensities over time at the firm level.

Companies' CO2 emissions reduction as a function of green financing

Graph 4.2



Note: the original sample includes close to 1,000 companies with approved decarbonisation targets as of 2022, which is subsequently narrowed down to only 420 for which it was possible to calculate actual decarbonisation progress (Scope 1 and 2 emissions relative to a base year); from these around 30 have placed green bond issues with at least $250 million outstanding; fitted regression line has R2 of 0.04.

Source: Bloomberg, Science-based Targets Initiative.

Echoing Ehlers et al (2020), this does not mean that green bonds haven't delivered – it is merely an indication that labelled instruments, while perhaps useful, are not a sine qua non of a successful decarbonisation strategy. In practice, sustainable projects may be financed only partially by green bonds, or even without recourse to such use-of-proceeds instruments altogether. A case in point is a recent purchase of Atchison Renewable Energy Center, a 300 MW wind farm, by Ameren Missouri, a large US energy utility. Of the project's purchase price of about $500 million, only $42.6 million was allocated from the proceeds of the issuance of a green bond. A large related project, a 400 MW farm called High Prairie Renewable Energy Center in Missouri, cost roughly $615 million, of which only $500 million was covered by the issuance of a green bond. Both projects clearly serve to reduce emissions in the region – a point we shall return to below – but their financing structure is determined by the company's broader funding plan which includes among others retained earnings and equity issuance and weighs their relative costs and merits against the company's objectives.

Although green bonds may fail as a proxy for the extent of decarbonisation taking place in the corporate sector en bloc, their underlying premise of ringfencing the use of proceeds and allocating them to specific projects invites investors to assess and compare green bonds not only in terms of their risk/return profile, but also in terms of the difference they are making to the corporations and societies in which they operate. In this context, Ehlers et al (2020) call for a carbon intensity-based rating system, while Partnership for Carbon Accounting Financials (PCAF), an industry-led initiative to improve greenhouse gas accounting standards for financial institutions, s actually suggested a related methodology to account for the carbon emissions of green bonds in December 2021. Although such approaches have yet to find their way into standard packages offered by commercial data vendors and rating agencies, and individual disclosures might be patchy and inconsistent, still post-issuance green bond reports coupled with academic studies and publicly available sources seem to actually provide quite a trove of useful information facilitating a relative-impact value analysis. The point of such comparisons would be to comprehensively assess the "greenness" of a particular issue, taking into account not just the nature of the project funded, or its alignment with a set of taxonomy criteria, but also the amount of greenhouse gas emissions it helped to prevent. To use a clichéd example – all things equal, one should probably prefer to finance a wind farm in a heavily coal-reliant emerging market economy than in an advanced economy already fully powered by renewable sources.

The question is, however, whether – and to what extent – things really are equal, and in particular whether a bond's environmental impact is already reflected in pricing. In a recent report, HSBC (2022) found little evidence to support that view and the updated results presented below confirm that initial finding.[15] The sample used in this example comprises 28 green bonds issued by eight largest US utilities[16] with sufficiently detailed disclosures to allow meaningful comparison. To reflect how bond proceeds are translated into real-world impact, and capture the distinction between the marginal benefit of a renewable project in a country with a clean vs a dirty grid, each issue is characterised by the emissions it helps displace or prevent. The latter category is scaled to account for the share of project cost covered by a given green bond. Finally, to account for the economic relative value of the bonds, each issue is represented by the ratio of its option-adjusted spread to duration (ie a measure showing roughly how much of a spread uplift an investor receives for a unit of duration risk exposure).

For example, consider the $550 million Ameren green bond issue mentioned above. The bond has a modified duration of 17 and trades at a spread of about 112 bp. But what is the climate "bang for the buck," ie how much real environmental change does the 112 bp help achieve and how does that compare with other available bonds? To get a handle on this, note that according to the post issuance report, the bond's net proceeds were allocated to fund the acquisition of two wind farms – the 400 MW High Prairie Renewable Energy Center ($500 million) and the 300 MW Atchison Renewable Energy Center ($42.6 million). Taking into account project costs (estimated at $555 million and $416 million, respectively, based on US averages reported by Irma), the green bond can be estimated to have financed 390.92 MW of total capacity. Now, according to the avoided emissions calculator provided by the US Environmental Protection Agency – which takes into account the grid composition in the US Midwest – the bond-financed renewable energy generation is estimated to have replaced roughly 1 million tons of CO_2 emissions. Relating the avoided emissions to green bond proceeds yields 1,946 CO_2 displaced per year per million dollars. Repeating an analogous analysis for all bonds in the sample yields a climate relative value estimate for US utilities (Graph 4.3).[17]

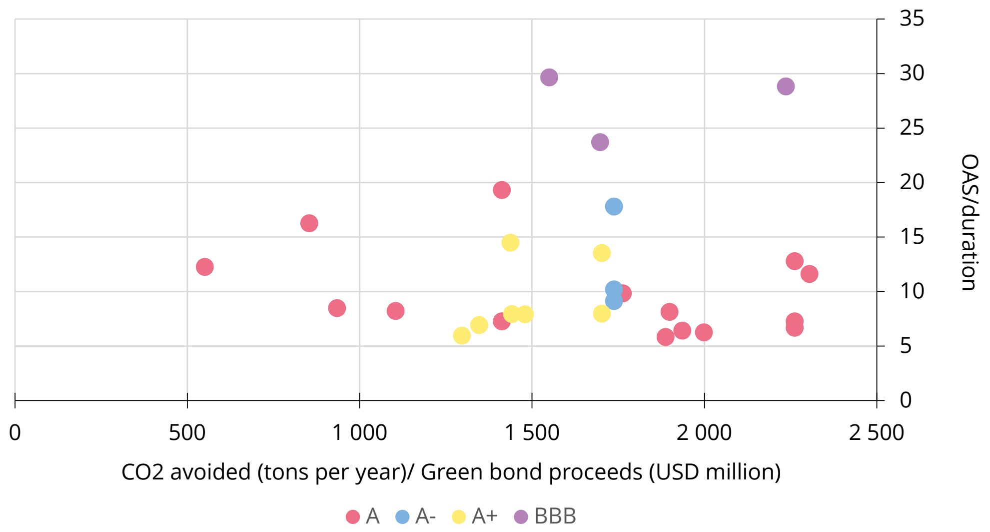
15 In a related recent study, Jarno and Richardson (2023) found that there was little differentiation in bond and CDS pricing levels for companies in the high-emitting oil and gas sector despite companies' widely differing progress in decarbonisation.

16 Xcel, DTE, Southern, Algonquin, MidAmerican, Duke, AEP and Ameren.

17 Note that this is a scoring exercise aimed not so much at *explaining* differences in bond spreads per unit of duration, but rather at comparing bonds in terms of their direct environmental impact, while taking into account their market risk characteristics. The latter, of course, can also be affected by other factors, over and above any environmental impact priced in by investors.

Climate-relative value: bond spread (per unit of duration) as a function of attributable avoided emissions

Graph 4.3



Note: the sample includes green bonds issued by the major US utility companies: Xcel, DTE, Southern, Algonquin, MidAmerican, Duke, AEP and Ameren for which relevant data could be ascertained.

Sources: adapted and updated from HSBC (2022), Bloomberg data.

Viewed against comparables, the Ameren bond delivers a substantial impact, but one could do better by considering debt issued by Southwestern Public Services, which finances a project associated with greater amount of averted emissions while offering a similar risk-return profile (a 118 bp spread with modified duration of 16). But an even more attractive proposition, if one were willing to go down a notch on the rating spectrum, might be the bond of Algonquin Power & Utilities Corporation, which funds a particularly high-impact wind farm in Kansas and offers an attractive spread of about 190 bp at moderate duration of just over 6.

While the scoreboard is just a starting point, and data availability/reliability issues might limit its broader use across the entire green bond universe, it does nonetheless demonstrate how one can go about building or tilting portfolio so as to achieve maximum impact subject to risk-return constraints. More importantly, perhaps, the scoreboard also demonstrates how investors can go about constructing/tilting their portfolios for maximum environmental impact, because the latter does not seem to be fully priced into the green bond market.

Concluding thoughts

As many central banks and public investors already invest in green bonds and many others are considering whether to start the journey, this article has attempted to discuss a number of issues that may be important in forming a coherent view of the labelled asset class and possibly also in formulating a green investment strategy. Not all of the questions listed in the introduction have clear-cut and straightforward answers (and arguably some answers provided above raise a whole set of new questions). It seems fairly obvious, for example, that over the past couple of years the green bond market has made tremendous gains in size, breadth and overall investability, making it an attractive proposition for public investors, typically focused on safety and liquidity of their portfolios. However, issues surrounding risk characteristics and return profile of green bonds are trickier.

We have seen that it would be naïve to expect that green bonds should outperform their conventional counterparts on a comparable basis, and there also seems to be little of a "safety premium" attached to them, and no meaningful diversification benefits – as evidenced in the Covid pandemic and later during a tumultuous 2022. In fact, a small greenium – ie spread give-up – seems to be more typical in the both the primary and secondary markets, leading over time to the erosion of carry and long-term returns. But we have also argued that a proper interpretation of this phenomenon requires a little bit of nuance. Although the pricing pattern is hardly an advantage from the narrow perspective of a risk-return-oriented portfolio optimisation, it does nevertheless spark hope that green bonds could actually be making a difference by improving the NPV, and hence also the attractiveness, of environmentally friendly projects. Yet it is not easy to pinpoint the extent to which green bonds as an asset class are making an impact in meaningfully speeding up the transition towards net zero.

What does seem clear, though, is that investors should be aware of shades of green, in that green bonds can significantly differ in the degree of impact they generate, eg in terms of emissions they help to avert. Perhaps more importantly, there is little evidence that these differences in carbon profile and impact level are fully reflected in the market pricing of green bonds. This suggests that there may still be avenues for investors to materially improve the impact they are exercising through their investments without necessarily sacrificing much in terms of pure risk-return trade-off.

References

Berg, F, J Koelbel and R Rigobon (2022): "Aggregate confusion: The divergence of ESG ratings" *Review of Finance*, vol 26, no 6.

Boffo, R, R Patalano (2020): *ESG investing: Practices, progress and challenges*. OECD.

Busch, T, P Bruce-Clark, J Derwall, R Eccles, T Hebb, A Hoepner and O Weber (2021): "Impact investments: a call for (re) orientation" *SN Business & Economics*, vol 1.

Ehlers, T, B Mojon and F Packer (2020): "Green bonds and carbon emissions: exploring the case for a rating system at the firm level", *BIS Quarterly Review*, September.

Erlandsson, U (2020): *How green bonds are (not) supposed to work. The case of State Bank of India and potential Carmichael lending*, Anthropocene Fixed Income Institute, 25 November.

Fender I, M McMorrow, V Sahakyan and O Zulaica (2020): "Reserve management and sustainability: the case for green bonds?", *BIS Working Papers*, no 849, March.

Generation Investment Management (2021): *Listed Company Emissions*, 11 October, www.generationim.com/our-thinking/insights/listed-company-emissions.

Godemar, M and J Luan (2022): *Green Bond Pricing: Gauging the Greenium*, Bloomberg NEF, 5 December.

Goldman Sachs (2020): *Carbonomics: The Green Engine of Economic Recovery*, Goldman Sachs Equity Research, 16 June.

HSBC (2022): *Global Credit & ESG. New ways to pursue net zero*, Global Thematic Research, September.

Jabłecki, J (2022): "Doing well while doing good: the elusive quest for green bond returns", *The Journal of Impact and ESG Investing*, Winter.

Jones, B, M Mendelson and L Pomorski (2023): *How Portfolios Can Impact the Real Economy*, AQR, February.

Kini, D, J White, C Attfield and F Will (2023): *Hold the Halo*, Green Bond Insights, HSBC, January.

Kini, D, L Lam, J White and S J Lee (2021): *Green Bond Outlook 2022*, HSBC, December.

MacAskill, E, B Roca, R Liu, A Stewart and O Sahin (2021): "Is there a green premium in the green bond market? Systematic literature review revealing premium determinants." *Journal of Cleaner Production*, vol 280, part 2.

Morgan Stanley (2019): *Decarbonisation: The Race to Net Zero*, Morgan Stanley Research, 21 October.

White, J, S J Lee, D Kini and T Russell (2022): *Tackling hard-to-abate sectors*, Fixed Income Credit Research, HSBC, 29 June.

Part 2

Climate risk and environmentally-aware investment strategies

Green sentiment, stock returns, and corporate behaviour*

Marie Brière[†] and Stefano Ramelli[‡]

Abstract

In this paper, we propose a new method for estimating non-fundamental demand shocks for green financial assets based on the arbitrage activity of exchange-traded funds (ETFs). By estimating the monthly abnormal flows into environmentally friendly ETFs, we construct a Green Sentiment Index that captures shifts in investor appetite for environmental responsibility which are not yet priced into the value of the underlying assets. Our measure of green sentiment differs significantly from the news-based climate indexes proposed by the existing literature, and it has additional explanatory power for both stock returns and corporate decisions. Over the period 2010–20, shifts in green sentiment anticipated a persistent stock price outperformance for more environmentally responsible firms, as well as an increase in their capital investments and cash holdings, particularly for the more equity-dependent ones.

JEL classification: G12, G32, G41.

Keywords: Climate finance, corporate behaviour, ESG, responsible investments, investor sentiment, non-fundamental demand, stock returns.

* The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. The authors thank Douglas Araujo, Jean-Jacques Barberis, Söhnke M. Bartram, Pascal Blanqué, Marco Ceccarelli (discussant), Stefano Colonnello (discussant), Shaun W. Davies, Giuliano De Rossi, Ana Mao de Ferro, Julien Fouquau, Emilia Garcia-Appendini, Francesco Franzoni, Anandakumar Jegarasasingam, Philippe Krüger (discussant), Tegwen Le Berthe, Ulf Lewrick, Robert Macrae, Benoît Mojon, Mike McMorrow, Frank Packer, Peter Pope, Sébastien Pouget, Steve Satchell, Koji Takahashi, Laurent Trottier, Alexander Wagner, Dora Xia, Omar Zulaica, Olivier David Zerbib for insightful comments.

† Amundi. Université Paris Dauphine, PSL Research University; Université Libre de Bruxelles. Email: marie.briere@amundi.com.

‡ University of St. Gallen and Swiss Finance Institute. Email: stefano.ramelli@unisg.ch.

1. Introduction

Environmental considerations, especially those related to climate change, are playing an increasingly prominent role in financial markets.[1] Factors driving this trend include the intensification of extreme weather events, the increase in public awareness of environmental issues, and recent regulatory developments.

Despite the growing emphasis on the importance of green finance by both practitioners and policymakers, our collective understanding of the effect of environmental concerns on financial markets and corporate decisions remains limited. Theoretical works indicate that investors' environmental preferences can affect asset prices and, in turn, corporate behaviour (Heinkel et al (2001); Pástor et al (2020)). From an empirical perspective, however, identifying and studying the real impact of investors' environmental preferences is challenging for at least two reasons. First, changes in such environmental preferences are not easily observable, let alone and measurable. Second, it is difficult to disentangle changes in environmental preferences from changes in expectations about a firm's fundamentals (cash flows and uncertainties), which are obviously also influenced by environmental-related factors, for instance regulatory risks (Xia and Zulaica (2022)).[2]

In this paper, we propose a novel method for estimating changes in investor appetite for green assets that addresses both problems. Our approach is based on the analysis of arbitrage activity in exchange traded funds (ETFs) – ie the creation and redemption of shares in the ETF primary markets, which leads to observable flows into ETFs – that previous works have shown to reflect non-fundamental investor demand shocks (see, in particular, Ben-David et al (2017); Brown et al (2021); Davies (2020)). The main reason why ETFs are likely to be more exposed to non-fundamental demand shocks than their underlying securities is because their ownership structure is more tilted towards retail clients and short-term institutional investors.[3] The main focus of our paper is on ETFs with explicitly environmentally friendly features – which we define as "green ETFs". In our sample, these green ETFs have a median institutional ownership of approximately 24%, compared with roughly 42% for conventional ETFs and above 70% for individual stocks.

Using data on a comprehensive sample of US equity ETFs from January 2010 through June 2020, we estimate, for each month, the differential flows into green ETFs relative to flows into conventional ETFs, net of the effects of other fund characteristics. We use the estimated abnormal flows into green ETFs to build our Green Sentiment Index. We present the details of the computation in Section 2.

We argue that the Green Sentiment Index reflects changes over time in investor taste for green assets that are not motivated by fundamental information. We show that it differs significantly from other proxies of attention to climate change used in

[1] As of 2020, $40.5 trillion of assets were managed accounting for environmental, social and governance (ESG) factors screening, representing close to 25% of total managed assets in the US and 50% in Europe. See www.gsi-alliance.org/trends-report-2018/.

[2] For instance, Krüger et al (2020) provide clear survey evidence that institutional investors consider climate risks, particularly regulatory ones, to have material financial implications for their portfolios.

[3] Given the differences in ownership structure, non-fundamental demand shocks affect an ETF's price differently from the net asset value (NAV) of its underlying securities. The resulting wedge creates an incentive for the ETF's Authorised Participants (APs) to create or redeem ETF shares, generating observable flows.

the existing literature, such as the Google search activity on "climate change" (Choi et al (2020); Ilhan et al (2021)) and the news-based climate risk indexes adopted by Engle et al (2020).

These measures are likely to reflect an undefined mix of both fundamental and non-fundamental information related to climate change and the environment. The key advantage of the approach we propose here is that it allows us to identify changes in investor demand for environmentally friendly assets that are mostly not motivated by changes in expected firm fundamentals. If they were, the value of the ETFs' underlying securities would have adjusted accordingly, without triggering the arbitrage mechanism behind the observed flows into green ETFs.

We use our Green Sentiment Index to establish two key results on the role of investor green sentiment. First, in Section 3, we study how green sentiment influences the value that investors attach to corporate environmental responsibility as priced by the stock market. We use the environmental score from the ESG data provider Sustainalytics, as in Engle et al (2020). We find that a one standard deviation stronger green sentiment is associated with an outperformance of one standard deviation for environmentally responsible firms, of approximately 27 basis points over a one-month horizon and 53 basis points over a six-month horizon, net of the effects of other firm characteristics and sector.

Importantly, the effect of green sentiment is independent from, and additional to, the effect of the news-based climate risk index used by Engle et al (2020). Indeed, both Engle et al's climate risk measure and green sentiment predict an outperformance of environmental responsibility, but for different reasons. While the former also predicts a positive revision in analysts' earnings forecasts on environmentally responsible firms, green sentiment does not, further confirming the validity of our approach. We also confirm that our results are not mechanically driven by the price pressure created by the rebalancing of ETFs themselves, ie the propagation channel that is explored and documented in Ben-David et al (2018).

Second, in Section 4 we use the Green Sentiment Index to study the effects of investor environmental preferences on real corporate decisions. We find that in quarters with higher green sentiment, environmentally responsible firms increase both their capital investments and cash holdings. A one standard deviation higher green sentiment is associated with 0.21% higher capex and 0.31% higher cash holdings – equal to approximately 5% and 3.4% of their respective sample means – for a one standard deviation higher environmental score. We do not observe any effect of green sentiment on firms' R&D activities.

Interestingly, the "real impact" of green sentiment on capex and cash holdings seems to vary across firms on the basis of their access to capital, as proxied by their credit rating. In particular, the influence of green sentiment on capex is focused on low- (non-investment grade) and medium-rated firms ("BBB", "BBB+", and "BBB–", based on the S&P scale). Conversely, the influence on cash holdings is focused on low-, and to a lesser extent, high-rated firms. These results confirm the importance of financial frictions in mediating the impact of responsible investing on firm behaviour.

Our paper contributes to three strands of research. First, we add to the literature on the effects of environmental preferences on financial markets. Several theoretical works suggests that investors' green preferences affect stock prices (Heinkel et al (2001); Fama and French (2007); Gollier and Pouget (2014); Landier and Lovo (2020); Luo and Balvers (2017); Oehmke and Opp (2020); Pástor et al (2020); Pedersen et al (2020); Zerbib (2022)). In particular, the model in Pástor et al (2020) predicts that green assets should outperform following unexpected upward shifts in investors' environmental preferences (even though, in equilibrium, green assets should experience lower returns – the opposite of what happens with "sin stocks", Hong and Kacperczyk (2009). Battiston et al (2021) and Gourdel et al (2021) provide climate stress tests of the financial system, and simulate how investors' expectations affect climate policy effectiveness. However, from an empirical perspective, identifying those shifts is far from obvious. Approaches based on climate-related attention and news-based measures (eg Choi et al (2020); Engle et al (2020); Huynh and Xia (2021)) are likely to partially or primarily reflect the arrival of new fundamental information.[4] In a contemporaneous work, Pástor et al (2022) use the spread between German green and non-green bonds to study the asset-pricing effects of changes in climate concerns, although they do not aim at disentangling the fundamental and non-fundamental drivers of green demand. Van der Beck (2021) estimates that the performance of ESG investments is strongly driven by price pressure arising from flows towards sustainable funds. The ETF-based approach that we propose has the advantage of specifically capturing shifts in investor taste for green assets that are *not driven* by firm-fundamental considerations. In addition to this methodological contribution, our paper applies the proposed approach to shed new light on the effects of investor environmental preferences on firm value and real corporate decisions, confirming some key predictions of theory (Pástor et al (2020)) and contributing to the flourishing empirical literature on climate finance (eg Anderson and Robinson (2020); Bartram et al (2022); Bolton and Kacperczyk (2021, 2022); Ceccarelli et al (2023); Choi et al (2020); Ilhan et al (2021); Pankratz and Schiller (2019); Ramelli et al (2021)).

Second, we contribute to the literature on the effects of investor sentiment. Sentiment is also known to affect firms' financing and investment decisions (Baker and Wurgler (2000); Henderson et al (2006); Kim and Weisbach (2008)). Da et al (2015) measure market-level sentiment based on Google search behaviour. There are, of course, different types of investor sentiment. For instance, Baker et al (2012) and Ben-Rephael et al (2019) study the effects of foreign sentiment. We contribute to this literature by measuring and studying a new class of investor sentiment, the one pertaining to environmentally related considerations. Again, the main advantage of our ETF-based approach is its ability to control for changes in expectations about firm fundamentals. By studying how green sentiment influences corporate decisions, we also link to the debate on the real effects of financial markets (eg Morck et al (1990); Luo (2005); Bakke and Whited (2010); Bond et al (2012); Dessaint et al (2019)).

Finally, the paper also contributes to the growing literature on ETFs. Although ETFs represent one of the most important financial innovation of the last decades, research on this market remains relatively scarce. Ben-David et al (2017) provides an interesting review of the early literature. Ben-David et al (2018) show that the arbitrage mechanism of ETFs propagate liquidity shocks to the underlying securities, increasing their volatility. Glosten et al (2021) find that ETF activity increases informational efficiency for stocks with weak information environments and imperfectly competitive equity markets. Ben-David et al (2023) find that specialised

4 Other recent works analysing news-based measures of attention to climate change include Bessec and Fouquau (2020), Faccini et al (2021) and Santi (2023). Other papers propose to capture firm-level exposures to climate risks – but not changes in investor environmental sentiment – based on the text analysis of corporate earnings calls (Sautner et al (2023); Li et al (2020)) or adopting machine learning techniques on annual reports (Bingler et al (2022)).

ETFs compete for flows by catering to the attention of unsophisticated investors, and deliver negative risk-adjusted returns. Rather than studying the direct effects of ETFs, in our paper we exploit their unique arbitrage mechanism to proxy a market sentiment. Works applying a similar approach are Brown et al (2021) and Davies (2022). Brown et al (2021) show theoretically and empirically that the creation and redemption of ETF shares provide observable signals of non-fundamental pressure on prices. Davies (2022) exploits the arbitrage activity of leveraged ETFs to build a "speculation sentiment index" proxying for the magnitude and direction of speculative demand shocks. In a similar spirit, we exploit the arbitrage activity on green ETFs to proxy for the magnitude and direction of shocks of non-fundamental demand for green financial assets. Given the popularity of these financial products among retail investors, they are particularly likely to reflect non-fundamental demand pressure for environment-friendly assets. Our paper aims at providing insight on the desirable and undesirable consequences of green sentiment.

2. Identifying green sentiment from ETF arbitrage activity

This section presents a proposed methodology for identifying green sentiment based on ETF flows, describes the data used in the empirical investigation, and illustrates the main properties of the estimated Green Sentiment Index.

2.1 Empirical strategy

Exchange-traded funds (ETFs) are pooled investment vehicles that track an index or a basket of underlying securities. They represent one of the most important financial innovations of recent decades. As of year-end 2020, the ETF market had more than $7.9 trillion of assets under management worldwide, with 69% concentrated in the approximately 2,200 ETFs domiciled in the United States (Investment Company Institute (2021)). ETFs account for approximately 18% of all assets managed by US investment companies, progressively eroding the space traditionally held by mutual funds.[5] The ETF market is also very liquid, with an average trading equal to approximately 26% of the trading of US securities (Investment Company Institute (2021)). We refer to Ben-David et al (2017), Lettau and Madhavan (2018), Ben-David et al (2018), and Pagano et al (2019) for a more comprehensive overview of ETFs and some of their documented effects on financial markets.

A key feature of ETFs is their arbitrage mechanism. In the secondary market, ETFs are traded like ordinary stocks, without involving any trading of the underlying securities. The price at which an ETF is exchanged can freely deviate from the asset-weighted net asset value (NAV) of the underlying securities. This potential mispricing is corrected by the activity of third-party arbitrageurs – known as the "authorised participants" (APs) – which can demand the ETF to issue or redeem shares, causing observable flows of capital into or out of the ETF.

[5] See eg Bloomberg, "Mutual funds bleed $469 billion as ETFs triumph in zero-sum 2020", 13 December 2020.

Brown et al (2021) show theoretically and empirically that ETF arbitrage activities reflect non-fundamental demand shocks. The main reason for this result is that ETFs have an ownership structure that is usually quite different from the ownership of the underlying securities, with a larger component of retail (non-sophisticated) investors. Given the difference in ownership structure, non-fundamental demand shocks are likely to affect the price of ETF more than the value of its underlying securities. The resulting mispricing between the ETF and the NAV (a premium or a discount) incentivises APs to create or redeem ETF shares in the primary market, causing observable ETF flows. These flows – contrary to the flows in eg mutual funds – reveal the presence of non-fundamental demand shocks, otherwise hard to disentangle from the effect of new fundamental information. Our main intuition is to exploit the unique features of the ETF market, in the spirit of Brown et al (2021) and Davies (2022), to measure changes in green non-fundamental demand shocks, ie shifts in investors' appetite for environmental responsibility not yet incorporated in the values of the underlying assets. We do that by studying the primary market of "green" ETFs, ie ETFs allowing environmentally conscious investors to replicate a basket of environmentally responsible securities.

Investor appetite for ESG investments has been growing rapidly in recent years. Assets under management of ESG-focused funds worldwide have risen from some $340 billion in 2015 to over $1.6 trillion in 2020. The size of US-domiciled ESG funds market has doubled since 2015, and now amounts to over $230 billion, with over 15% in the form of ETFs (Morningstar (2020)). As for sustainable funds, the size of the green ETF market has grown tremendously. The US segment of the green ETF market, for example, increased almost fourfold between January 2015 and June 2020 (from $1.1 billion to $4.1 billion).

The main assumption behind our approach is that the demand for green ETFs is more sensitive to non-fundamental information than the demand for individual stocks. There are good reasons to believe this assumption to be true. In general, ETFs are used predominantly by retail investors (Ben-David et al (2017); Brown et al (2021); Ben-David et al (2023)), with an average institutional ownership significantly lower than the institutional ownership of individual stocks. Following Stambaugh (2014)'s argument that uninformed traders are mostly present among retail investors, this suggests a higher density of liquidity traders in the ETF investor base.

This is even more true for specialised products such as green ETFs, which particularly appeal to retail and sentiment-driven investors (Ben-David et al (2023)). The average share of institutional ownership of the green ETFs in our sample is 31%, significantly lower than for conventional equity ETFs (49%). The dominant presence of retail investors, combined with the fact that, as for other ETFs, green ETFs can be used by institutional investors to gain a short-term exposure to the green segment of the market, may help to make green ETFs more sensitive to non-fundamental demand shocks than are the underlying green securities. A shock related to an exceptionally high demand for green investment can give rise to a relative mispricing, and the subsequent creation or redemption of ETF shares to correct it. The segmentation of investors between green ETFs and the underlying green stocks' markets is likely to create a wedge (a premium or discount) between the price of green ETFs and the value of the underlying securities, triggering a change in flows to green ETFs.

To measure the creation/redemption activity by APs in the ETF market, we define $Flows_{i,t}$ as the monthly percentage change in ETF shares outstanding for fund i at time t:

$$Flows_{i,t} = \frac{SharesOutstanding_{i,t}}{SharesOutstanding_{i,t-1}} - 1$$

We measure green sentiment as the differential inflows in green ETFs (non-fundamental demand on green ETFs) compared with the inflows of other ETFs, net of the effects of other observable ETF characteristics. Specifically, for every month in our sample, we run the following T cross-sectional regressions of monthly ETF flows:

$$Flows_{i,t} = c_t + \gamma_t \times GreenETF_{i,t} + \delta_t \times controls_{i,t} + E_{i,t}, \ \forall t$$

(1) where *GreenETF* is an indicator for green ETFs and *controls* is a vector of ETF characteristics: past month ln(NAV), return and volatility.[6] We define the standardised time series of estimated coefficients on *GreenETF* as our Green Sentiment Index.

2.2 Data

For all equity ETFs domiciled in the United States, we retrieve survivorship bias-free data (shares outstanding, volume traded, net asset value, last price and the percentage of institutional ownership) from Bloomberg.[7] We identify a total of 3,887 individual ETFs (of which, 406 are exchange traded notes, ETN) over the period from January 2010 through June 2020.

From Morningstar Direct, we obtain information on ETFs' categories, keeping only funds classified as "equity funds" and dropping funds investing exclusively outside the US and long/short equity funds.[8] We also retrieve the following additional information from the ETF Global data set: inception date, net expenses, creation fees, and whether the fund is levered or not. The final sample includes 1,195 individual ETFs.

Table 1 displays summary statistics for our sample of ETFs over the period from January 2010 through June 2020. The average AUM is close to $2 billion and the average number of shares outstanding is $31 billion. Both distributions are highly skewed to the right, confirming the very high concentration of the ETFs market (Pagano et al (2019)). Monthly flows represent on average 2% of the total number of shares outstanding.

6 As a robustness check, we added several additional characteristics in the regressions, such as ETF age (number of years since inception), percentage of institutional ownership, net expense ratio, creation fee and a dummy for levered ETFs. As an additional robustness check, we also used weighted least squares (WLS) regressions where observations were weighted by ETFs AUM, with similar results.

7 Bloomberg is recognised as the most accurate source for ETF data (Ben-David et al (2018)).

8 More precisely, we drop the following categories: Europe Equity Large Cap, Japan Equity, Latin America Equity, Asia ex-Japan Equity, Asia Equity, Global Emerging Markets Equity, Greater China Equity, Canadian Equity Large Cap, Africa Equity, Thailand Equity, India Equity, Korea Equity, Mexico Equity, Australia & New Zealand Equity, Long/Short Equity. Equity, Canadian Equity Large Cap, Africa Equity, Thailand Equity, India Equity, Korea Equity, Mexico Equity, Australia & New Zealand Equity, Long/Short Equity.

Interestingly, equity ETFs appear to be equally used by institutional and retail investors, with an average institutional ownership of approximately 49%. Importantly, this institutional ownership is significantly lower than the average institutional ownership of individual stocks (approximately 65%). The average net expense ratio is 46 basis points and creation fee $1,588. Fewer than 1% of the ETFs in our sample are levered.

Descriptive statistics of ETF variables — Table 1

	p5	p25	mean	p50	p75	p95	sd	N
AUM	0.01	0.05	2.26	0.20	0.89	9.42	10.62	81,929
Shares outstanding	0.25	1.50	31.19	4.85	19.10	144.30	94.38	81,929
Flows (%)	-9.29	-0.89	2.05	0.00	3.29	17.83	10.59	81,929
NAV	16.91	26.84	52.24	38.58	62.19	130.25	45.32	81,929
Return	-1.07	-0.25	0.06	0.11	0.41	1.03	0.68	81,929
Volatility	0.07	0.11	0.17	0.15	0.20	0.31	0.08	81,929
Age	0.88	2.90	7.10	6.34	10.54	16.27	4.88	67,322
Institutional Ownership (%)	10.24	29.00	49.03	45.17	66.62	100.00	26.33	80,846
Net expense ratio	8.40	25.00	45.67	44.00	60.00	80.00	42.56	65,917
Levered ETF	0.00	0.00	0.00	0.00	0.00	0.00	0.06	67,286
Creation fee	250.00	500.00	1,588.26	500.00	1,500.00	7,000.00	2,768.64	64,749

Sources: Bloomberg, Morningstar Direct, ETF Global; authors' calculations.

A critical choice in our empirical investigation is how to identify "green ETFs". We classify as green those ETFs whose names include one of the following keywords: "climate", "carbon", "clean", "solar", "fossil", "renewable", "environment", "wind", "ecological", "green energy", "progressive energy". In addition, we perform a manual check on the names and the prospectus of ETFs to avoid omitting any additional funds with explicit and salient environmentally conscious features. We identify a total of 23 green ETFs, listed in Table 2.

List of green ETFs Table 2

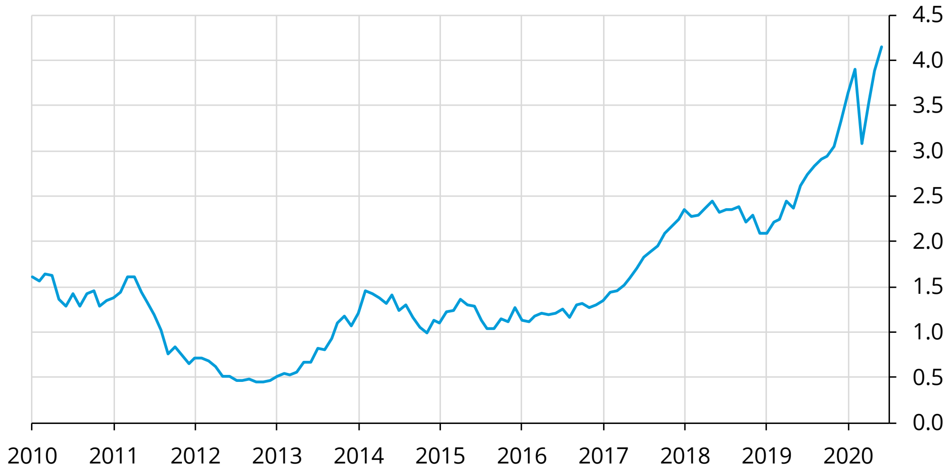
Ticker	ETF name	Net expense ratio (bp)	Inception - delisting	Morningstar sustainable?
ICLN	iShares Global Clean Energy	48	2008 -	yes
TAN	Invesco Solar	70	2008 -	yes
SPYX	SPDR S&P 500 Fossil Fuel Reserves Free	20	2015 -	yes
CRBN	iShares MSCI ACWI Low Carbon Target	20	2014 -	yes
PBW	Invesco WilderHill Clean Energy	70	2005 -	yes
QCLN	First Trust NASDAQ Clean Edge Green Energy	60	2007 -	yes
PZD	Invesco Cleantech	68	2006 -	yes
ACES	ALPS Clean Energy	65	2018 -	yes
SMOG	VanEck Vectors Low Carbon Energy	63	2007 -	yes
EFAX	SPDR MSCI EAFE Fossil Fuel Free	20	2016 -	yes
FAN	First Trust Global Wind Energy	60	2008 -	yes
ETHO	Etho Climate Leadership US	47	2015 -	yes
PBD	Invesco Global Clean Energy	75	2007 -	yes
LOWC	SPDR MSCI ACWI Low Carbon Target	20	2014 -	yes
YLCO	Global X YieldCo&Renewable Engy Income	65	2015 -	no
EVX	VanEck Vectors Environmental Services	55	2006 -	yes
CNRG	SPDR Kensho Clean Power	45	2018 -	yes
VEGN	US Vegan Climate	60	2019 -	yes
CHGX	Change Finance US LargeCap FossilFuel Free	49	2017 -	yes
PUW	Invesco WilderHill Progressive Energy	70	2006 - 2019	yes
HECO	Strategy Shares EcoLogical Strategy	95	2012 -	yes
RENW	Pickens Morningstar Renewable Energy Response	65	2019 -	yes
ECLN	First Trust EIP Carbon Impact	95	2019 -	yes

Sources: Bloomberg, Morningstar Direct, ETF Global; authors' calculations.

Importantly, among the identified green ETFs, only one is not classified as "sustainable" on the Morningstar Direct platform. The largest green fund is the iShares Global Clean Energy, with $721 million in assets under management as of June 2020. The oldest fund, Invesco WilderHill Clean Energy, was created in 2005.

As of June 2020, the total assets under management for green ETFs is above $4 billion. The size is relatively small compared with conventional equity ETFs (more than $4,000 billion), but it has been rapidly growing, with assets under management more than doubling over our sample period, as shown in Graph 1.

Evolution of green ETF assets under management — Graph 1.1
In billion USD



Source: Bloomberg, authors' calculations.

2.3 The Green Sentiment Index

We here compare the ETF flows-based Green Sentiment Index estimated based on Equation 1 with two measures of climate-related attention/risk proposed in the existing literature: the "Crimson Hexagon negative climate news" index proposed by Engle et al (and then employed also in Huynh and Xia (2021) and Ceccarelli et al (2023) and the Google search volume index (SVI) for the topic "climate change", as used, for instance, in Choi et al (2020) and Ilhan et al (2021).

The negative climate news index of Engle et al is particularly interesting for our purposes because it is meant to proxy climate risk, ie climate-related fundamental information.[9]

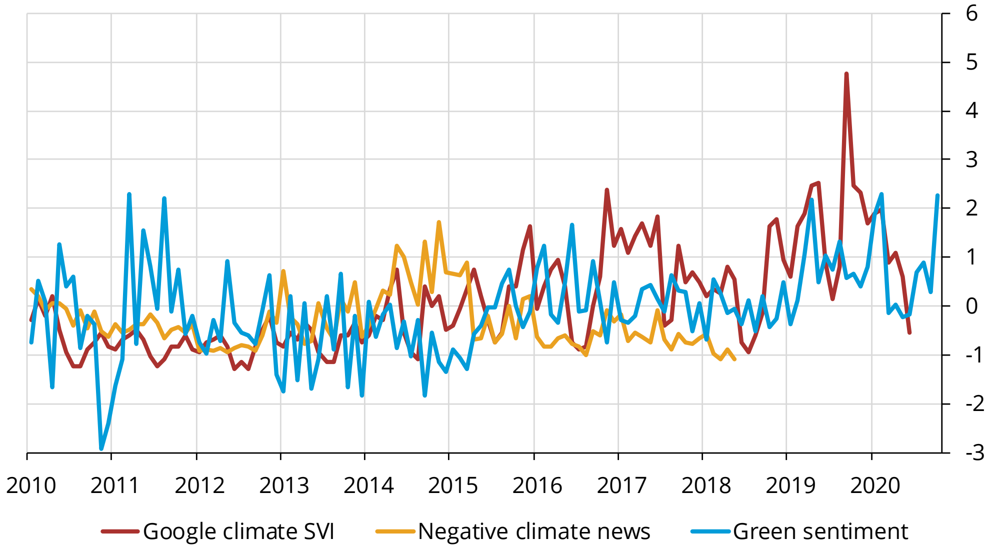
Our Green Sentiment Index aims at capturing the opposite side of the demand for green financial assets, ie that not related to fundamental considerations. The Google SVI is a good proxy for the level of public attention on specific topics, climate change in our case, and it is therefore likely to reflect a mix of both fundamental and non-fundamental information. Graph 2 plots the three indexes, all standardised to facilitate a comparison. We observe quite different patterns over time. All the three indices spike around the signature of the Paris Agreement in December 2015, but with slightly different timing. *Google climate SVI* and *Green sentiment* reflect the rising awareness of climate change in more

[9] Engle et al obtain this index from the data provider Crimson Hexagon (CH). The index represents the share of all news articles in major outlets that are both about "climate change" and have a "negative sentiment" as categorised by CH. The index is available from January 2008 through May 2018. We thank Stefano Giglio and Johannes Stroebel for making these data available on their websites.

recent years (especially after 2018). Interestingly, *Green sentiment* also spikes in early 2020 in correspondence with the Covid-19 crash.[10]

Evolution of the Green Sentiment Index — Graph 2.1

Green sentiment is measured as the differential inflows in green ETFs compared with the inflows of other ETFs, net of the effects of other observable ETF characteristics. The three indices are normalised.



Note: *Negative climate news* is the standardised negative climate news index proposed by Engle et al (2020). *Google climate SVI* is the Google search volume index for the topic "climate change" in the United States.

Source: Engle et al (2020); Google SVI; authors' calculations.

Table 3 shows the pairwise correlation between the three indexes. As expected, *Green sentiment* correlates *positively* with *Google climate SVI* (0.29, $p < 0.001$) but it correlates *negatively* with Engle et al's *Negative climate news* (-0.28, $p < 0.01$). *Google climate SVI* and *Negative climate news* do not significantly correlate with each other (0.08, $p > 0.1$).

[10] We interpret this evidence as suggesting that the increased investor attention to environmental issues following the outbreak of the pandemic – that existing research has identified also in terms of stock price (out-)performance of firm environmental responsibility (Albuquerque et al (2020); Pastor and Vorsatz (2020); Garel and Petit-Romec (2021)) – is not driven primarily by fundamental information and the behaviour of institutional investors.

Pairwise correlation between indexes			Table 3
	1.	2.	3.
1. Green sentiment	1		
	(125)		
2. Negative climate news	-0.28***	1	
	(101)	(101)	
3. Google climate SVI	0.29***	0.08	1
	(125)	(125)	(101)

Note: Number of observations in parentheses. ***, **, and * indicate that the correlation is significantly different from zero at the 1%, 5%, and 10% level, respectively.

Source: Engle et al (2020); Google SVI; authors' calculations.

Overall, this descriptive evidence supports the claim that the Green Sentiment Index captures investor demand for green assets not driven by hedging-climate-risk purposes or changes in expectations about firms' fundamentals. If it were – reflecting, for instance, the potential benefit of green ETFs in facilitating "price discovery" after fundamental shocks (Glosten et al (2021)) – we would have observed a positive correlation with news-based climate risk measures capturing the arrival of financially material information.

In the next sections, we investigate what are the effects of this type of investor sentiment on both firm value and corporate behaviour.

3. Effects of green sentiment on stock returns

In this section, we use the Green Sentiment Index to shed light on the channels behind the effects of corporate environmental responsibility on stock prices. The theoretical literature on responsible investing predicts that shifts in investor environmental preferences should be followed by a decrease in the cost of capital of more environmentally responsible firms (Heinkel et al (2001); Fama and French (2007); Pástor et al (2020)). Our approach to measuring green sentiment provides a powerful tool to empirically test whether this is really the case, net of the effects of changes in expectations about firm fundamentals.

3.1 Data

We retrieve monthly stock prices for common shares listed on major US stock exchanges (NYSE, NYSE Arca, AMEX, and NASDAQ) from January 2010 through June 2020, from the Compustat Capital IQ North America Daily database. We adjust prices for dividends through the monthly multiplication factor and the price adjustment factors provided by Compustat. In cases of dual listings, we keep only the firm's security with the highest market capitalisation. For every month, we trim returns at the first and 99th percentiles to reduce the impact of outliers. We also use monthly returns to compute buy-and-hold returns in windows of up to six months (eg *Cumulative return* $t+6$).

For each stock, we estimate the *Market beta* from regressions of monthly returns in excess of the 1-month Treasury-bill rate on the excess market return using a 36-month moving window, when at least 24 months of non-missing returns are available. We use the excess returns on the market factor available from Kenneth French's website. For each month, we also compute *Momentum* as the average individual stock return from month t-12 to t-2, as in Bessembinder et al (2019).

From Compustat, we also retrieve the following firm-level annual accounting characteristics: *Leverage* (long-term debt plus debt in current liabilities, divided by total assets, in percentage points: (dltt + dlc) $\times$ 100/at), *Size* i.e. *Log(market cap)* (ln(prcc f $\times$ csho)), *Book-to-market* (book value of equity divided by market valuation: ceq/(prcc f $\times$ csho)), and *Profitability* (annual income before extraordinary items over total assets: ib$\times$100/at).

We merge the above Compustat data with the firms' ESG scores from Sustainalytics, which are also employed in Engle et al.[11] To facilitate the economic interpretation of the results, we standardise the environmental scores from Sustainalytics to have mean of 0 and a unit standard deviation. As an alternative proxy for environmental responsibility, we compute the firms' environmental score using the MSCI KLD database.[12] Specifically, *ENV (kld)* is defined as the fraction of covered environmental "strengths" indicators equal to one minus the fraction of covered environmental "concerns" indicators equal to one, following a common practice in the ESG literature (eg Krüger (2015); Lins et al (2017)).

We end up with a sample of approximately 95,000 firm-month observations from January 2010 through June 2020 with available stock returns, accounting information and environmental score. Table 4 reports descriptive statistics of the main variables used in our analyses. We omit a detailed discussion of these statistics for the sake of brevity.

11 Given that the Sustainalytics scores at our disposal are available for the period from 2010 through 2017, we expand the latest available score through June 2020, relying on the stickiness of ESG scores.

12 The MSCI KLD data set, which we access through WRDS, provides a series of dummy variables indicating, for each firm and year, the presence of strengths or concerns on several environmental, social and governance factors.

Descriptive statistics of firm-level characteristics								Table 4
	p5	p25	mean	p50	p75	p95	sd	N
	Firm-level characteristics (monthly observations)							
Return	-12.88	-3.55	0.98	1.12	5.54	14.36	8.92	95,248
Cumulative return t+6	-28.64	-6.18	6.38	6.33	18.35	40.89	21.74	87,756
Env score	-1.32	-0.78	0.04	-0.14	0.73	1.90	1.01	95,248
Env score (kld)	-0.39	-0.39	0.49	-0.39	1.17	3.21	1.33	84,995
Leverage	0.00	13.34	28.95	26.82	40.60	64.10	21.62	95,248
Market beta	0.21	0.71	1.12	1.07	1.47	2.18	0.61	95,248
Log(market cap)	7.43	8.26	9.12	8.97	9.84	11.41	1.21	95,248
Book-to-market	0.03	0.20	0.45	0.37	0.62	1.13	0.41	95,248
Profitability	-3.24	1.55	4.98	4.38	8.28	16.16	7.41	95,248
Momentum	-2.83	-0.13	1.08	1.18	2.41	4.63	2.35	95,248
Green ETF ownership	0.00	0.00	0.01	0.00	0.00	0.01	0.06	95,248
	Firm-level characteristics (quarterly observations)							
Capex/PPE	-10.35	0.95	4.15	3.96	7.80	19.03	9.19	25,055
Cash/Assets	0.35	2.31	9.00	6.11	12.62	27.46	9.35	32,599
R&D/Assets	0.00	0.00	1.29	0.64	1.90	4.70	1.89	15,277

Source: authors' calculations.

3.2 Main results on stock returns

Table 5 reports the result of OLS regressions of individual stock returns on the interaction between the Green Sentiment Index and the firm's environmental score, as well as standard firm characteristics (*Leverage*, *Market beta*, *Log(market cap)*, *Book-to-market*, *Profitability* and *Momentum*).[13] The regressions also include sector fixed effects based on the GICS industry group classification (comprising a total of 26 industries). We cluster standard errors at the firm level to control for the correlation of residuals within firms.[14]

[13] We control for firm characteristics instead of a stock's estimated loadings on the size, value and quality factors following Kelly et al (2019) and Bessembinder et al (2019). However, we obtain very similar results when controlling for factor loadings instead of firm characteristics, or even using model-adjusted returns on the left-hand side of the regressions.

Green sentiment and the pricing of corporate environmental responsibility

Table 5

Dependent variable:	Return in t	Cumulative return through:					
		t+1	t+2	t+3	t+4	t+5	t+6
Green sentiment × Env score	0.068**	0.272***	0.290***	0.440***	0.495***	0.397***	0.527***
	(2.36)	(5.32)	(4.26)	(5.55)	(5.34)	(3.60)	(4.14)
Env score	0.039	0.033	0.098	0.189	0.153	0.175	0.179
	(1.16)	(0.48)	(0.97)	(1.42)	(0.94)	(0.89)	(0.78)
Green sentiment	1.153***	-2.163***	-2.248***	-2.275***	-2.572***	-2.106***	-1.952***
	(-36.58)	(-40.95)	(-34.09)	(-29.16)	(-28.18)	(-20.02)	(-16.79)
Leverage	0.003	0.004	0.004	-0.001	0.000	0.000	0.001
	(1.59)	(1.21)	(0.76)	(-0.09)	(0.01)	(0.02)	(0.10)
Market beta	0.245***	0.178	0.158	-0.210	-0.345	-0.544	-0.729*
	(3.47)	(1.36)	(0.85)	(-0.91)	(-1.20)	(-1.58)	(-1.80)
Log(marketcap)	0.006	0.011	-0.056	-0.163	-0.152	-0.201	-0.222
	(0.20)	(0.19)	(-0.65)	(-1.46)	(-1.10)	(-1.22)	(-1.16)
Book-to-market	-0.133	-0.405	-0.565	-0.682	-1.000*	-1.011	-0.965
	(-1.00)	(-1.54)	(-1.51)	(-1.48)	(-1.84)	(-1.57)	(-1.30)
Profitability	-0.003	-0.017	-0.030*	-0.048**	-0.070***	-0.090***	-0.112***
	(-0.59)	(-1.50)	(-1.78)	(-2.27)	(-2.72)	(-2.89)	(-3.07)
Momentum	-0.145***	-0.255***	-0.292***	-0.146***	-0.147**	-0.089	-0.027
	(-9.27)	(-8.57)	(-6.64)	(-2.61)	(-2.14)	(-1.09)	(-0.29)
Constant	0.857***	2.131***	3.824***	5.967***	7.168***	8.812***	10.216***
	(2.72)	(3.49)	(4.31)	(5.21)	(5.10)	(5.24)	(5.23)
Observations	95,248	93,972	92,704	91,444	90,199	88,969	87,756
R-squared	0.018	0.035	0.030	0.028	0.032	0.025	0.024
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: t-statistics based on standard errors clustered at the firm level are reported in parentheses. ***, **, and * indicate that the coefficient estimate is significantly different from zero at the 1%, 5%, and 10% level, respectively.

Source: authors' calculations.

The coefficient of interest is on the interaction between *Green sentiment* and the firm's environmental performance, *Env score*. We observe that, in months with a one standard deviation higher green sentiment, firms with a one standard deviation stronger environmental[14] performance experience a 7 basis points higher return (Column 1 of Table 5). The effect is statistically significant, but economically small. We interpret this result as indicating that the pressure on the price of environmentally responsible shocks caused directly by green ETFs' arbitrage activity

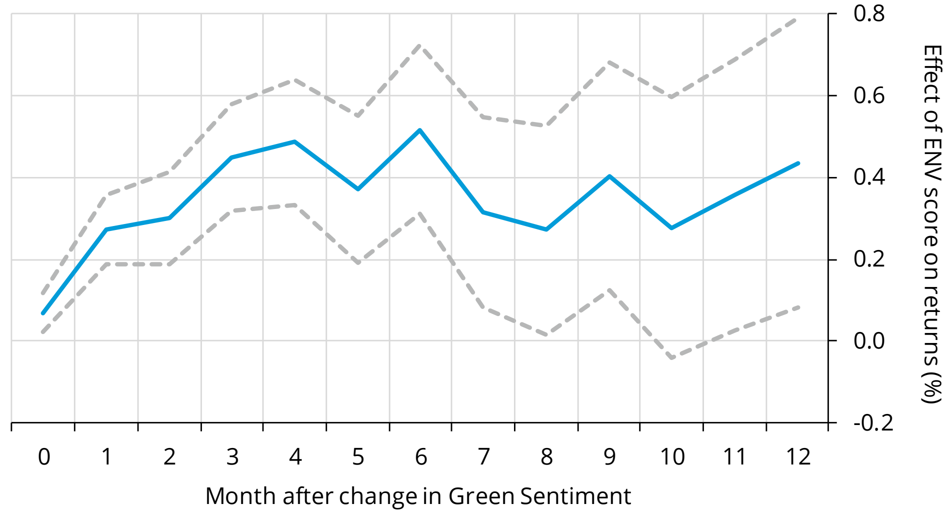
[14] As discussed in Section 3.5, our findings remain statistically significant even when double-clustering standard errors both at the firm and time dimensions (Petersen (2009); Thompson (2011)). We present our main results clustering at the firm level because, given the relatively short period analysed, clustering standard errors (also) at the time level risks to be excessively restrictive.

– a channel similar to the one documented in Ben-David et al (2018) – is, in our setting, limited. Indeed, not all environmentally responsible firms are included in green ETFs and hence directly affected by their arbitrage activity. We confirm this intuition in Section 3.4 by directly controlling for green ETFs' ownership.

When looking at the effect in *t+1* (column 2), we find that green sentiment predicts a strong outperformance associated with a firm's environmental responsibility. A one standard deviation higher green sentiment leads to approximately 27 basis points per additional standard deviation of environmental responsibility. The effect is highly statistically significant.

Effect of Green Sentiment on the pricing of corporate environmental responsibility — Graph 3.1

Effect of one standard deviation higher green sentiment on the stock returns of firms with one standard deviation stronger environmental performance



Source: authors' calculations.

Interestingly, the effect does not appear to revert in the following months (columns 3 to 7). It slightly increases in magnitude through *t+4* and remains stable in *t+5* and *t+6*. A one standard deviation higher green sentiment in *t* leads to an outperformance of one standard deviation more environmentally responsible firms approximately equal to 0.53% through *t+6*. Graph 3 illustrates the evolution of this effect using cumulative returns through *t+12*. Even when looking at such extended time frame, the stock price effect of green sentiment persists, with only a mild reversal.

One may have expected the stock price effects of green sentiment to be only temporary, as for other forms of investor sentiment (Baker and Wurgler (2006)). However, as also noticed in Pástor et al (2020), shifts in green tastes, although not driven by fundamental considerations, are likely to be persistent and, hence, drive a long-lasting effect on stock prices.

3.3 Green sentiment vs climate risk

We here conduct three tests further supporting the claim that the effect of green sentiment on the value of corporate climate responsibility is not driven by new fundamental environmentally related information.

First, in Table 6, we split the sample period before and after the Paris Agreement was signed in December 2015. Previous works document that the Paris Agreement significantly increased the salience and materiality of climate transition risks for institutional investors (Bolton and Kacperczyk (2022); Delis et al (2019); Seltzer et al (2021)). We expect the effect of green sentiment to be independent from such an important regulatory development. Table 6 shows that this is actually the case: green sentiment mediates the value of environmental responsibility both before and after the Paris Agreement.

Green sentiment and stock prices: before and after the Paris Agreement — Table 6

			Panel A: Before December 2015				
Dependent variable:	Return in t	Cumulative return through:					
		t+1	t+2	t+3	t+4	t+5	t+6
Green sentiment ×	0.093**	0.428***	0.342***	0.628***	0.642***	0.506***	0.628***
Env score	(2.25)	(7.08)	(4.49)	(7.10)	(6.69)	(4.93)	(5.68)
Observations	49,054	48,879	48,699	48,518	48,337	48,160	47,995
R-squared	0.021	0.033	0.033	0.036	0.043	0.033	0.037
			Panel B: After December 2015				
Green sentiment ×	0.181***	0.476***	0.578***	0.445***	0.516***	0.422**	0.335*
Env score	(3.42)	(5.12)	(5.24)	(3.70)	(3.36)	(2.53)	(1.74)
Observations	46,193	45,092	44,004	42,925	41,861	40,808	39,760
R-squared	0.027	0.068	0.055	0.044	0.048	0.044	0.047
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: ***, **, and * indicate that the coefficient estimate is significantly different from zero at the 1%, 5%, and 10% level, respectively.

Source: authors' calculations.

Second, in Table 7 we re-run our regressions of individual stock returns by also including the interaction between the environmental score and the negative climate news index of Engle et al, aimed at capturing variation in (perceived) climate risk (the sample is reduced because Engle et al's measure is available only through May 2018). We find that, even when accounting for news-based climate risk, green sentiment is associated with a significant increase in the value of corporate environmental responsibility. We observe that the negative climate news measures has important effects on stock prices, in line with what documented by Engle et al. The results are particularly striking when recalling that *Green sentiment* and *Negative climate news* correlate negatively with each other (see Table 3), suggesting that they affect stock prices through different channels.

Green sentiment and stock prices: Accounting for negative climate news — Table 7

Dependent variable:	Return in t	Cumulative return through:					
		t+1	t+2	t+3	t+4	t+5	t+6
Green sentiment × Env score	0.072*	0.480***	0.442***	0.686***	0.722***	0.660***	0.738***
	(1.89)	(8.84)	(6.96)	(9.24)	(9.12)	(6.98)	(6.73)
Negative climate news ×Env score	0.049*	0.221***	0.287***	0.378***	0.363***	0.356***	0.388***
	(1.89)	(5.00)	(4.82)	(5.07)	(3.78)	(3.26)	(3.06)
Env score	0.105***	0.241***	0.348***	0.477***	0.526***	0.575***	0.604**
	(3.03)	(3.55)	(3.42)	(3.46)	(3.05)	(2.77)	(2.49)
Green sentiment	-1.249***	-1.413***	-1.433***	-1.438***	-2.076***	-1.385***	-1.576***
	(-30.48)	(-25.55)	(-22.27)	(-19.52)	(-24.78)	(-14.46)	(-14.95)
Negative climate news	-0.339***	-0.388***	-0.688***	-0.577***	-0.936***	-0.970***	-1.288***
	(-11.52)	(-7.73)	(-10.27)	(-7.08)	(-9.22)	(-8.29)	(-9.39)
Observations	73,280	72,986	72,691	72,397	72,111	71,833	71,563
R-squared	0.020	0.018	0.015	0.015	0.021	0.016	0.020
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: t-statistics based on standard errors clustered at the firm level are reported in parentheses. ***, **, and * indicate that the coefficient estimate is significantly different from zero at the 1%, 5%, and 10% level, respectively.

Source: authors' calculations.

Third, we shift our attention from stock prices to analysts' earnings forecasts, which are meant to reflect changes in cash flows expectations (Brown and Rozeff (1978)). Although stock prices and forecast revisions are generally highly positively correlated (Kothari et al (2016)), we expect green sentiment to drive a divergence of the two dimensions, ie to cause an increase in stock prices that is not accompanied by a positive update of earnings forecasts. For this exercise, we retrieve data on earnings forecasts from the IBES Summary Statistics database, which provides snapshots as of the day before the third Friday of each month of individual firms' expected earnings per share (EPS) at different horizons. For each firm-month observation, we compute the monthly change in average earnings forecasts, Δ *EPS forecast,* at one-, two-, and three-year horizons as done, eg in Landier and Thesmar (2020).[15]

[15] Specifically, for each horizon h and firm i, we compute the earnings revisions as $\Delta EPSForecast_h = E_{t+1}(EPS_{i,h}) - E_t(EPS_{i,h})$ scaled by 100, when $E_t(EPS_{i,h}) > 0$. We trim the resulting values at the 1st and 99th percentiles. The horizon is computed on the basis of the distance between the forecast's statistical period (variable "statpers") of the end date of the accounting period covered by the forecast (variable "fpedats"): one year (fiscal year ending between one and 12 months after the forecast's statistical period), two years (fiscal year ending between one and two years after the forecast's statistical period), three years (fiscal year ending between two and three years after the forecast's statistical period).

Green sentiment and analysts' forecast revisions Table 8

	(1)	(2)	(3)
Dependent variable:		Δ EPS forecast	
Horizon:	1-year ahead	2-year ahead	3-year ahead
Green sentiment ×Env score	-0.017	-0.004	0.048
	(-0.45)	(-0.15)	(1.51)
Negative climate news × Env score	0.032	0.066***	0.064**
	(1.06)	(2.79)	(2.52)
Env score	-0.116	-0.067	-0.072
	(-1.56)	(-1.18)	(-1.43)
Observations	61,055	62,102	58,608
R-squared	0.026	0.045	0.035
Firm controls	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes

Note: t-statistics based on standard errors clustered at the firm level are reported in parentheses. ***, **, and * indicate that the coefficient estimate is significantly different from zero at the 1%, 5%, and 10% level, respectively.

Source: authors' calculations.

Table 8 shows the results of OLS regressions of forecast revisions between months *t* and *t+1* on green sentiment in month *t* interacted with firms' environmental scores, controlling for firm characteristics, as well as sector and month fixed effects. The regressions also include the interaction between the environmental score and Engle et al's negative climate news index. As expected, green sentiment does not appear to have any explanatory power on the revisions of earnings forecasts, despite its effects on stock returns. Conversely, the Engle et al negative climate news index is associated with a statistically significant increase in the average forecast at the two-year and three-year horizons.

3.4 Green sentiment vs. ETF price pressure

Throughout the paper, we use the observed abnormal flows into green ETFs as a proxy for market-wide green sentiment. In particular, we argue that the observed stock price effect of green sentiment is the result of changes in investor appetite for environmental responsibilities, and not merely of the price pressure exerted directly by the ETF arbitrage activity, the propagation mechanism identified by Ben-David et al (2018).

To rule out the possibility that our results are mechanically driven by (green) ETFs' arbitrage activity, in Table 9, we replicate our main regressions by interacting the green sentiment index also with the percentage of common stocks held by green ETFs (*Green ETF ownership*). To compute this variable, we first retrieve green ETFs' portfolio holdings from the CRSP survivor-bias-free US mutual fund database. For each stock-month observation, we then divide the sum of green ETFs holdings in USD over the total market capitalisation.

Green sentiment and direct price pressure from green ETFs							Table 9
Dependent variable:	Return in t	Cumulative return through					
		t+1	t+2	t+3	t+4	t+5	t+6
Green sentiment ×	0.069**	0.273***	0.292***	0.445***	0.500***	0.403***	0.533***
Env score	(2.41)	(5.33)	(4.28)	(5.59)	(5.39)	(3.65)	(4.19)
Green sentiment ×	-0.287	0.003	0.153	-0.687	-0.773	-0.907	-0.905
Green ETF ownership	(-0.60)	(0.01)	(0.16)	(-0.56)	(-0.44)	(-0.42)	(-0.33)
Env score	0.044	0.042	0.112	0.207	0.176	0.201	0.211
	(1.29)	(0.61)	(1.10)	(1.56)	(1.07)	(1.03)	(0.92)
Green sentiment	-1.150***	-2.160***	-2.246***	-2.266***	-2.561***	-2.094***	-1.939***
	(-36.21)	(-40.56)	(-33.85)	(-28.88)	(-27.98)	(-19.93)	(-16.75)
Green ETF ownership	-2.014***	-3.988***	-6.184***	-8.413***	-10.024***	-11.873***	-14.615***
	(-4.86)	(-4.44)	(-4.63)	(-5.03)	(-4.63)	(-4.53)	(-4.43)
Observations	95,248	93,972	92,704	91,444	90,199	88,969	87,756
R-squared	0.019	0.035	0.030	0.029	0.033	0.026	0.026
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: t-statistics based on standard errors clustered at the firm level are reported in parentheses. ***, **, and * indicate that the coefficient estimate is significantly different from zero at the 1%, 5%, and 10% level, respectively.

Source: authors' calculations.

Green sentiment does not appear to predict any stock return effect on green ETFs' constituents, at least not over our sample period. On the contrary, green sentiment continues to have a significant predictive power on the pricing of environmental responsibility, in line with our main results.

3.5 Additional robustness checks

This subsection investigates the robustness of the stock price effects of green sentiment in four relevant dimensions.

First, Appendix Table A1 shows that our estimates remain statistically significant even when we double-cluster standard errors at the firm and the month levels to allow for potential correlation of residuals across both dimensions (Petersen (2009); Thompson (2011)).[16]

Second, Appendix Table A2 shows that our results are robust to including month fixed effects, to account for potential effect of macroeconomic conditions on

[16] With this double-clustering, the specification nests the classical Fama-MacBeth procedure (Fama and MacBeth (1967)) which controls for time effect in the correlation of residuals, but not for potential firm effect.

the pricing of environmental responsibility.[17] Notice that in these regressions the direct effect of *Green sentiment* is absorbed by the month indicators.

Third, given the diffuse concerns on the disagreement between ESG scores from different providers (Berg et al (2022); Gibson et al (2021)), in Appendix Table A3 we replicate our results using the alternative definition of environmental responsibility based on the MSCI KLD data set (*Env score (kld)*).[18] Note that in these regressions the sample is considerably larger given the broader coverage offered by the MSCI KLD database. Despite these differences, we obtain regression estimates that are statistically and economically similar to the ones obtained with our main proxy of environmental responsibility.

4. Effects of green sentiment on corporate behaviour

One of the most common narratives in the ESG industry is that sustainable investing can trigger positive societal change by influencing a firm's cost of capital, which in turn should allow more socially responsible firms to make more and better investments than other firms.

The above "financing channel mechanism" of responsible investing is identified and discussed in several theoretical works (Heinkel et al (2001); Pástor et al (2020); Oehmke and Opp (2020); Landier and Lovo (2020); De Angelis et al (2020)), but related empirical evidence remains scarce. In this section, we exploit the properties of our ETF-based Green Sentiment Index to shed light on the effects of investor non-fundamental demand shocks for green assets on corporate behaviour.

4.1 Main results on corporate behaviour

We focus on two important corporate decisions: investment and saving, by examining the impact of green sentiment on the levels of capital investments and cash holdings, which are useful for precautionary (Bates et al (2009); Almeida et al (2014)) and repurchase (Wang and Nyborg (2021)) motives, as well as to finance future investment (Bolton et al (2013)).

Based on the Compustat Accounting Quarterly database, we compute the variable *Capex/PPE* as the percentage of capital investments scaled by lagged Property, Plant and Equipment (capexq ×100/L1.ppentq), the variable *Cash/Assets* as the percentage of cash holdings over total assets (chq×100/atq), and the variable *R&D/Assets* as the percentage of research and development expenses over total assets (xrdq×100/atq).[19] We trim these variables at the 1–99 percentiles to control for extreme values. For the purposes of this analysis, we bring our data from the monthly

17 For instance, Bansal et al (2018) argue that stocks of socially responsible firms outperform in good economic times, whereas Lins et al (2017) and Albuquerque et al (2020) provide evidence that stocks of socially responsible firms performance relatively well in crisis times.

18 In our sample, the environmental scores from Sustainalytics and from MSCI KLD have a correlation of .58, statistically significant at the 1% level.

19 We normalise capex by lagged property, plant and equipment following Dessaint et al (2019). However, our results are robust to normalising capex by lagged total assets.

to the quarterly level. Summary statistics on *Capex/PPE*, *Cash/Assets*, and *R&D/Assets* are reported in Table 4.

In Table 10, we report the results of OLS regressions of quarterly capex (column 1), cashholdings (column 2), and R&D (column 3) on the average quarterly green sentiment, the firm's environmental score, and the interaction of the two. The regressions also control for firm characteristics and sector fixed effects (we obtain similar results when also adding quarter fixed effects, absorbing the direct effect of the green sentiment index). Standard errors are clustered at the firm level to account for the correlation of error terms across firms.[20]

Effects of green sentiment on real corporate decisions — Table 10

Dependent variable:	Capex/PPE	Cash/Assets	R&D/Assets
Green sentiment (q) ×Env score	0.214***	0.315***	-0.006
	(2.94)	(3.03)	(-0.35)
Env score	-0.233***	-0.089	-0.081
	(-2.89)	(-0.44)	(-1.27)
Green sentiment (q)	-1.108***	-0.085	0.042*
	(-13.69)	(-0.81)	(1.80)
Leverage	-0.007	-0.083***	-0.016***
	(-1.10)	(-6.04)	(-5.75)
Market beta	-0.060	0.637**	0.156*
	(-0.43)	(2.12)	(1.72)
Log(marketcap)	-0.035	-0.823***	-0.016
	(-0.46)	(-4.41)	(-0.24)
Book-to-market	-1.149***	-4.560***	-1.299***
	(-4.97)	(-6.64)	(-4.10)
Profitability	-0.009	0.038	-0.057***
	(-0.79)	(0.83)	(-5.28)
Momentum	-0.040	0.098**	0.015
	(-1.54)	(2.52)	(1.44)
Constant	5.371***	19.854***	2.478***
	(6.79)	(9.66)	(3.61)
Observations	23,569	30,018	14,136
R-squared	0.031	0.281	0.475
Industry FE	Yes	Yes	Yes

Note: t-statistics based on standard errors clustered at the firm level are reported in parentheses. ***, **, and * indicate that the coefficient estimate is significantly different from zero at the 1%, 5%, and 10% level, respectively.

Source: authors' calculations.

The results indicate that higher green sentiment in a given quarter is associated with higher capital investments and accumulation of cash for more environmentally responsible firms, consistent with the idea that firms make more investments and hold larger cash balances when access to funds is easier (eg Dittmar et al (2003)). The effect is economically important: a one standard deviation stronger green sentiment is associated with 0.21% higher capex and 0.31% higher cash holdings for a one

[20] The estimated coefficients remain statistically significant even if we double-cluster standard errors at both the firm and quarter levels (Petersen (2009); Thompson (2011)).

standard deviation higher environmental score (compared with an average capex of 9% and cash holdings of 4%, this represents respectively a 5% and 3% relative increase). We do not observe any effect of green sentiment on green firms' R&D activity.

4.2 Heterogeneity across credit ratings

What types of firm is green sentiment more likely to influence? The existing literature suggests that managers of more equity-dependent/credit-constrained firms are more likely to be influenced by stock prices in their decision-making (Baker et al (2003); Hau and Lai (2013)). We expect this principle to apply also in the context of the real impacts of responsible investing. Intuitively, the effects of green sentiment on corporate decisions should vary with a firm's ability to raise funds outside the stock market.[21]

To test for the heterogeneity of the effect of green sentiment on real corporate decisions, we split our sample on the basis of corporate credit ratings, which we use as a proxy of the firm's ability to access external capital on credit markets. We retrieve corporate long-term S&P credit ratings from Bloomberg, and we classify them in three groups: *Low credit rating* < "BBB-" (non-investment grade); *Middle credit rating* = "BBB", "BBB+", "BBB–"; *High credit rating* = "A","A+","A–","AA", "AAA","AA–","AA+".

Table A4 in the Appendix shows the number of firms × quarters with above- and below-median environmental score in each of the three credit rating groups. Not surprisingly, we observe a positive correlation between environmental responsibility and credit ratings, consistent with the evidence in Seltzer et al (2021). For instance, we find that firms with a high environmental score have a likelihood of 34% to also have a high credit rating, versus only 16% among firms with low environmental score.[22]

Table 11 shows the heterogeneity of the effects of green sentiment on firm behaviour along the credit rating. We obtain two intriguing results. First, in Panel A, we show that the effect of green sentiment on capital investment is concentrated in firms with low and medium credit ratings. No significant effect is observed for companies with a strong credit rating. This result is consistent with the idea that less equity-dependent firms are less influenced by stock prices in making investment decisions (Baker et al (2003)). Our results indicate that these firms are also less likely to be influenced by green sentiment. Second, in Panel B, we observe that the effect of green sentiment on cash holdings is driven primarily by the subsamples of low- (and to a less extent) highly rated firms, consistent with the idea that cash holdings are more valuable for financially constrained firms (Denis and Sibilkov (2010)). Conceivably, these firms take advantage of green sentiment to increase their precautionary buffers.

21 For instance, the model in Landier and Lovo (2020) suggests that, in order for ESG funds to force companies to partially internalise externalities, it is necessary to have significant frictions in financial markets.

22 Indeed, the environmental score even correlates positively with the likelihood of having the credit rating available in the first place, causing Table A4 to show relatively more firms with an above-median environmental score.

Green sentiment, corporate investments, and credit ratings

Table 11

	Low credit rating	Medium credit rating	High credit rating
Panel A: Dependent variable: Capex/PPE			
Green sentiment (q) ×Env score	0.546***	0.249**	0.242
	(2.89)	(2.24)	(1.60)
Env score	0.004	-0.094	-0.047
	(0.02)	(-0.74)	(-0.32)
Green sentiment (q)	-1.310***	-1.184***	-1.142***
	(-7.13)	(-9.26)	(-5.68)
Observations	4,464	7,629	4,331
R-squared	0.043	0.040	0.032
Panel B: Dependent variable: Cash/Assets			
Green sentiment (q) ×Env score	0.575***	0.030	0.257*
	(2.96)	(0.25)	(1.66)
Env score	0.243	0.519*	-0.027
	(0.73)	(1.91)	(-0.09)
Green sentiment (q)	-0.222	-0.146	-0.158
	(-1.25)	(-1.03)	(-0.95)
Observations	5,526	10,003	5,338
R-squared	0.263	0.362	0.371
Firm controls	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes

Note: t-statistics based on robust standard errors in parentheses. ***, **, and * indicate that the coefficient estimate is significantly different from zero at the 1%, 5%, and 10% level, respectively.

Source: Authors' calculations.

5. Conclusion

In the past decade, environmental considerations have taken centre stage in shaping the debate in global financial markets, a trend that is set to continue for many years to come. How do changes in investors' appetite for green assets influence the allocation of capital in the economy and the behaviour of firms? This question is a key policy issue, as many policymakers and regulators expect the redirection of capital market financing towards green firms to have a decisive impact on carbon emissions (eg Lagarde (2021)) and firms' cost of capital (eg Scatigna et al (2021)).

Studying the effects of investors' environmental preferences on economic outcomes is an empirical challenge due to the entanglement of fundamental and non-fundamental factors in driving firm value. In this paper, we proposed a new method for estimating non-fundamental demand shocks for green financial assets. The method exploits the unique arbitrage mechanism of ETFs' primary market, which the existing literature shows to be influenced by non-fundamental demand shocks due to differences in the ownership structure of ETFs compared with the underlying securities (Ben-David et al (2017); Brown et al (2023)). Specifically, using a comprehensive sample of US ETFs over the period from January 2010 through June 2020, we estimate the monthly excess flows into ETFs with explicit environmentally friendly features (green ETFs) relative to comparable conventional ETFs. The time series of these abnormal green flows – which we name the Green Sentiment Index – quantifies the direction and magnitude of changes in investors' appetite for green financial assets that are not priced in the value of the underlying securities.

After establishing the difference between our index and other climate-related measures used in the literature, we study the effects of green sentiment on the pricing of corporate environmental responsibility in the stock market and on corporate decisions. Using a sample of US firms over the period January 2010 to June 2020, we establish two key results.

First, we show that higher green sentiment is associated with a stock price outperformance of environmentally responsible firms. A one standard deviation stronger green sentiment in month t is followed by approximately 27 basis points higher returns in $t+1$ for a one standard deviation higher environmental score. The estimated outperformance considering returns through in $t+6$ is 53 basis points. A series of tests confirm that this effect does not reflect fundamental information: green sentiment predicts stock prices both before and after the signature of the Paris Agreement, a structural break in climate transition risks; its stock price effects are independent to variations in climate risk, as proxied by the negative climate news index used in Engle et al; finally, despite the fact that they both have similar stock price effects, Engle et al's index leads to positive revisions in analysts' earnings forecasts for environmentally responsible firms, while our green sentiment index does not.

Second, we document that an increase in green sentiment also affects corporate decisions. In quarters with strong green sentiment, environmentally responsible firms make higher capital investments (particularly firms with low and medium credit ratings) and accumulate more cash holdings (particularly firms with low and high credit ratings). The role of financial constraints in mediating the impact of (green) sentiment on corporate behaviour is in line with previous works on the real effects of financial markets (eg Baker et al (2003); Campello and Graham (2013)) and the theoretical literature on responsible investing. It is reasonable to expect that companies are more likely to do "good" when they are less financially constrained (Cohn and Wardlaw (2016); Hong et al (2012); Martin et al (2021)). In this sense, by (further) increasing the financial strength of environmentally responsible firms, green sentiment allows them to further increase their environmentally friendly investments. At the same time, we should also be aware of the risk that green sentiment may inadvertently divert resources away from firms that are not currently considered green but have high green innovation potential (Cohen et al (2020)).

How to encourage firms to contribute to the development of technologies useful to decarbonise our economies is a key question of our times.[23] While the effects of governmental policies (such as carbon pricing) and the role of public finance are more researched and understood (eg Aghion et al (2016); Gollier (2021)), the role of financial markets in stimulating firms to make more investments in green projects and technologies deserves investigation. Our results suggest that green sentiment

[23] Almost half of the emission reductions that are needed to reach the climate-neutrality goal by 2050 are expected to come from technologies that still need to be developed (International Energy Agency (2021)).

can reduce the relative cost of capital of more environmentally responsible firms, and increase their investment capacity. How exactly firms make use of these extra resources is a critical issue that we leave for future research.

References

Aghion, P, A Dechezlepr̂etre, D Hemous, R Martin and J Van Reenen (2016): "Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry", *Journal of Political Economy*, vol 124, no 1, pp 1–51.

Albuquerque, R, Y Koskinen, S Yang and C Zhang (2020): "Resiliency of environmental and social stocks: An analysis of the exogenous COVID-19 market crash", *The Review of Corporate Finance Studies*, vol 9, no 3, pp 593–621.

Almeida, H, M Campello, I Cunha and M Weisbach (2014): "Corporate liquidity management: A conceptual framework and survey", *Annual Review of Financial Economics*, vol 6, pp 135–62.

Anderson, A and D Robinson (2020): "Knowledge, fear and beliefs: Understanding household demand for green investments", Swedish House of Finance, Working Paper.

Baker, M, J Stein and J Wurgler (2003): "When does the market matter? Stock prices and the investment of equity-dependent firms", *The Quarterly Journal of Economics*, vol 118, no 3, pp 969–1005.

Baker, M and J Wurgler (2000): "The equity share in new issues and aggregate stock returns", *The Journal of Finance*, vol 55, no 5, pp 2219–57.

——— (2006): "Investor sentiment and the cross-section of stock returns", *The Journal of Finance*, vol 61, no 4, pp 1645–80.

Baker, M, J Wurgler and Y Yuan (2012): "Global, local, and contagious investor sentiment", *Journal of Financial Economics*, vol 104, no 2, pp 272–87.

Bakke, T-E and T Whited (2010): "Which firms follow the market? An analysis of corporate investment decisions", *The Review of Financial Studies*, vol 23, no 5, pp 1941–80.

Bansal, R, D Wu and A Yaron (2018): "Is socially responsible investing a luxury good?", Duke University, Working Paper.

Bartram, S, K Hou and S Kim (2022): "Real effects of climate policy: Financial constraints and spillovers", *Journal of Financial Economics*, vol 143, no 2, pp 668–696

Bates, T, K Kahle and R Stulz (2009): "Why do US firms hold so much more cash than they used to?", *The Journal of Finance*, vol 64, no 5, pp 1985–2021.

Battiston, S, I Monasterolo, K Riahi and B van Ruijven (2021): "Accounting for finance is key for climate mitigation pathways", Science, vol 372, no 6545, pp 918–20.

Ben-David, I, F Franzoni, K Byungwook and R Moussawi (2023): "Competition for attention in the ETF space", *The Review of Financial Studies*, vol 36, no 3, pp 987–1042.

——— (2017): "Exchange-traded funds", *Annual Review of Financial Economics*, vol 9, pp 169–89.

——— (2018): "Do ETFs increase volatility?", *The Journal of Finance*, vol 73, no 6, pp 2471–535.

Ben-Rephael, A, X Dong, M Massa and C Zhou (2019): "Foreign sentiment", mimeo.

Ben-Rephael, A, S Kandel and A Wohl (2012): "Measuring investor sentiment with mutual fund flows", *Journal of Financial Economics*, vol 104, no 2, pp 363–82.

Berg, F, J Koelbel and R Rigobon (2022): "Aggregate confusion: The divergence of ESG ratings", *Review of Finance*, vol 26, no 6, pp 1315–44.

Bessec, M and Julien F (2020): "Green sentiment in financial markets: A global warning", Proceedings of Paris December 2020 Finance Meeting EUROFIDAI – ESSEC, Université Paris-Dauphine Research Paper, no 3710489.

Bessembinder, H, M Cooper and F Zhang (2019): "Characteristic-based benchmark returns and corporate events", *The Review of Financial Studies*, vol 32, no 1, pp 75–125.

Bingler, J, Mathias K and M Leippold (2022): "Cheap talk and cherry-picking: What climatebert has to say on corporate climate risk disclosures", *Finance Research Letters*, vol 47.

Bolton, P, H Chen and N Wang (2013): "Market timing, investment and risk management", *Journal of Financial Economics*, vol 109, no 1, pp 40–62.

Bolton, P and M Kacperczyk (2022): "Do investors care about carbon risk?", *Journal of Financial Economics*, vol 142, no 2, pp 517–49.

——— (2021): "Global pricing of carbon-transition risk", *NBER Working Papers*, no 28510.

Bond, P, A Edmans and I Goldstein (2012): "The real effects of financial markets", *Annual Review of Financial Economics*, vol 4, pp 339–60.

Brown, D, S Davies and M Ringgenberg (2021): "ETF arbitrage, non-fundamental demand, and return predictability", *Review of Finance*, vol 25, no 4, pp 937–72.

Brown, L and M Rozeff, (1978): "The superiority of analyst forecasts as measures of expectations: Evidence from earnings", *The Journal of Finance*, vol 33, no 1, pp 1–16.

Campello, M and J Graham (2013): "Do stock prices influence corporate decisions? Evidence from the technology bubble", *Journal of Financial Economics*, vol 107, no 1, pp 89–110.

Ceccarelli, M, S Ramelli and A Wagner (2023): "Low-carbon mutual funds", *Review of Finance*, forthcoming.

Choi, D, Z Gao and W Jiang (2020): "Attention to global warming", *The Review of Financial Studies*, vol 33, no 3, pp 1112–45.

Cohen, L, U Gurun and Q Nguyen (2020): "The ESG-innovation disconnect: Evidence from green patenting", *NBER Working Papers*, no 27990.

Cohn, J and M Wardlaw (2016): "Financing constraints and workplace safety", *The Journal of Finance*, vol 71, no 5, pp 2017–58.

Da, Z, J Engelberg and P Gao (2015): "The sum of all FEARS investor sentiment and asset prices", T*he Review of Financial Studies*, vol 28, no 1, pp 1–32.

Davies, S (2022): "Speculation sentiment", *Journal of Financial and Quantitative Analysis*, vol 57, no 7, pp 2485–515.

De Angelis, T, P Tankov and O Zerbib (2020): "Environmental impact investing", Working Paper.

Delis, M, K de Greiff and S Ongena (2019): "Being stranded with fossil fuel reserves? Climate policy risk and the pricing of bank loans", *Swiss Finance Institute Research Papers*, no 18-10.

Denis, D and V Sibilkov (2010): "Financial constraints, investment, and the value of cash holdings", *The Review of Financial Studies*, vol 23, no 1, pp 247–69.

Dessaint, O, T Foucault, L Frésard and A Matray (2019): "Noisy stock prices and corporate investment", *The Review of Financial Studies*, vol 32, no 7, pp 2625–72.

Dittmar, A, J Mahrt-Smith and H Servaes (2003): "International corporate governance and corporate cash holdings", *Journal of Financial and Quantitative Analysis*, vol 38, no 1, pp 111–33.

Engle, R, A Giglio, B Kelly, H Lee and J Stroebel (2020): "Hedging climate change news", *The Review of Financial Studies*, vol 33, no 3, pp 1184–216.

Faccini, R, R Matin and G Skiadopoulos (2021): "Dissecting climate risks: Are they reflected in stock prices?", *Journal of Banking & Finance*, vol 152, July.

Fama, E and K French (2007): "Disagreement, tastes, and asset prices", *Journal of Financial Economics*, vol 83, no 3, pp 667–89.

Fama, E and J MacBeth (1967): "Risk, return, and equilibrium empirical tests", *Journal of Political Economy*, vol 81, no 3, pp 607–36.

Garel, A and A Petit-Romec (2021): "Investor rewards to environmental responsibility: Evidence from the COVID-19 crisis", *Journal of Corporate Finance*, vol 68.

Gibson, R, P Krueger and P Schmidt (2021): "ESG rating disagreement and stock returns", *Financial Analysts Journal*, vol 77, no 4, pp 104–27.

Glosten, L, S Nallareddy and Y Zou (2021): "ETF activity and informational efficiency of underlying securities", *Management Science*, vol 67, no 1, pp 22–47.

Gollier, C (2021): "The welfare cost of ignoring the beta", Fondazione Eni Enrico Mattei, *FEEM Working Papers*, no 309916.

Gollier, C and S Pouget (2014): "The 'washing machine': Investment strategies and corporate behavior with socially responsible investors", Toulouse School of Economics, Working Paper.

Gourdel, R, I Monasterolo, N Dunz, A Mazzocchetti and L Parisi (2021): "Assessing the double materiality of climate risks in the EU economy and banking sector", ECB Working Paper Series, no 2665.

Hau, H and S Lai (2013): "Real effects of stock underpricing", *Journal of Financial Economics*, vol 108, no 2, pp 392–408.

Heinkel, R, A Kraus and J Zechner (2001): "The effect of green investment on corporate behavior", *Journal of Financial and Quantitative Analysis*, vol 36, no 4, pp 431–49.

Henderson, B, N Jegadeesh and M Weisbach (2006): "World markets for raising new capital", *Journal of Financial Economics*, vol 82, no 1, pp 63–101.

Hong, H and M Kacperczyk (2009): "The price of sin: The effects of social norms on markets", *Journal of Financial Economics*, vol 93, no 1, pp 15–36.

Hong, H, J Kubik and J Scheinkman (2012): "Financial constraints on corporate goodness", *NBER Working Papers*, no 18476.

Huynh, T and Y Xia (2021): "Climate change news risk and corporate bond returns", *Journal of Financial and Quantitative Analysis*, vol 56, no 6, pp 1985–2009.

Ilhan, E, Z Sautner and G Vilkov (2021): "Carbon tail risk", *The Review of Financial Studies*, vol 34, no 3, pp 1540–71.

International Energy Agency, IEA (2021): *Net zero by 2050: A roadmap for the global energy sector*, www.iea.org.

Investment Company Institute (2021): *Investment company fact book*, www.ici.org.

Kelly, B, S Pruitt and Y Su (2019): "Characteristics are covariances: A unified model of risk and return", *Journal of Financial Economics*, vol 134, no 3, pp 501–24.

Kim, W and M Weisbach (2008): "Motivations for public equity offers: An international perspective", *Journal of Financial Economics*, vol 87, no 2, pp 281–307.

Kothari, S, E So and R Verdi (2016): "Analysts' forecasts and asset pricing: A survey", *Annual Review of Financial Economics*, vol 8, no 1, pp 197–219.

Krüger, P (2015): "Corporate goodness and shareholder wealth", *Journal of Financial Economics*, vol 115, no 2, pp 304–29.

Krüger, P, Z Sautner and L Starks (2020): "The importance of climate risks for institutional investors", *The Review of Financial Studies*, vol 33, no 3, pp 1067–111.

Lagarde, C (2021): "Towards a green capital market for Europe", speech at the European Commission's High-Level Conference on the Proposal for a Corporate Sustainability Reporting Directive, 6 May, www.ecb.europa.eu/press/key/date/2021/html/ecb.sp210506~4ec98730ee.en.html.

Landier, A and S Lovo (2020): "ESG investing: How to optimize impact? ", Working Paper.

Landier, A and D Thesmar (2020): "Earnings expectations in the COVID crisis", *Review of Corporate Finance Studies*, vol 4, pp 598–617.

Lemmon, M and E Portniaguina (2006): "Consumer confidence and asset prices: Some empirical evidence", *The Review of Financial Studies*, vol 19, no 4, pp 1499–529.

Lettau, M and A Madhavan (2018): "Exchange-traded funds 101 for economists", *Journal of Economic Perspectives*, vol 32, no 1, pp 135–54.

Li, Q, H Shan, Y Tang and V Yao (2020): "Corporate climate risk: Measurements and responses", Working Paper.

Lins, K, H Servaes and A Tamayo (2017): "Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis", *The Journal of Finance*, vol 72, no 4, pp 1785–1824.

Luo, H and R Balvers (2017): "Social screens and systematic investor boycott risk", *Journal of Financial and Quantitative Analysis*, vol 52, no 1, pp 365–99.

Luo, Y (2005): "Do insiders learn from outsiders? Evidence from mergers and acquisitions", *The Journal of Finance*, vol 60, no 4, pp 1951–82.

Martin, R, R De Haas, M Muuls and H Schweiger (2021): "Managerial and financial barriers to the net-zero transition", *EBRD Working Papers*, no 254.

Morck, R, A Shleifer, R Vishny, M Shapiro and J Poterba (1990): "The stock market and investment: Is the market a sideshow?", *Brookings Papers on Economic Activity*, pp 157–215.

Morningstar (2020): *Global sustainable fund flows: Q4 2020 in review*, www.morningstar.com/lp/global-esg-flows.

Oehmke, M and M Opp (2020): "A theory of socially responsible investment", *Swedish House of Finance Research Papers*, no 20-2.

Pagano, M, A Sánchez Serrano and J Zechner (2019): "Can ETFs contribute to systemic risk?", European Systemic Risk Board, Report of the Advisory Scientific Committee, no 9.

Pankratz, N and C Schiller (2019): "Climate change and adaptation in global supply-chain networks", Federal Reserve Board, *Finance and Economics Discussion Series*, no 56.

Pástor, L, R Stambaugh and L Taylor (2022): "Dissecting green returns", *Journal of Financial Economics*, vol 146, no 2, pp 403–24.

——— (2020): "Sustainable investing in equilibrium", *Journal of Financial Economics*, vol 142, no 2, pp 550–71.

Pástor, L and M Blair Vorsatz (2020): "Mutual fund performance and flows during the COVID-19 crisis", *The Review of Asset Pricing Studies*, vol 10, no 4, pp 791–833.

Pedersen, L, S Fitzgibbons and L Pomorski (2020): "Responsible investing: The ESG-efficient frontier", *Journal of Financial Economics*, forthcoming.

Petersen, M (2009): "Estimating standard errors in finance panel data sets: Comparing approaches", *The Review of Financial Studies*, vol 22, no 1, pp 435–80.

Qiu, L and I Welch (2004): "Investor sentiment measures", Working Paper.

Ramelli, S, A Wagner, R Zeckhauser and A Ziegler (2021): "Investor rewards to climate responsibility: Stock-price responses to the opposite *shocks of the 2016 and 2020 U.S. elections", Review of Corporate Financial Studies,* vol 10, no 4, pp 748–87.

Santi, C (2023): "Investors' climate sentiment and financial markets", *International Review of Financial Analysis*, vol 86, March.

Sautner, Z, L van Lent, G Vilkov and R Zhang (2023): "Firm-level climate change exposure", *Journal of Finance*, forthcoming.

Scatigna, M, D Xia, A Zabai and O Zulaica (2021): "Achievements and challenges in ESG markets", *BIS Quarterly Review,* December.

Seltzer, L, L Starks and Q Zhu (2021): "Climate regulatory risks and corporate bonds", Federal Reserve Bank of New York, Staff Reports, no 1014.

Stambaugh, R (2014): "Presidential address: Investment noise and trends", *The Journal of Finance*, vol 69, no 4, pp 1415–53.

Stambaugh, R, J Yu and Y Yuan (2012): "The short of it: Investor sentiment and anomalies", *Journal of Financial Economics*, vol 104, no 2, pp 288–302.

Thompson, S (2011): "Simple formulas for standard errors that cluster by both firm and time", *Journal of Financial Economics*, vol 99, no 1, pp 1–10.

Van der Beck, P (2021): "Flow-driven ESG returns", *Swiss Finance Institute Research Papers*, no 21-71.

Wang, J and K Nyborg (2021): "The effect of stock liquidity on cash holdings: The repurchase motive", *Journal of Financial Economics*, vol 42, no 2, pp 905–27.

Xia, D and O Zulaica (2022): "The term structure of carbon premia", *BIS Working Papers*, no 1045.

Zerbib, O (2022): "A sustainable capital asset pricing model (S-CAPM): Evidence from environmental integration and sin stock exclusion", *Review of Finance*, vol 26, no 6, pp 1345–88.

Appendix

Green sentiment and stock prices: Double-clustering standard errors — Table A1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:	Return in t	Cumulative return through:					
		t+1	t+2	t+3	t+4	t+5	t+6
Green sentiment ×	0.068	0.272*	0.290**	0.440***	0.495**	0.397**	0.527**
Env score	(0.88)	(1.67)	(2.41)	(2.75)	(2.47)	(2.00)	(2.10)
Env score	0.039	0.033	0.098	0.189	0.153	0.175	0.179
	(0.47)	(0.26)	(0.59)	(1.02)	(0.68)	(0.67)	(0.63)
Green sentiment	-1.153***	-2.163**	-2.248***	-2.275***	-2.572***	-2.106**	-1.952*
	(-2.88)	(-2.59)	(-2.67)	(-2.93)	(-3.37)	(-2.48)	(-1.93)
Leverage	0.003	0.004	0.004	-0.001	0.000	0.000	0.001
	(0.44)	(0.59)	(0.44)	(-0.07)	(0.00)	(0.02)	(0.08)
Market beta	0.245	0.178	0.158	-0.210	-0.345	-0.544	-0.729
	(0.96)	(0.53)	(0.37)	(-0.49)	(-0.71)	(-1.01)	(-1.25)
Log(market cap)	0.006	0.011	-0.056	-0.163	-0.152	-0.201	-0.222
	(0.06)	(0.08)	(-0.32)	(-0.86)	(-0.68)	(-0.79)	(-0.79)
Book-to-market	-0.133	-0.405	-0.565	-0.682	-1.000	-1.011	-0.965
	(-0.51)	(-1.02)	(-1.04)	(-1.01)	(-1.24)	(-1.05)	(-0.86)
Profitability	-0.003	-0.017	-0.030	-0.048*	-0.070**	-0.090**	-0.112***
	(-0.27)	(-0.95)	(-1.30)	(-1.89)	(-2.35)	(-2.52)	(-2.69)
Momentum	-0.145	-0.255*	-0.292	-0.146	-0.147	-0.089	-0.027
	(-1.37)	(-1.73)	(-1.50)	(-0.81)	(-0.80)	(-0.40)	(-0.12)
Constant	0.857	2.131	3.824**	5.967***	7.168***	8.812***	10.216***
	(0.81)	(1.50)	(2.28)	(3.28)	(3.36)	(3.52)	(3.67)
Observations	95,248	93,972	92,704	91,444	90,199	88,969	87,756
R-squared	0.018	0.035	0.030	0.028	0.032	0.025	0.024
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Green sentiment and stock prices: Adding month fixed effects

Table A2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:	Return in t	Cumulative return through:					
		t+1	t+2	t+3	t+4	t+5	t+6
Green sentiment × Env score	0.028	0.165***	0.190***	0.210***	0.250***	0.249**	0.307**
	(0.96)	(3.44)	(3.01)	(2.68)	(2.68)	(2.26)	(2.42)
Env score	-0.048	-0.098	-0.111	-0.112	-0.122	-0.109	-0.105
	(-1.65)	(-1.60)	(-1.21)	(-0.91)	(-0.78)	(-0.58)	(-0.47)
Observations	95,248	93,972	92,704	91,444	90,199	88,969	87,756
R-squared	0.292	0.273	0.257	0.235	0.221	0.214	0.202
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Green sentiment and stock prices: Alternative environmental score

Table A3

Dependent variable:	Return in t	Cumulative return through:					
		t+1	t+2	t+3	t+4	t+5	t+6
Green sentiment ×	0.063***	0.065*	0.075*	0.108**	0.169***	0.209***	0.326***
Env score (kld)	(3.52)	(1.91)	(1.72)	(2.14)	(2.79)	(3.06)	(4.27)
Env score (kld)	-0.047**	-0.107***	-0.182***	-0.227***	-0.265***	-0.254**	-0.254*
	(-2.30)	(-2.60)	(-3.07)	(-2.96)	(-2.79)	(-2.22)	(-1.91)
Green sentiment	-1.211***	-2.486***	-2.530***	-2.628***	-2.885***	-2.495***	-2.410***
	(-52.92)	(-64.40)	(-51.49)	(-44.84)	(-41.51)	(-31.27)	(-27.25)
Observations	232,622	228,156	223,806	219,560	215,394	211,315	207,340
R-squared	0.014	0.031	0.025	0.023	0.025	0.019	0.018
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Sample distribution by environmental score and credit rating

Table A4

	Credit rating			
Env score	Low	Medium	High	Total
Below or equal median	4,267	5,481	1,887	11,635
Above median	2,693	6,290	4,653	13,636
Total	6,960	11,771	6,540	25,271

The term structure of carbon premia

Fan Dora Xia and Omar Zulaica[1]

Abstract

This paper explores the carbon premium – the extra yield investors demand to buy bonds issued by firms with more greenhouse gas emissions – in the US corporate bond market. We analyse the carbon premium along two channels, via panel regression. One is the preference channel, under which the premium reflects investors' preference for firms that they perceive as being more environmentally responsible, all else equal. The other is the risk channel, where investors perceive more carbon-intensive firms as being more prone to default. We test the preference channel by investigating the relationship between corporate bond yields and carbon emissions, while controlling for proxies of the probability of default (PD) and for other bond characteristics. We examine the risk channel by analysing how carbon emissions affect the PD. We validate the existence of carbon premia in both channels, with the premium being larger for firms in more energy-intensive sectors. Moreover, the premium differs across maturities, giving rise to a hump-shaped term structure of carbon premia, reaching its highest level at the belly of the curve (maturities of 15–20 years).

JEL classification: G12, G30, Q54.

[1] Bank for International Settlements (BIS). The authors (dora.xia@bis.org, omar.zulaica@bis.org) would like to thank Pierre Cardon, Stijn Claessens, Ingo Fender, Frank Weikai Li, Mark Mason, Michela Scatigna, Hyun Song Shin, Ilhyock Shim, and participants of the BIS Research Meeting,the Banking Department's Reserve Management Seminar and the Seventh Public Investors Conference in Ottawa as well as the tenth ABFER Annual Conference in Singapore for helpful comments; and Adam Cap and Alberto Americo for excellent research assistance. The views expressed are those of the authors and do not necessarily reflect those of the BIS.

1. Introduction

It is an emerging consensus that shifting the global economy to a low-carbon growth path is essential. As more greenhouse gases (GHGs) are accumulated in the atmosphere, global temperatures will continue to rise, with some of the resulting effects being irreversible (IPCC (2022)). Global warming catalyses the frequency and intensity of natural disasters such as droughts and storms. Indeed, in the past three decades, adverse weather events have become more frequent on the back of higher global temperatures. These extreme weather events often cause widespread losses and damage to nature, people and economic activities.[2]

In pricing this transition, investors will likely demand compensation for investing in firms with higher carbon footprints.[3] Why? Investors seem to agree on carbon emissions as a reasonable proxy for gauging the exposure to climate-related transition risk.[4] And they may demand higher yields for financing companies that will be affected by the transition from fossil fuels to renewable energy, for example, thereby giving rise to a *carbon premium*.[5]

Any carbon premium may take into account the following two channels. The first one is the preference channel, reflecting that investors who want to support sustainable growth might have a preference, all else equal, for firms that they perceive as helping to achieve this goal. Seen conversely, investors may dislike firms they perceive as more harmful to the environment. This channel captures aspects of the investment process such as negative screening, which excludes firms that score poorly on environmental factors such as GHG emissions (Elsenhuber and Skenderasi (2019)).

The second is the credit risk channel (the "risk channel"), where investors perceive more carbon-intensive firms as more prone to default.[6] This is because these firms are likely to face larger transition risks, which are related to regulatory policies, advances in technology, and changes in consumer preferences that may impair their financial health. This channel captures practices in banks and credit rating agencies which explicitly take into account environmental factors – such as carbon emissions – in assigning risk grades on their scorecards.

It is important to examine how much of a carbon premium is priced into financial assets for at least two reasons. First, because financial markets can support the transition to a more sustainable economy by reallocating their resources towards economic activities that foster it, it is crucial to evaluate the extent to which this reallocation has affected asset pricing. Naturally, a precondition for this mechanism to work is that investors differentiate between financial assets that fund activities with

2 In addition to transition risk, firms also face physical risk – which directly results from the effects of climate change on economic activities. The threat of disrupted production from rising sea levels to factories located close to the sea is an example. Compared with transition risk, physical risk is harder to quantify. It depends in complex ways on firms' geospatial characteristics.

3 In this paper, we use the term carbon footprint to refer to greenhouse gas (GHG) emissions caused by a firm.

4 For a review, see Giglio et al (2021). That said, one shortcoming of using carbon emissions to quantify exposure to transition risk is the fact that they are backward-looking instead of forward-looking. Some forward-looking indicators are available but do not seem to be widely available/accepted yet.

5 Henceforth we use "carbon premium", "carbon risk premium" and "carbon transition risk premium" interchangeably.

6 We define a *more* carbon-intensive firm as one with *higher* CO_2 emissions than others.

different degrees of environmental impact. On the other hand, climate change can have a substantial impact on financial stability (Bolton et al (2020)). For instance, if carbon risk is not sufficiently priced in, financial assets are vulnerable to sharp re-pricings that could lead to systemic risk episodes. Estimating the carbon premium embedded in current prices sheds light on any risk of such re-pricing.

In this paper, we ask whether such a carbon premium exists in corporate bond prices – specifically, via their yield spread to the risk-free rate. The two channels described above are relatively straightforward to test in this market, given that spreads can be decomposed into a component related to default risk and one capturing all other relevant risk factors, offering us a foundation to dissect the preference and risk angles.[7] To do so, we measure a firm's carbon footprint through its GHG emissions, which we draw from the widely used S&P Global Trucost database. We test the preference channel by analysing the relationship between corporate bond spreads and carbon emissions, while controlling for default risk and other bond characteristics. We test the risk channel by looking into the impact of carbon emissions on the probability of default while filtering out the impact of other firm characteristics. Our analysis focuses on firms in the United States, as it is the jurisdiction best covered by the Trucost database. The market value of US companies covered by Trucost accounts for well over 90% of total US market capitalisation. In those same terms, the firms in the United States account for the largest share (around 40%) of the firms in the database.

Our main finding is that carbon emissions affect the spreads of corporate bonds issued by US firms via both the preference and risk channels.[8] When controlling for the probability of default, we find a positive and statistically significant carbon premium on firm-level scope 1 and scope 2 emissions.[9] We interpret this finding as the *credit risk-adjusted part* of the carbon premium. In a second step, we find a statistically significant and positive relationship between a firm's carbon emissions and its probability of default. We interpret this non-linear relationship as the *credit-risk part* of the carbon premium, which holds for all emission scopes.

Combining both preference and risk channels – credit risk-adjusted *and* credit-risk carbon premia – we can derive *total* carbon premia. For a typical firm in our sample, halving carbon emissions would narrow corporate spreads by around 2 to 4.5 basis points, with the more significant contribution coming from the credit risk-adjusted part of the premium – that is, the preference channel. The larger contribution from investor preferences makes sense at the current juncture. The simplest way to introduce a sustainable investment approach is through screening (ie securities are left outside an investment universe due to their more negative environmental impacts). To the best of our knowledge, frameworks for quantifying the impact of climate events on default risk remain under development.

7 Other advantages of focusing on corporate bonds include that downside risks from climate change are likely to matter more to bond investors than to equity investors. Moreover, investors in corporate bonds are more sophisticated and thus more likely to consider carbon risk. See Duan et al (2021) for details.

8 In addition to our main analysis, we investigate the impact of carbon intensities on spreads through the preference and risk channels, and report these results in Appendix 3. Specifically, we examine how carbon intensities affect corporate bond spreads, and show that these effects are significant through both channels.

9 Scope 1 emissions are direct emissions generated from a firm's activities, while scope 2 emissions are indirect emissions resulting from a firm's purchases of electricity, steam and heating/cooling.

Furthermore, carbon emissions are priced in bonds issued by both non-energy-intensive and energy-intensive firms, with larger impacts for the latter. In particular, through the preference channel, a 50% decrease in the sum of scope 1 and 2 emissions predicts a drop of 8.2 and 4.2 basis points in the spread of bonds issued by energy-intensive firms and non-energy-intensive firms, respectively. For the risk channel, a 50% reduction in either scope 1 or scope 1+2 emissions would reduce the probability of default of a typical firm in an energy-intensive sector by 6 basis points, which can be translated to around 2.4 basis point decline in option-adjusted spreads. The impact for a typical firm in non-energy-intensive sectors is around 2 basis points on the probability of default and around 0.6 basis points on option-adjusted spreads. Putting the two channels together, we find that the combined impact on the bond spread is around 7–13 basis points for firms in energy-intensive sectors, and less than 4 basis points for those in non-energy-intensive sectors. This impact is non-negligible. Taken literally, the result means that, by halving their emissions, firms can improve the credit rating implied by their spread by up to one notch, on average.[10]

More importantly, we find that, for the preference channel, carbon risk loads differently across maturities. The interaction between bond maturity and firm-level emissions is relevant at high levels of statistical significance. We dub this finding the *term structure of credit risk-adjusted carbon premia*. The term structure is hump-shaped. Carbon premia increase with maturity up to the belly of the curve (15- to 20-year maturity) and decline thereafter. We offer two conjectures on the curve's shape. The first is the long-term nature of environmental risks, which, despite requiring critical action today, will become inevitable only in the future. The second is the preferred habitat of investors who operate in this market. For example, institutional investors with a sustainable investment mandate (eg pension funds) may use longer-term bonds to match their liabilities but may not go for the ultra-long segments due to liquidity and interest rate risk considerations. As a consequence, the term structure of total carbon premia is also hump-shaped, because the risk channel introduces (roughly) parallel shifts in the term structure of credit risk-adjusted carbon premia.

To-date, not much is known about whether a carbon premium is reflected in asset prices, and our paper contributes to the small but growing literature on the topic. Bolton and Kacperczyk (2021a, 2021b) focus on the carbon risk premium in equity markets. They document the existence of a widespread carbon risk premium in equities: firms with higher carbon emissions offer higher returns across sectors and countries. Briere and Ramelli (2021) show that a green sentiment index – which captures shifts in investor appetite for environmental responsibility – has explanatory power for stock price performance.[11] Ehlers et al (2022) test whether banks demand a premium when lending to firms with higher carbon emission intensity. They find a statistically significant carbon premium in lending rates across industries in the syndicated loan market since the Paris Agreement was struck in 2015. Huynh and Xia (2021) examine whether climate change news risk are priced in corporate bonds, and find that bonds with a higher climate change news "beta" earn lower future returns. Duan et al (forthcoming) explore the pricing of carbon risk in US corporate bond returns. While we also look at the corporate bond market, we differ from Duan et al (forthcoming) in several aspects. First, we focus on the spread instead of the return because bonds are quoted and traded on this variable. The spread-level angle allows

[10] In our sample, the mean spread difference between A– and BBB+ is 10 basis points.

[11] The authors, in fact, predict that green sentiment may anticipate a stock outperformance of more environmentally responsible firms.

us to explore carbon risk pricing within and outside default risk. Second, our findings are also different. While Duan et al (forthcoming) conclude that bonds from firms with more carbon emissions offer significantly lower returns, we show evidence of a positive carbon risk premium consistent with what Bolton and Kacperczyk (2021a, 2021b) and Ehlers et al (2022) find in equities and syndicated loans, respectively. At the same time, our observation that both risk and preference channels contribute to a carbon premium is consistent with Briere and Ramelli's (2021) finding that higher investor demand for environmentally responsible stocks is explained by both fundamental and non-fundamental (ie sentiment) motives. Carbone et al (2021) look into how carbon emissions and mitigating measures, such as climate disclosure and emission reduction targets, influence firm credit risk as measured by credit ratings and the distance to default. Consistent with our findings on the risk channel, they also document that high emissions tend to be associated with higher credit risk.

Our paper also contributes to the line of literature investigating whether the environmental and social commitments of firms, more generally, affect their cost of debt. Goss and Roberts (2011) investigate the impact of corporate social responsibility (CSR) performance on the cost of private bank loans and find that banks charge more for loans to firms with social responsibility concerns. Chava (2014) finds a similar relationship between the cost of bank loans and firms' environmental performance. On public debt markets, Ge and Liu (2015) find that firms' better CSR performance is associated with lower spreads after controlling for credit ratings. Polbennikov et al (2016) study the historical relationship between environmental, social and governance (ESG) ratings and corporate bond spreads, finding that bonds with higher ESG ratings have slightly lower spreads, all else equal. More recently, using data from Sustainalytics, Seltzer et al (2022) find that firms' with lower environmental scores tend to have higher yield spreads – carbon emissions being one of the components.

Finally, our paper also adds to the literature on the determinants of corporate spreads. Since at least Collin-Dufresne et al (2001), it has been recognised that spreads on corporate bonds tend to be several times wider than would be implied by expected default losses alone. The phenomenon is widely known as the credit spread puzzle (Amato and Remolona (2003)). To investigate the puzzle, two types of model have been used to estimate corporate spreads: structural and empirical. Our work falls in the latter camp. Empirically, determinants other than default risk such as taxes (Elton et al (2021)), firm-level equity return volatility (Campbell and Taksler (2002)) and liquidity (Chen et al (2007)) have been found useful in explaining US corporate spreads. The above results have also been validated for other markets such as US mortgage securitisations (Fender and Scheicher (2009)). For the euro area, Boss and Scheicher (2002) show that, among other variables, liquidity and equity return volatility are useful in explaining changes in corporate bond spreads. For China, Chen and Jiang (2019) conclude that liquidity risk significantly affects corporate bond pricing, though its contribution is much smaller than its US counterparts.

The debate on the puzzle is very much alive to this day, with papers arguing in favour of or against its existence (see, for example, the contrasting views of Feldhütter and Schaefer (2018) and Bai et al (2020)). We contribute to the ongoing discussion by adding the carbon risk angle. From a specification perspective, our model most closely resembles that of Gilchrist and Zakrajšek (2012), where credit spreads are written as a function of a proxy of default risk and other variables. However, as established above, our primary purpose is not to predict macroeconomic conditions.

The remainder of the paper is structured as follows. Section 2 describes the carbon emissions, and firm-level and bond-level data sets required for our analysis. Section 3 begins the empirical anaysis with the preference channel Section 4 presents the discussion on the risk channel. Section 5 puts together both channels, showcasing the total effect of carbon risk on corporate bonds. Section 6 concludes.

2. Data

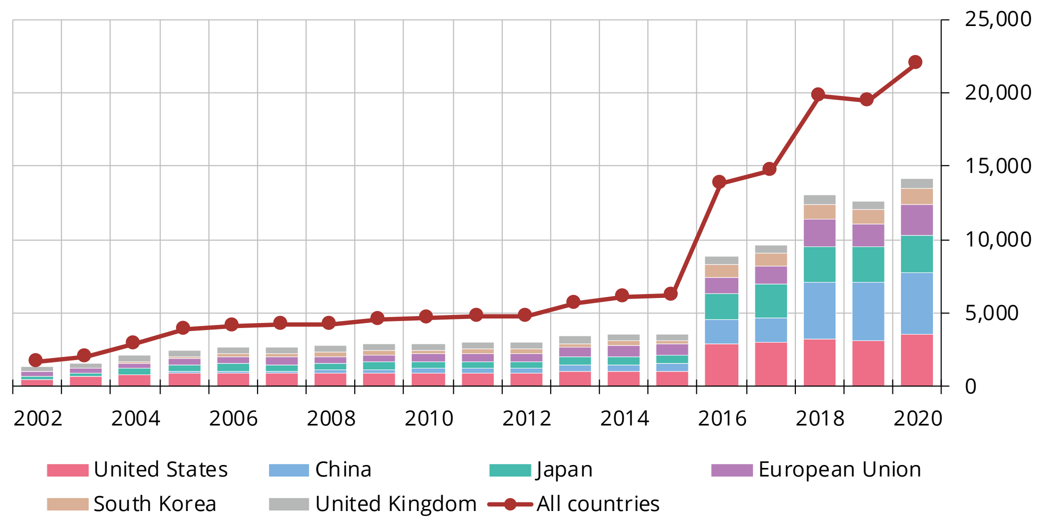
Our analysis seeks to explain corporate bond spreads as a function of a firm's carbon footprint, while controlling for bond and firm level characteristics. For this purpose, we collect three types of data: carbon emissions data, firm-level financial data and security-level data, which are matched and merged to produce the database for the analysis. In this section, we describe each of the three types, while pointing to their particular sources.

2.1 Carbon emissions data

Following others in the literature (Bolton and Kacperczyk (2021a, 2021b) and Ehlers et al (2022), for example), we obtain data on carbon emissions from S&P Global Trucost ("Trucost"). The database provides firms' annual carbon emissions for each fiscal year since 2002. Graph 2.1 summarises firm coverage in the data set. The number of firms in the database has expanded from fewer than 2,000 in 2002 to over 20,000 in 2020 (Graph 2.1, red line). Coverage has also broadened in terms of geographic locations, going from firms mainly in advanced economies to firms in both advanced and emerging market economies. In our analysis, we focus on the United States, given that its companies are the lion's share of the database in market value terms. Altogether, these firms account for 40% of total market value of all firms in the database in 2020. These same firms are also approximately 90% of the total US public market capitalisation.

Firm coverage of the S&P Global Trucost emissions database — Graph 2.1

Number of firms



Source: S&P Global Trucost.

Although this database goes back almost two decades, firm composition changed dramatically in fiscal year 2016 due to additions. The red line in Graph 2.1 shows a more-than-twofold jump from 2015 to 2016. Since we do not wish for this change in sample to bias our results, we choose to start our analysis in 2016, where the number of companies is much richer. This also helps keep our carbon emission time series stable over time. Other research has shown that such sample changes can lead to very different conclusions.[12]

We now define the three types of emission that follow the Greenhouse Gas Protocol (GGP) and that are used in our analysis:[13]

- Scope 1 or direct GHG emissions occur from sources that are owned or controlled by the company. For example, emissions from combustion in owned (or controlled) vehicles and emissions from chemical production in owned (or controlled) process equipment. Scope 1 emissions are part of the disclosure requirements in accordance with the GHG Protocol Corporate Standard.
- Scope 2 or indirect GHG emissions. This accounts for emissions coming from the purchased electricity, steam and heating/cooling consumed by the company. According to the GGP, for many firms, purchased electricity represents one of the largest sources of GHG emissions and the most significant opportunity to reduce them. They are also part of disclosure requirements.
- Scope 3 emissions or other indirect GHG emissions. They are a consequence of the activities of a company, but occur from sources not owned by the company (eg extraction of purchased materials and transportation of fuels). The GGP establishes that disclosure of scope 3 is optional, but provides an opportunity to be innovative in GHG management. Given how difficult they can be to measure, the GGP recommends focusing on one or two major GHG-generating activities, instead of performing a life cycle analysis of all products.[14]

To vary our language, we sometimes use the terms “direct”, “indirect” and “value-chain” emissions to refer to scopes 1, 2 and 3, respectively. In practice, scope 1 and 2 emissions are widely reported across different data providers, including Trucost. Across providers, these two scopes are highly correlated (by >90%; see Busch et al (2022)), which is a sign of consistency. However, this is not the case for scope 3 emissions, given the optional nature of their reporting. As a consequence, one needs to estimate them and methodologies vary across suppliers (eg Trucost uses an input-output method). Given that the data quality of scope 3 emissions is questionable,

12 In their work, Bolton and Kacperczyk (2021a) explain that the important shift on average carbon emissions from 2015 to 2016 is due to the inclusion of new firms. When analysing the effects before and after the Paris Agreement, they also find that excluding these new firms, the carbon premium they find with the full sample becomes statistically insignificant.

13 The Greenhouse Gas Protocol establishes comprehensive global standardised frameworks to measure and manage greenhouse gas emissions from private and public sector operations, value chains and mitigation actions.

14 See GGP (2020), Chapter 4 “Setting Operational Boundaries”. Scope 3 emissions include both upstream and downstream emissions. In our analysis, we focus on upstream emissions, as they are relatively easier to estimate and therefore have longer time series available.

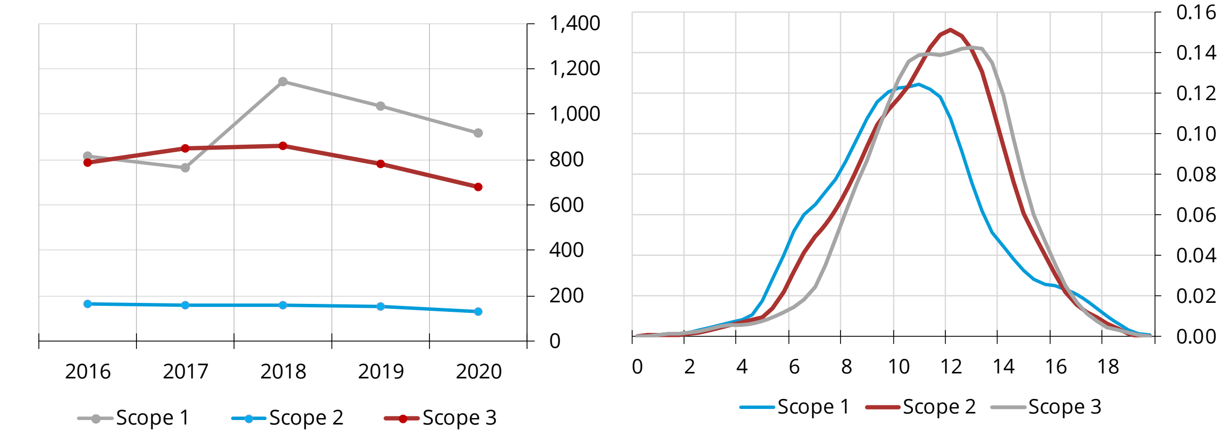
anecdotal evidence suggests only a very limited use of this measure in investment decision-making.

Carbon emissions vary over time and across firms

Graph 2.2

Average carbon emissions in the US by fiscal year
Thousands of tonnes of CO_2

Distribution of US log-carbon emissions in 2020[1]
Probability density



[1] The distribution of carbon emissions is highly skewed. Therefore, a natural logarithm transformation is applied before building the kernels.

Sources: S&P Global Trucost; authors' calculations.

The left-hand panel of Graph 2.2 plots average carbon emissions by scope for all US firms in the Trucost database. Analysing their magnitudes, we see that scope 1 emissions appear to be the highest at about 920 thousand tonnes of CO_2 in 2020; this is followed by scope 3 (~680 ktCO_2e), and scope 2 (~130 ktCO_2e). We also observe that average emissions declined by some 20% from 2018 to 2020, probably reflecting corporations' efforts to reduce their share of GHG. However, focusing on the mean masks an important fact: emissions have wide and asymmetrical distributions. The right-hand panel of Graph 2.2, illustrates this point. Once we take the natural logarithm of firm-level emissions (which are skewed), we see wide and rich probability density functions. Surprisingly, all appear rather continuous and any skew left is not overly pronounced.[15] The panel also depicts the higher mean of scope 3 (yellow line) emissions, although they are defined in a tighter range, than that of scope 1 emissions (red line), for instance.

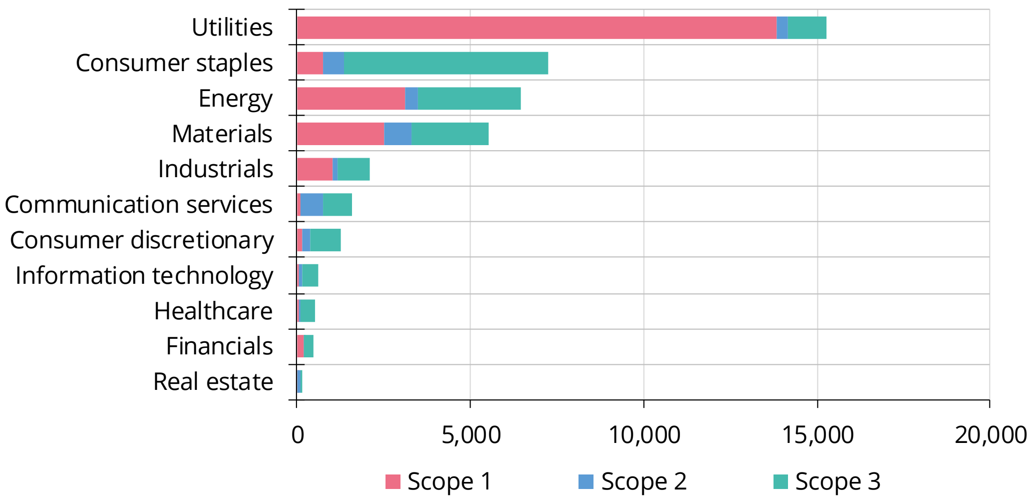
Across sectors too, carbon emissions vary substantially. Graph 2.3 shows average emissions in the United States by GICS sector in 2020.[16] For direct emissions (red bars), sectors traditionally perceived as brown such as utilities, energy and materials stand out as the top three. Yet, when indirect (blue bars) and value-chain (green bars) emissions are taken into account, consumer staples becomes one of the top carbon-intensive sectors. Based on this, and for the purpose of our analysis, we classify utilities, energy and materials as "energy-intensive" sectors when we talk about scope

[15] The probability densities of carbon emissions in tonnes of CO_2 are highly skewed, and therefore require an adjustment before being used in a regression model. The brief exercise illustrates the case of applying a log transformation.

[16] GICS stands for the Global Industry Classification Standard – a method for assigning companies to a specific economic sector and industry group that best defines its business operations. It consists of 11 sectors.

1 emissions and the sum of scopes 1 and 2. And, we replace materials with consumer staples as part of the energy-intensive category when we include scope 3 in our firm-level total emissions.[17] It is also interesting to analyse how total emissions are distributed *within* sectors. When computing the share each scope represents as a percentage of total, we find that scope 3 makes up for a great share in many sectors. Focusing only on emissions with disclosure requirements, however, the results are mixed. Depending on the sector, total emissions at the firm level may be driven by their direct (eg in the energy sector) or indirect emissions (eg in real estate), which is certainly dependent on the nature of the business. We keep this in mind when we consider the existence of the carbon premium.

Average carbon emissions by sector in 2020 Graph 2.3
In thousands of tonnes CO_2e



Sources: S&P Global Trucost; authors' calculations.

2.2 Firm-level data

In addition to carbon emissions data, we make use of two other types of firm-level data. The first is a measure of credit risk: the probability of default, our key variable to examine the risk channel. The second includes other firm characteristics affecting their credit risk that are well established in the literature. We need to control for these variables when testing whether a firm's carbon emissions play a role in its credit risk. We gather these variables for companies with carbon emissions data.

2.2.1 Probability of default

To measure a firm's credit risk, we rely largely on default probability data from Bloomberg. Bloomberg provides forward-looking real-world probabilities of default for publicly traded firms. As these are updated daily, the estimates are up to date with current market conditions. A logistic model is used to estimate the probability of

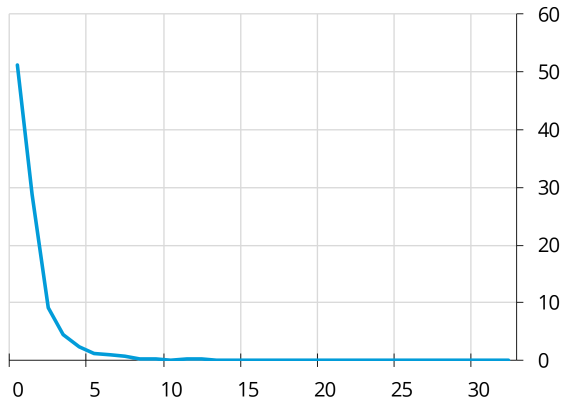
[17] A dummy variable distinguishing between "energy-intensive" and "non-energy-intensive" firms is used in our analysis.

default based on factors that best capture credit risk.[18] For firms in our sample, the annualised five-year-ahead probability of default ranges from near zero to around 32%, with mass concentration being in the range of 0–4% (Graph 2.4, left-hand panel).

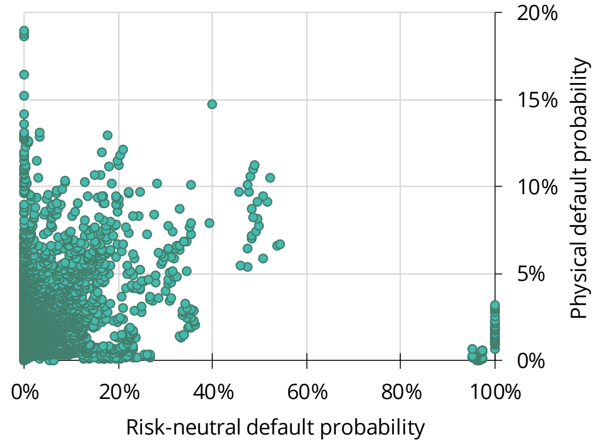
For our robustness checks (in section 6), we also compute our own probabilities of default, which we derive from the Merton (1974) model. Details of the computations can be found in Appendix 1. The two measures of default probability are plotted in Graph 2.4 (right-hand panel). The default probabilities are different because our estimates consist of risk-neutral default probabilities, while Bloomberg's estimates are physical or "real-world" probabilities. The difference between the two reflects risk premia.[19] While the latter is more relevant for corporate bond pricing, the former has the advantage of being a cleaner measure of credit risk.

Five-year-ahead probability of default — Graph 2.4

Distribution of five-year-ahead probability of default
Per cent



Risk-neutral versus physical default probability
Per cent



Sources: Bloomberg; S&P Capital IQ; authors' calculations.

An alternative measure of credit risk is the credit rating provided by agencies such as S&P Global, Moody's and Fitch.[20] We prefer the probability of default to credit ratings because the latter are coarsely grained, with the rating designed to remain broadly static over time. Probabilities of default fluctuate over the short term, reflecting information at a higher frequency. We further analyse firm-level default probability when we look at the security-level data in section 2.3.

18 The risk factors include relevant accounting ratios such as return-on-assets, non-performing loans for financial firms, and interest coverage for non-financial firms as well as the distance to default. The distance to default is calculated using the Black-Cox model, which writes it as a function of total debt (proxied by the sum of short-term debt and 50% of long-term debt), value of assets and the implied volatility of assets.

19 Under the Merton model, default takes place in a contingent claims framework, which means all probabilities are risk-neutral. Bloomberg's model (see Bondioli et al (2021)) adds an extra step, where risk-neutral probabilities are mapped into physical ("real-world") probabilities via a logistic model. Nonetheless, under the structural modelling framework, it is easy to show that the spread can be written as a function of the risk-neutral default probability, giving us permission to use this quantity as a regressor.

20 For example, Carbone et al (2021) examine how a firm's carbon footprint affects its credit rating.

2.2.2 Financial variables on firms

To control for variables that affect a company's credit risk, we obtain relevant firm-level financial variables from S&P Capital IQ.[21] The control variables we consider include: size of assets, the long-term debt-to-asset ratio, the earnings-to-asset ratio, the capital-to-asset ratio and return-on-assets. We choose these variables to be consistent with the literature on corporate credit risk (eg Carbone et al (2021)). Intuitively, larger, more profitable and better-capitalised firms are typically less likely to default. By contrast, more indebted firms are more prone to default. Summary statistics for these variables are provided in Table 2.1. Note that these variables tend to be defined in wide ranges relative to their respective standard deviations. Therefore, we winsorise them at the 2.5 percentile before conducting our regression analysis.

We also obtain daily equity price data from Bloomberg. We use these data to compute the volatility of equity returns, a firm-level characteristic that we will control for in the preference channel analysis.

Summary statistics for financial variables on firms

Table 2.1

	Mean	Std. dev.	Min	Median	Max
ln(asset)	7.61	1.96	–4.26	7.62	15.14
Long term debt/asset	0.29	0.52	0	0.24	30.26
Earnings/asset	–0.86	22.72	–1,312.97	0.05	7.65
Capital/asset	0.09	1.8	–119.33	0.13	1
Return on asset (%)	0.39	22.43	–3,487.93	2.45	677.93

Summary statistics are computed across 2,813 firms between January 2017 and December 2021.

Sources: S&P Capital IQ; authors' calculations.

2.3 Security-level data

To build our bond data panel, we start from the universe of all firms with emissions data from 2016 to 2020 in the S&P Global Trucost database. We conduct our data-gathering process in two steps: first, we make a list with the corporate bonds of firms with carbon emissions data; then, we fetch the relevant data fields for this list of securities.

To find the securities issued by each firm in our carbon data set, we use Refinitiv. Its search function allows us to use company-level ISINs (found in the Trucost database) to generate individual CUSIP lists for each of the firms. We include bonds issued by both the parent company and its subsidiaries. To query these lists, we apply a series of filters, in the spirit of Bai et al (2019). We exclude the following:

- Bonds that are not listed or traded in the US public market, which includes bonds issues through private placement and bonds issued under the 144A rule;[22]
- Structured notes, asset-backed, equity- or index-linked securities;

[21] Macroeconomic variables affect a firm's credit risk as well. We include time-fixed effects to this end.

[22] Unlike Bai et al (2019), we preserve bonds traded in a currency other than the US dollar.

- Convertible bonds;
- Floating coupon rate securities;
- Securities with a maturity lower than one year;
- Unrated securities; and
- Bonds trading under $5 or above $1,000.

For the remaining securities, we download a set of static and historical data fields by combining two sources: Bloomberg and Refinitiv. And, as we will be using emissions data lagged by a year (which start in 2016) in our model, these historical fields are obtained from January 2017 to December 2021.

From Bloomberg, where trade data are more widely available, we fetch monthly option-adjusted spreads and daily close prices. The latter will be used to compute our time-varying measure of bond liquidity (see Annex 1 for details) – an important determinant of corporate bond spreads.[23] From Refinitiv, we download monthly data for the amount outstanding, maturity, age, duration and credit rating of each bond. From this same source, we draw static data on coupon and whether the security is callable or not, which we store as a dummy variable (equal to one if the bond is callable and zero otherwise).

We then perform some transformations. First, to ensure that our results are not driven by a small number of extreme observations, we winsorise option-adjusted spread data at the 2.5% fraction. Next, we assign a numerical score to credit ratings. Our score goes from 0 to 20, where the AAA rating on the S&P scale is assigned the highest value (20) and a rating of C is the lowest (zero). This way, we ensure that our variable has the following interpretation: higher credit quality entails a higher credit score.

Summary statistics for corporate bonds

Table 2.2

Variable	Mean	Std. dev.	Min	Median	Max
Option-adjusted spread[1] (bps)	141	114	17	111	619
Maturity (years)	10.4	10	1	6.8	101.4
Duration (years)	7.1	5.1	0	5.7	35
Age (years)	5.3	5.6	0	3.7	32.6
Coupon (pp)	4.2	1.8	0	4	12.3
Amount outstanding (USD mn)	1,050	5,000	0	535	250,000
Credit rating[2]	13	3	0	13	20
Liquidity[3] (bps)	27	27	0	18	122
Callable (binary)	62%	-	0	-	1

Sample period: January 2017 to December 2021; observations = 263,797; number of bonds: 7,599. 1 Option-adjusted spreads are winsorised at the 2.5% fraction. 2 Credit ratings are converted to a numerical S&P scale equivalent from 1 to 20, where 20 = AAA+ and 1 = CCC. 3 Absolute roll measure in basis points (see Annex 1).

Sources: Bloomberg; Refinitiv; authors' calculations.

Our final sample comprises 7,599 securities issued by 779 unique firms. Table 2.2 shows the set of cross-sectional summary statistics we use to characterise our bond

[23] In fact, our full list of determinants is based on the corporate bond literature. This is covered in depth in section 3.

sample. We draw two important observations from the table. First, option adjusted spreads are straddled in a wide range from 17 to 619 basis points, with the average being about 140 basis points. Second, the average credit rating is 13 (equivalent to BBB+), two notches above the investment grade cutoff. With regards to other features, we see that, by construction, bond maturity is above 12 months and is 10 years on average. In line with its definition, modified duration stands somewhat lower, at 7.1 years. We also see the average coupon sitting at about 4 percentage points per annum, and the outstanding issue size slightly above $1 billion on average. It is also important to note that over 60% of our sample is constituted by callable bonds, which asserts our choice of *option-adjusted* spreads to account for this optionality.

We can also look at our bond data under different sample splits. Table 2.3 shows the average corporate spread across the different GICS sectors, which we know are important when looking at firm-level GHG emissions. Across industries, the mean spread sits between 100 and 250 basis points, with the lowest in information technology (103 basis points) and the highest in the energy sector (242 basis points). In theory, this heterogeneity reflects the differences in credit risk across sectors, which underscores the need to control for firm-level default probability. Indeed, this ordinal relationship is preserved when we look at firm-level default probability (fourth column in Table 2.3). Furthermore, spread dispersion, captured by the standard deviation, appears to differ from one sector to another, highlighting the nuances *within* sectors. When looking at bond maturity, we find that the sector average is close to the full-sample number of 10 years in most cases. An exception appears to be the real estate sector, with an average maturity of seven years. Finally, when we look at the number of observations to be included in the model, we see a slight dominance of the financial and industrial sectors. The industry categories with fewer observations are materials, utilities and real estate.

Corporate bond summary statistics by sector

Table 2.3

Sector	Mean spread[1]	Std. dev. of spread	Firm's probability of default[2]	Bond maturity	Number of observations
Communication services	155	102	0.50%	13	21,723
Consumer discretionary	182	140	0.70%	9	25,819
Consumer staples	108	91	0.30%	10	27,914
Energy	242	176	1.20%	10	17,782
Financials	126	92	0.40%	9	40,693
Healthcare	108	81	0.30%	11	31,322
Industrials	138	108	0.60%	12	34,735
Information technology	103	82	0.30%	10	28,551
Materials	174	120	0.50%	10	12,402
Real estate	166	116	0.40%	7	10,910
Utilities	126	73	0.50%	10	11,831
All sectors	141	114	0.50%	10	263,682

Sample period: January 2017 to December 2021; observations = 263,797; number of bonds: 7,599. [1] Option-adjusted spreads are winsorised at the 2.5% fraction. [2] Annualised, five-year-ahead probability of default from Bloomberg.

Sources: Bloomberg; Refinitiv; authors' calculations.

Finally, exploring the data by rating (table omitted for brevity), we find a strictly monotonic relationship between credit rating and corporate spreads: a higher credit quality translates into a lower mean spread. For instance, the mean spread on AAA–rated securities is 54 basis points, with the spread tripling at the investment grade cutoff of BBB–. The result validates the magnitude of the spread as the market's proxy for the perceived level of default risk. Naturally, credit ratings are only a qualtitative (or "soft") indicator, and a model is required to quantify default risk. Indeed, we again present the average value of the firm-level probability of default (fourth column of Table 2.4). The relationship of this variable with the mean spread is almost strictly monotonic, except for the break in the BB– notch, where the average default probability is 0.94%. Nonetheless, these differences are explained by the reduction in sample size as we approach the lowest credit ratings. Indeed, these are the least represented in our sample: altogether, credit ratings of BB+ or below represent about 14% of all our observations, which means our bond sample best (yet not exclusively) represents the investment grade corporate debt spectrum.

Putting the above findings altogether, we conclude that it is important to control for sector-, firm- and issue-level features when performing our regression analysis, which is covered in Sections 3 and 4. Section 3 is dedicated to the analysis at the corporate spread level – the preference channel – and Section 4 details the default risk probability model – the risk channel – which complements our headline results.

3. The preference channel

In this section, we establish a model which, controlling for default risk, is able to explain over 80% of the variation of credit spreads in the United States. We then extend it to include carbon transition risk. Our hypothesis is that, default risk considerations aside, investors trade on information about firms' carbon emissions – a gauge of their environmental footprint and, thereby, of their exposure to carbon transition risk. The consideration of carbon transition risk beyond credit risk captures a genuine preference on the part of investors for environmentally friendly firms, whether motivated by the firms' reputations or investor mandates. Such preferences may be reflected in practice, for example, in the screening of issuers when building investment portfolios. Whether default risk *itself* is affected by the carbon footprint of a firm is addressed in Section 4.

3.1 Underlying theory

According to theory, the price of a corporate bond must reflect the spot rate of a default-free bond (ie government bond) plus a risk premium paid for facing default risk and any options embedded in the issue. This risk premium is known as the corporate spread, and is computed as the difference between the risk-free rate and the yield to maturity on the corporate bond. We denote the spread of bond j issued by firm i at time t as $s_{i,j,t}$, and the firm's probability of default with $P_{i,t}$.

Our empirical methodology is based on the premise that the spread on a bond is directly proportional to the issuing firm's probability of default ($\beta > 0$ a constant):

$$s_{i,j,t} = \beta P_{i,t} \tag{1}$$

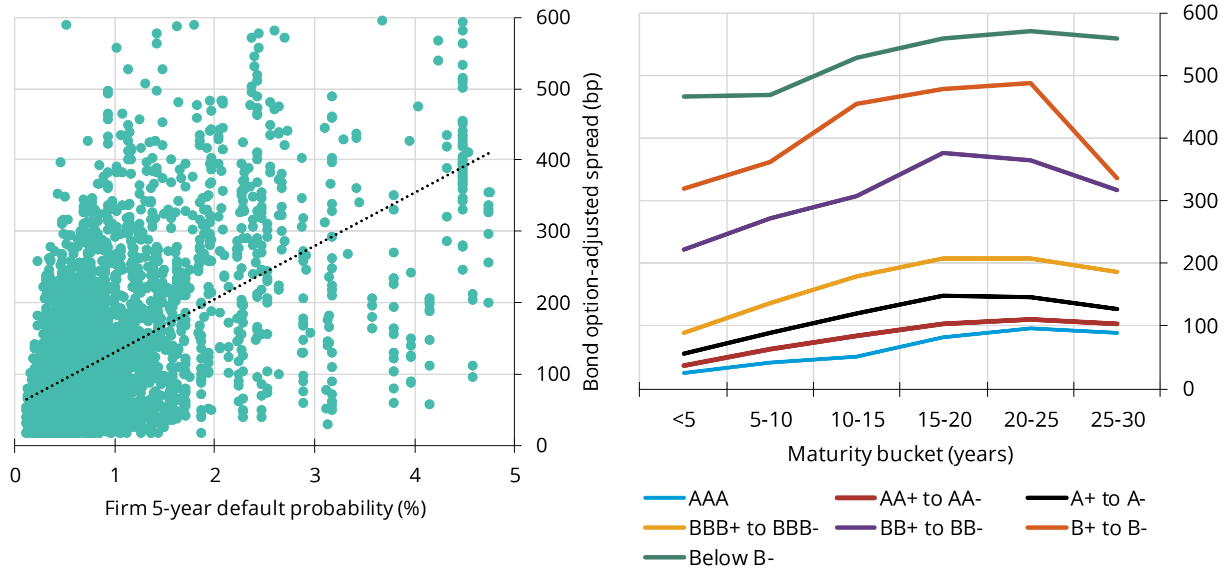
The higher $s_{i,j,t}$, the greater the expectation that the firm will fail on its payments. We can validate this by taking a look at the relationship between these variables in practice. The left-hand panel of Graph 3.1 shows, for January 2021, firm-level default probabilities in the United States (proxied here by the five-year probability described in Section 2) plotted against the option-adjusted spreads on a firm's bonds.[24]

The relationship appears positive: the greater a firm's default probability, the wider the spread on its securities. There are some nuances, however. For example, looking at firm-specific information ignores important bond features, such as the maturity of the issue and its particular credit rating. It is well known that credit spreads increase with respect to credit ratings, which are qualitative assessments about a bond's serviceability. The right-hand panel of Graph 3.1 illustrates how the average spread increases as a function of these two variables: maturity (going from left to right) and rating (going from bottom to top). The function linking maturity with spread is known as the *term structure* of credit spreads.

US credit spreads and their relationships to firm- and bond- specific variables — Graph 3.1

Relationship between a firm's default probability and its bond credit spread (as of Jan 2021)[1]
Percentage; basis points

Term structure of credit spreads, by bond rating and maturity buckets (as of Jan 2021)[2]
Basis points



[1] Dashed line denotes a simple regression line $ax + b$. [2] Computed as the average credit spread across all bonds within each combination of maturity and credit rating buckets.

Sources: Bloomberg; Refinitiv; authors' calculations.

The fact that the default probability alone cannot explain all of the variation in spreads is an established feature from the literature, known as the credit spread puzzle.[25] More specifically, the puzzle derives from the fact that neither levels nor changes in the yield spread of corporate bonds over Treasury bonds can be fully explained by credit risk determinants proposed by structural models (eg a firm's

[24] Given that the credit risk premium can reflect any options embedded in the issue, we use the option-adjusted spread in our analysis. This is important because, as explained in Section 2, above 60% of our sample comprises callable bonds.

[25] This is an empirical finding held since at least Collin-Dufresne et al (2001). For a much more recent discussion on the credit spread puzzle, see Bai et al (2020).

financial health, macroeconomic conditions). As a consequence, the model for spreads is much more complex than in equation (1), looking more like equation (2) below:[26]

$$s_{i,j,t} = \alpha + \beta_P P_{i,t} + \beta_Z Z_{i,j,t} + FE + \epsilon_{i,j,t} \tag{2}$$

where α is the constant of the regression model; β_P the coefficient on the firm-specific probability of default $P_{i,t}$; $Z_{i,j,t}$ a vector of bond- and/or firm-specific controls (with β_Z their respective coefficients); FE a set of fixed effects (typically related to the macroeconomic environment and firm-specific characteristics) and $\epsilon_{i,j,t}$ a zero-mean disturbance or "pricing error". A body of research has been dedicated to defining and expanding the set of controls $Z_{i,t}$, adding missing pieces to the puzzle. We review these next, in chronological order:

- Elton et al (2001) propose the tax premium as a determinant of yield spreads. According to their work, this premium arises because of the higher taxation of corporate bonds compared with sovereign bonds. This effect was later debated by Longstaff (2005), who finds weak support for the hypothesis that the non-default component of spreads is due to taxes.
- Campbell and Taksler (2002) find that idiosyncratic firm-level equity volatility is directly related to the cost of borrowing for corporate issuers. According to them, volatility should drive up the yields of bonds, given that volatility of firm value hurts bondholders. Their study suggests that volatility can indeed explain cross-sectional variation in yields as much as credit ratings. This result has been carried forward to more recent models, such as Rossi (2014), who works with realised volatility.
- Chen et al (2007) argue in favour of the existence of a liquidity premium. They show that several measures of corporate bond liquidity such as the bid-ask spread or the percentage of zero returns are key determinants of bond yield spreads. A wide array of studies has included liquidity as a standard variable in corporate bond modelling; a more recent example is He and Milbradt (2014).

Our research adds to this list by considering a measure capturing a firm's carbon emissions. Our hypothesis is that carbon transition risk is priced in the cross section of corporate spreads, thereby granting investors a carbon premium. As in the case of the liquidity premium, the risk arises from holding a bond that is *less preferred* by investors, given a firm's heavier environmental footprint relative to others. As argued in Ehlers et al (2020), environmental factors – and, most importantly, carbon emissions – are a material financial risk for creditors, which invites exploration.[27]

3.2 Baseline estimation

Our exercise consists of estimating equation (2) via panel regression at the bond level. We estimate $s_{i,j,t}$, the spread of the bond j issued by firm i at month t, as a function of $P_{i,t}$, the firm-level estimate of five-year-ahead default probability at month t, plus the following other variables:

[26] See Gilchrist and Zakrajsek (2012) for example, where option-adjusted spreads are a function of distance to default (a variable representing default), plus bond- and firm-specific variables.

[27] This section covers the preference angle, however. The credit risk angle is handled separately in Section 4.

- $Z_{i,j,t}$ is a vector of six bond-specific variables and one firm-level variable. On the bond side, we use duration, age, coupon and (the natural logarithm of) the amount outstanding as controls. Because liquidity is a well known determinant, we also compute the Roll measure of liquidity, which serves as our proxy for bid-ask spreads.[28] Furthermore, we include a dummy variable which is equal to 1 when the issue is callable and zero otherwise.[29] For the firm side, we compute company-level equity return volatility, as in Campbell and Taksler (2002).[30]
- FE, is a battery of fixed effects. Time-, firm- and credit rating- fixed effects are included.[31] As in the literature, time fixed effects serve as controls for macroeconomic effects (eg state of the yield curve, business cycle). The latter, as in Gilchrist and Zakrajšek (2012), are meant to capture the "soft information" regarding the firm's financial health, which is complementary to our default probability measure.
- Finally, to test our key hypothesis: that carbon transition risk is priced in corporate bond spreads, we need a metric of carbon emissions.

Which measure(s) should be included in the model? We take a practitioner's view and assume that, when making their decisions, investors care about whether the company pollutes the environment or not, regardless of profit.[32] We also suppose that, when they look at greenhouse gases, they think of them on a cumulative basis.[33] In other words: investors do not consider indirect emissions (scope 2) independently of direct emissions (scope 1, the baseline). Instead, they care about the total level of pollution: the sum of both scopes altogether. It is also important to note that the reliability of value chain emissions (scope 3) is at this stage questionable, given their lack of wide availability and inconsistency across data providers.[34] We keep this in mind when analysing our results.

[28] This measure of illiquidity was originally proposed by Roll (1984). More recently, Christopoulos (2021) introduced a version which addresses the presence of positive autovariance in the original formula. See Appendix 2 for computational details.

[29] Duffee (1998) finds that the relation between credit spreads and Treasury rates is stronger for callable bonds than for non-callable bonds. It is thereby important to make a distinction between these two types of instruments in any spread model.

[30] As in Campbell and Taksler (2002), equity return volatility is calculated as the 180-day trailing standard deviation of the firm's stock return at the end of each month.

[31] The composite credit rating is the average rating across three providers: S&P, Moody's and Fitch, when available. Furthermore, we assume that, if present, the effects of taxes are absorbed by the fixed effects, given their static nature.

[32] It is debated whether the explanatory variable representing emissions should be a level or a ratio (eg intensities). We pose that investors care whether a firm is pollutes *more* or *less* than others, regardless of their level of profitability. This is because, when the ultimate goal is "net-zero", firms who emit more greenhouse gases into the atmosphere are not less exposed to a carbon tax, technological change or investor dispreference simply because they generate more income.

[33] Nonetheless, we investigate the impact of carbon intensities (defined as the ratio of carbon emissions to revenue) on spreads and report these results in Appendix 3.

[34] Scope 1 and scope 2 emissions have been more systematically reported because of disclosure requirements. However, scope 3 emissions are still estimated by data providers, such as Trucost. Busch et al (2022) find that the complexity of carbon accounting increases from direct emissions to indirect emissions (scope 2 and 3), and the consistency of data between third-party providers decrease: correlations among them drop from >90% to <60% across providers. They suggest that

Therefore, in our regression, we include a firm-level term capturing total GHG emissions in tonnes for each financial year. We work, first, with scope 1 emissions; then, with the sum of scopes 1 and 2; and finally, with the sum of scopes 1, 2 and 3. As others in the literature, we lag these numbers by 12 months, to reflect the availability of this information for the average investor.[35] For easier interpretation of the coefficient, we take their natural logarithm.

Our estimated model is as follows:

$$E(s_{i,j,t}) = \hat{\alpha} + \hat{\beta}_P P_{i,t} + \hat{\beta}_Z Z_{i,j,t} + \hat{\beta}_{P,Carbon} \ln(\text{Emissions}_{i,t-12}) + \widehat{FE} \qquad (3)$$

The terms following $\hat{\beta}_P P_{i,t}$ in the equation represent spread determinants beyond credit risk. This allows us to conjecture that the effect captured by our estimate $\hat{\beta}_{P,Carbon}$ is due to investor preferences, all else equal. We call the effect of this coefficient the preference channel.

We present the results from four regressions in Table 3.1. The first is a specification without carbon emissions and the latter three introduce emission scopes 1 to 3 in a cumulative fashion. We start by focusing on specification (1) to analyse the effect of well known bond- and firm-level controls.

Effects of carbon emissions on US corporate bond spreads

Table 3.1

	(1)	(2)	(3)	(4)
ln(scope 1 emissions)		1.61**		
		[0.74]		
ln(scope 1+2 emissions)			5.22***	
			[1.16]	
ln(scope 1+2+3 emissions)				2.44
				[1.75]
Default probability (%)	32.01***	31.99***	31.95***	32.04***
	[1.26]	[1.26]	[1.26]	[1.26]
Duration	5.22***	5.22***	5.22***	5.22***
	[0.12]	[0.12]	[0.12]	[0.12]
Age	0.56***	0.56***	0.56***	0.56***
	[0.13]	[0.14]	[0.14]	[0.14]
Coupon	10.53***	10.54***	10.53***	10.52***
	[0.47]	[0.47]	[0.47]	[0.47]
ln(amount outstanding)	-2.92***	-2.92***	-2.92***	-2.92***
	[0.32]	[0.32]	[0.32]	[0.32]
Equity return volatility (%)	17.78***	17.74***	17.73***	17.79***
	[0.93]	[0.93]	[0.93]	[0.93]
Liquidity	0.44***	0.43***	0.43***	0.43***
	[0.17]	[0.17]	[0.17]	[0.17]
Callable	-8.00***	-7.97***	-7.95***	-7.98***
	[0.99]	[0.99]	[0.99]	[0.99]
Number of bonds	7599	7599	7599	7599
Observations	263,682	263,682	263,768	263,797
R-squared	0.84	0.84	0.84	0.84

*** p<0.01, ** p<0.05, * p<0.1. Specifications with time-, firm- and credit rating fixed effects. Standard errors clustered at the security level.

Sources: Bloomberg; Refinitiv; authors' calculations.

requesting firms to follow a standardised approach is even more important in the complex scope 3 realm.

35 See eg Ehlers et al (2022) and Duan et al (forthcoming).

The probability of default is, as expected, highly significant in explaining corporate bond spreads. Their magnitude is also powerful: an increase of 1 percentage point in this probability raises spreads by 32 basis points (bp), on average. Investors need to be heavily compensated for facing higher credit risk. Moving on to bond-specific features, we find that duration, age and coupon are all positively related to spreads, in line with theory. One more year of interest rate risk entails a 5 bp higher spread; a bond one year older (less "on-the-run") delivers a yield of half a basis point higher; and, having a coupon of one more 1 pp increases corporate yields by 10 bp due to higher income. In turn, the amount outstanding (a measure of size and therefore supply) has a negative effect; this makes sense, as we would expect bonds with a greater supply to bear lower yields.

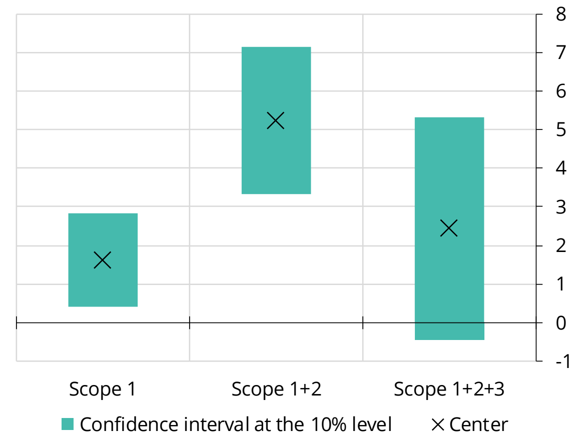
Next, we turn to our computed variables. We start with equity return volatility, a measure of firm value. In line with our prior, it is positive at high levels of statistical significance. An increment of 1 pp in the standard deviation of the company's stock return pushes spreads upwards by approximately 18 bp, in line with the original work cited in Section 3.2. Our second variable is the Roll measure of liquidity which also performs well, bearing a positive sign at the 1% level. The coefficient predicts a rise of 0.43 bp in corporate spreads for every basis point increase in our *synthetic* bid-ask spread. The order of magnitude of our result is strikingly similar to that of the *observed* bid-ask spread in the work of Chen et al (2007). They find a coefficient of about 0.42 bp.

Having validated that all controls and historical determinants of corporate spreads behave as expected, we now focus on our hypothesis regarding carbon transition risk. Specification (2) shows that, at the 5% level, scope 1 emissions predict corporate spreads – evidence of a *credit risk-adjusted* carbon premium. Concretely, a 1% increase in GHG directly emitted entails a 0.02 basis point yield increment. We can also look at a positive message: what happens when firms reduce their emissions? For instance, by halving their direct carbon emissions (ie reducing them by 50%), firms can reduce their funding costs by 1.1 basis points, on average. Despite statistical significance, the economic effect seems low.

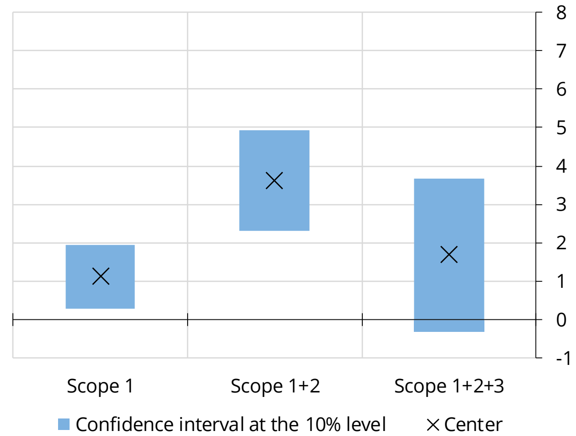
Effect of carbon emissions on corporate spreads

Graph 3.2

Regression coefficient on lagged carbon emissions[1]
Units



Spread impact of a 50% reduction in emissions[1,2]
Decrease in basis points



[1] If the bar touches zero, the null hypothesis that the coefficient is zero cannot be rejected. [2] Computed as the coefficient $\hat{\beta}_{P,Carbon}$ on carbon emissions, multiplied by $ln(0.5)$.

Sources: Bloomberg; Refinitiv; Trucost; authors' calculations.

Turning to specification (3), which puts direct *and* indirect emissions together, we see both the statistical power and the economic significance rise. A joint 50% reduction of scope 1 and 2 emissions predicts a 3.6 basis points decrease in the cost of debt at the 1% confidence level. This result makes our carbon premium findings more meaningful. Finally, we note that specification (4), which covers indirect emissions along the value chain, bears a lower coefficient and strips out any statistical significance. We take this result as confirmation of our word of caution about using scope 3 emissions – which are neither widely available nor consistent across providers – as an explanatory variable.

Non-energy-intensive vs energy-intensive carbon premium

Table 3.2

	(1)	(2)	(3)	(4)	(5)	(6)
	Scope 1 emissions			Scope 1+2 emissions		
ln(emissions)	0.86	3.64*		4.54***	4.40*	
	[0.74]	[2.13]		[1.16]	[2.58]	
Non-energy- intensive x ln(emissions)			0.38			3.88***
			[0.73]			[1.17]
Energy- intensive x ln(emissions)			6.96***			8.94***
			[2.39]			[2.93]
Default probability (%)	29.24***	38.25***	31.89***	29.26***	38.22***	31.87***
	[1.25]	[3.37]	[1.26]	[1.26]	[3.37]	[1.26]
Duration	5.10***	6.53***	5.22***	5.11***	6.53***	5.23***
	[0.12]	[0.28]	[0.12]	[0.12]	[0.28]	[0.12]
Age	0.39***	1.28***	0.56***	0.39***	1.28***	0.56***
	[0.15]	[0.31]	[0.14]	[0.15]	[0.31]	[0.14]
Coupon	10.87***	9.36***	10.52***	10.87***	9.35***	10.52***
	[0.52]	[0.99]	[0.47]	[0.52]	[0.99]	[0.47]
ln(amount outstanding)	-2.26***	-6.63***	-2.91***	-2.27***	-6.64***	-2.92***
	[0.31]	[1.28]	[0.32]	[0.31]	[1.28]	[0.32]
Equity return volatility	15.57***	28.75***	17.68***	15.63***	28.74***	17.70***
	[0.98]	[2.63]	[0.93]	[0.98]	[2.62]	[0.93]
Liquidity	0.40***	0.47***	0.43***	0.40***	0.47***	0.43***
	[0.02]	[0.04]	[0.02]	[0.02]	[0.03]	[0.02]
Callable	-6.82***	-12.88***	-7.97***	-6.80***	-12.88***	-7.94***
	[1.04]	[2.75]	[0.99]	[1.04]	[2.75]	[0.99]
Number of bonds	6330	1269	7599	6330	1269	7599
Observations	221,667	42,015	263,682	221,782	42,015	263,768

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in brackets; clustered at the security level.

Sources: Bloomberg; Refinitiv; Trucost; authors' calculations.

Our results are summarised visually in Graph 3.2. The credit risk-adjusted premium appears statistically significant (left-hand panel, our regression coefficients) when we account for direct and indirect emissions that are required disclosures by the Greenhouse Gas Protocol. And the economic effect (right-hand panel, basis points) appears highest when the sum of scope 1 and 2 emissions are considered.

3.3 Exploring sector effects

In Section 2, we showed an important step difference in the order of magnitude of emissions between companies considered "energy-intensive" and those that are not. So, should bonds from all firms bear the same carbon premium? In this subsection, we seek to answer this question.

To analyse differences in the carbon premium between energy-intensive and non-energy-intensive sectors, we conduct two different exercises:

a. We split the bond sample into two subsamples. One with securities from non-energy-intensive firms and another with bonds from energy-intensive ones. This allows us to vary the coefficients on the control variables depending on the firm's sector.

b. We run a full-sample exercise which interacts our carbon emissions variable with an energy-intensive sector dummy. The value of the dummy is equal to 1 when the company belongs in the category and zero otherwise.

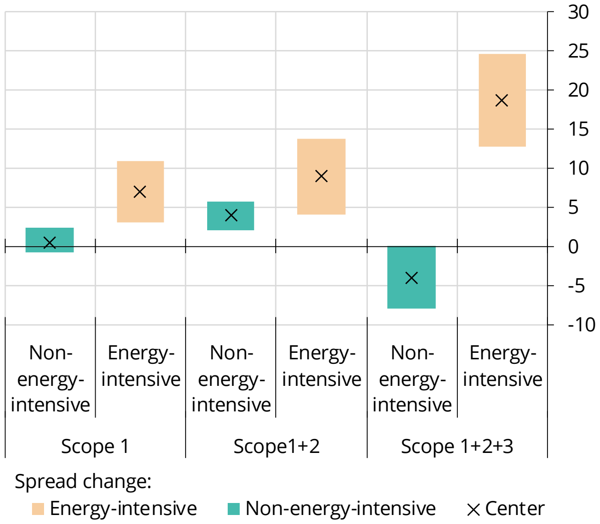
Table 3.2 shows the results for exercises (a) and (b) across the different groups of emissions. We start by describing the subsample results for scope 1. Specifications (1) and (2) bear very different coefficients in front of the log-emissions variable. For non-energy-intensive sectors (model 1), the coefficient is below one and does not appear statistically significant; yet for energy-intensive sectors (model 2), $\hat{\beta}_{P,Carbon}$ grows fourfold and becomes significant at the 10% level. When we apply the dummy variable (model 3), we find a similar result: the carbon premium is much higher for firms in energy-intensive sectors.

This result changes somewhat when looking at scope 1 and 2 emissions jointly. Models 4 to 6 show how including indirect emissions in the computation gives its coefficient statistical significance across all kinds of firms, regardless of their energy consumption. Though the subsample results show similar premia for both cases (models 4 and 5), our dummy specification in particular (model 6) offers a carbon premium at least twice as big for bonds from energy-intensive companies. Finally, by looking at models 7 to 9, we once again see that using scope 3 emissions in our modelling introduces awkward dynamics (eg a negative sign on $\hat{\beta}_{P,Carbon}$ for non-energy-intensive bonds (model 7)).

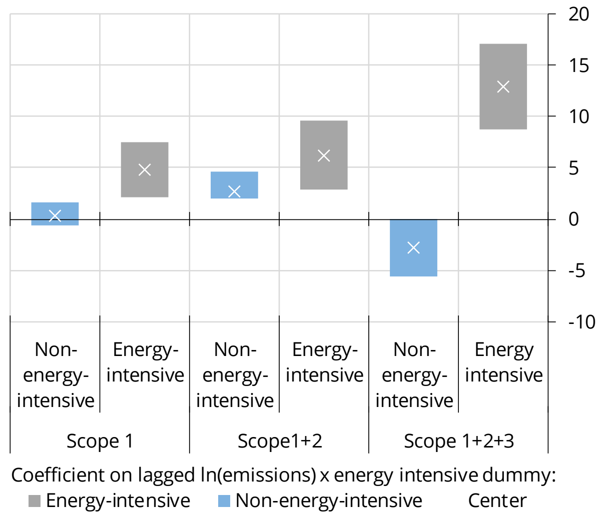
The risk-adjusted carbon premium by sector

Graph 3.3

Effect of carbon emissions on corporate spreads[1]
Coefficient



Spread decrease implied by −50% in emissions
Basis points



[1] Specifications where lagged log-carbon emissions are interacted with an indicator variable equal to 1 when the firm is from an energy-intensive sector. Bars denote a confidence interval at the 10% level. If the bar touches zero, the null hypothesis that the coefficient is zero cannot be rejected. [2] Computed as the coefficient $\hat{\beta}_{P,Carbon}$ on carbon emissions, multiplied by $ln(0.5)$.

Sources: Bloomberg; Refinitiv; Trucost; authors' calculations.

A graphical summary of specifications (b) with the interaction is offered in Graph 3.3. We highlight a few differences with our previous exercise: first, there are important changes in the carbon premium between companies considered energy-intensive and those that are not. The coefficients (left-hand panel) are at least twice the size for more polluting ("browner") firms.

As a consequence, the spread impact is much increased. A reduction of 50% in 1+2 emissions could help energy-intensive firms' bonds trade 9 bp cheaper. This is equivalent to a rating upgrade of 0.8 notches.[36] Given the relevance of both scope 1 and scope 2 emissions in compulsory reporting, we consider model (6) the key finding of this section. Zooming in on firms in energy-intensive sectors reveals a more meaningful preference channel than found in the overall sample. The impact within the set of energy-intensive firms is statistically significant and of non-negligible economic importance.

[36] This result is the average of individual notch changes across all bonds. To estimate each individual notch change, the 9 bp impact from a 50% change in emissions is divided by the differential between two (mean) spreads: that of the bond's credit rating and that of the adjacent notch below. Our sample comprises bonds with ratings AAA– to C on the Fitch scale.

3.4 The effects of maturity

One defining characteristic of bond spreads is the existence of the term structure. As Graph 3.1 showed, there is a relationship between corporate yields and bond maturity. In this section, we ask ourselves whether carbon risk compensation may also be related to the term of each instrument. The analysis is possible thanks to the security-level approach we have taken in our models.

Looking by maturity

Table 3.3

	(1) Scope 1 emissions	(2) Scope 1+2 emissions	(3) Scope 1+2+3 emissions
Maturity < 5 years x ln(emissions)	0.11	3.87***	1.16
	[0.72]	[1.14]	[1.71]
Maturity 5-10 years x ln(emissions)	2.24***	5.99***	3.14*
	[0.71]	[1.13]	[1.70]
Maturity 10-15 years x ln(emissions)	3.03***	6.82***	3.95**
	[0.72]	[1.13]	[1.70]
Maturity 15-20 years x ln(emissions)	4.05***	7.87***	4.94***
	[0.73]	[1.14]	[1.70]
Maturity 20-25 years x ln(emissions)	3.64***	7.58***	4.70***
	[0.72]	[1.13]	[1.70]
Maturity 25-30 years x ln(emissions)	2.75***	6.83***	4.07**
	[0.73]	[1.14]	[1.70]
Maturity > 30 years x ln(emissions)	3.05***	7.16***	4.41***
	[0.76]	[1.15]	[1.71]
Number of bonds	7599	7599	7599
Observations	263,682	263,797	263,797
R-squared	0.84	0.84	0.84

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in brackets; clustered at the security level. Coefficients on other variables omitted for brevity.

Sources: Bloomberg; Refinitiv; Trucost; authors' calculations.

To answer our question, we classify bonds according to maturity. To do so, we create a dummy variable within buckets, using five-year steps. We start with bonds of less than five years in maturity, then with bonds between five and 10 years, and so forth – our last bucket comprises all bonds above 30 years. Next, we rerun our model, by interacting carbon emissions with each of these maturity buckets.

Table 3.3 presents the results, which provide evidence of a term structure of carbon premia. Across models 1, 2 and 3 – which differ in emission scopes being taken into account – we find statistical significance in most coefficients on the interaction. In other words, maturity and emissions *together* help explain the cross-section of corporate bond spreads. And the magnitude varies by term. Looking at the magnitudes, carbon risk appears to impact bonds in the 15–20 years bucket the most and shorter maturities (securities with <5 years) the least (see statistical significance).

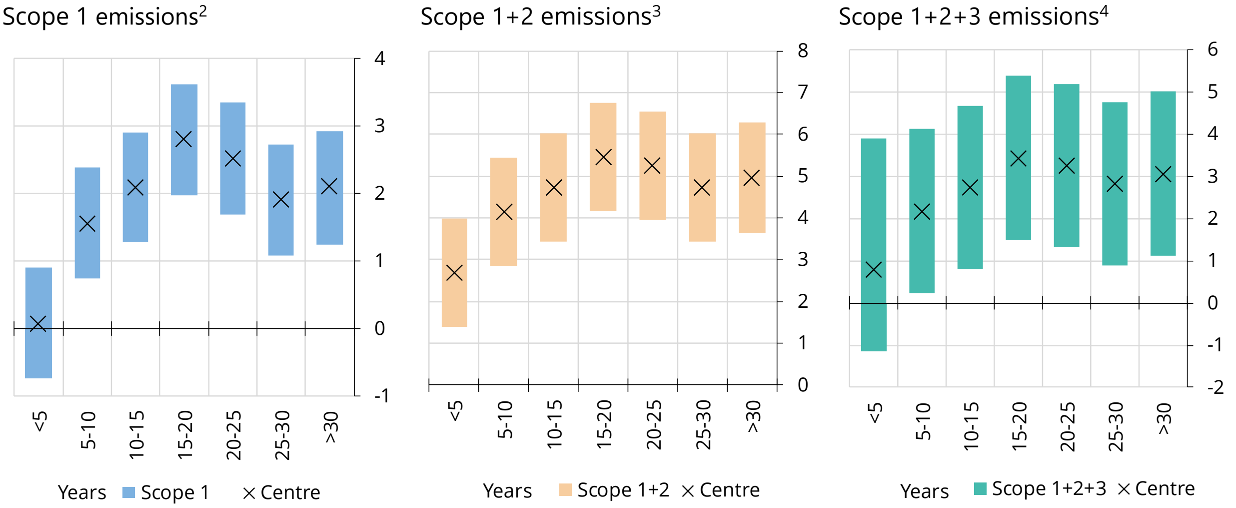
To better showcase our results, Graph 3.4 contains the decreases in spread implied by the halving of firm-level emissions under these models. By controlling scope 1 GHG (left-hand panel), firms may reduce their financing costs by some 0–3 bp, depending on maturity, on average. The effect is up to 5.5 bp – in expectation –

for scope 1 and 2 emissions together (centre panel). However, the confidence intervals denoted by the bars show that an effect of up nearly 7 bp for the belly of the curve is possible. With slightly more uncertainty (right-hand panel, wider blue bars) this effect is up to 4.7 bp, on average for maturities between 15 to 20 years when all emission scopes are taken into account.

Term structure of credit risk-adjusted carbon premia[1]

Spread decrease induced by a 50% reduction in carbon emissions, in basis points

Graph 3.4



[1] Computed as the coefficient $\hat{\beta}_{P,Carbon}$ on carbon emissions, multiplied by $ln(0.5)$.

[2] Corresponds to model (1) in Table 3.3. [3] Corresponds to model (2) in Table 3.3. [4] Corresponds to model (3) in Table 3.3.

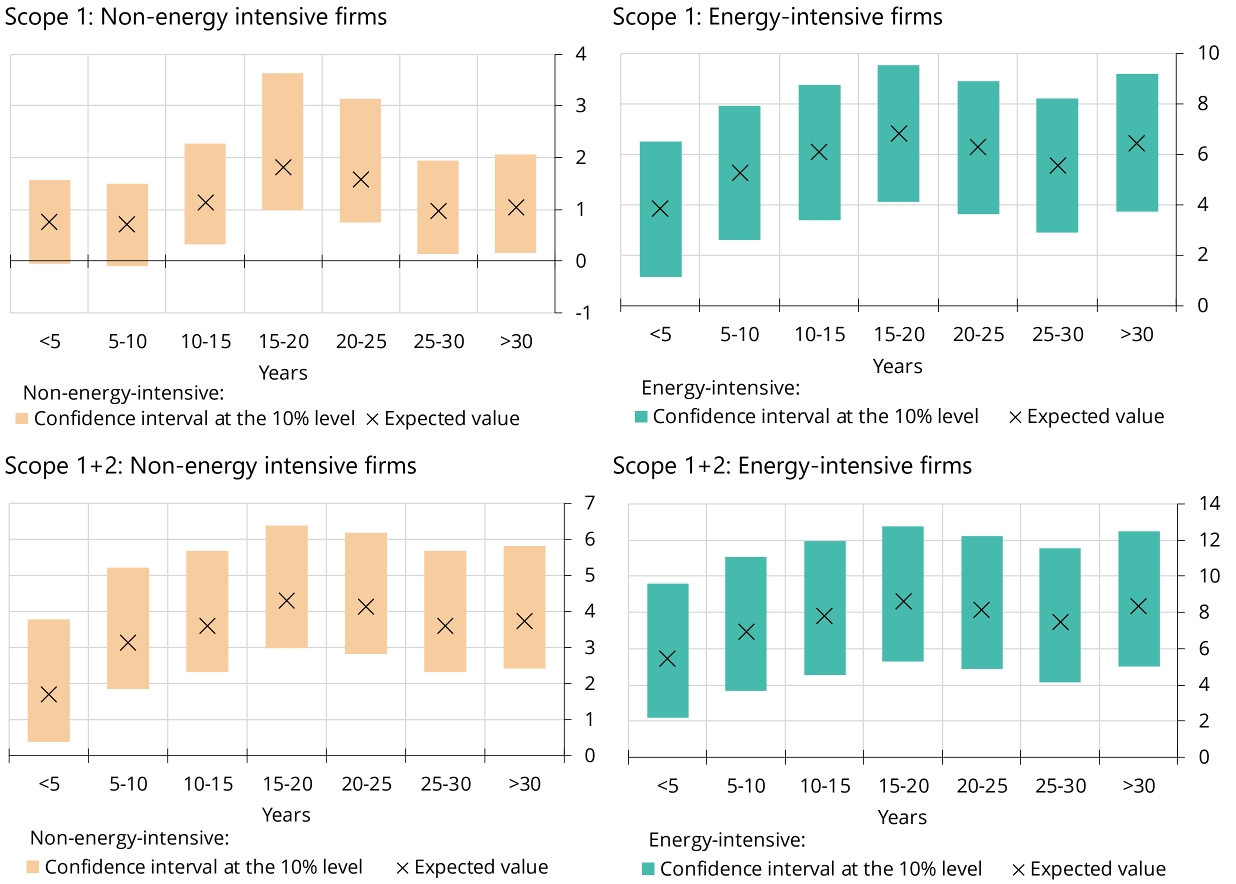
Sources: Bloomberg; Refinitiv; Trucost; authors' calculations.

Up to this point, we have found that both sector and maturity matter, independently. What happens when we explore both effects simultaneously? In what follows, we interact both indicator variables with carbon emissions, to investigate whether there are two term structures: one for energy-intensive firms and another for their complement. This time, we skip the formalities, going straight to our term structure computations for the regulatory emissions (scope 1 and scopes 1+2). We show these in Graph 3.5.

Our key finding here is that, indeed, cross-industry results do mask important differences in the term structures of non-energy-intensive (Graph 3.5, left-hand side) and energy-intensive firms (right-hand side). Focusing on scope 1+2 emissions (two bottom panels) we see that firms' bonds trade up to 4.2 basis points lower when total emissions are cut back 50% (Graph 3.5, bottom left-hand panel, white crosses, maximum value). The result is strikingly higher for energy-intensive firms (Graph 3.5, bottom right-hand panel, white crosses), where the effect can be up to 8.2 bp. In fact, our confidence intervals take our estimates – which are all statistically significant at the 1% level – to a spread effect of up to 13 bp. As in the overall sample results, the effects are more pronounced for the maturities after 15 years, the 15–20 year bucket being the greatest. The aspects of this so-called term structure of carbon premia appears rather hump-shaped.

A second glance at the term structure: by sector[1,2]

Spread decrease induced by a 50% reduction in emissions, in basis points — Graph 3.5



[1] The impact is calculated as the product between the coefficient on $ln(\text{emissions})$ and a 50% change in emissions (ie as $\hat{\beta}_{P,Carbon} \times ln(0.5)$). [2] The results corresponds to model (2) in Table 3.4.

Sources: Bloomberg; Refinitiv; Trucost; authors' calculations.

Table 3.4 presents a regression table with all specifications, grouping results by energy-intensiveness, where model (2) corresponds to Graph 3.5 above. As usual, results for scope 1 on its own (model (1)) show a smaller carbon premia than those for scope 1 and 2 together. The result holds regardless of whether a company is energy-intensive or not. When we look at the coefficients which include scope 3 GHG, the term structure appears to be inverted for non-energy-intensive firms (ie decreasing with maturity) and defined in the [2,4] bp range. For the energy-intensive case the function is mostly upward sloping and defined roughly in the range [12,16]. However, there are large differences in statistical significance for this model. For the majority of maturity groups, the coefficient fails to pass any significance test when the bond spread is from a non-energy-intensive company (upper half of the table, column 3), with some values bearing a negative sign. We interpret this erratic behaviour as one more piece of evidence that using scope 3 emissions for empirical analysis requires careful deliberation.

Twin term structures under the preference channel

Table 3.4

	(1) Scope 1 emissions	(2) Scope 1+2 emissions	(3) Scope 1+2+3 emissions
Non-energy-intensive x ln(emissions) x			
Maturity < 5 years	-1.09	2.45**	-5.76***
	[0.71]	[1.14]	[1.83]
Maturity 5-10 years	1.02	4.51***	-3.69**
	[0.71]	[1.13]	[1.83]
Maturity 10-15 years	1.64***	5.19***	-2.93
	[0.72]	[1.13]	[1.83]
Maturity 15-20 years	2.61***	6.19***	-1.93
	[0.72]	[1.14]	[1.83]
Maturity 20-25 years	2.26***	5.94***	-2.14***
	[0.73]	[1.14]	[1.82]
Maturity 25-30 years	1.40**	5.22***	-2.73
	[0.73]	[1.14]	[1.83]
Maturity > 30 years	1.49**	5.37***	-2.57
	[0.76]	[1.15]	[1.84]
Energy-intensive x ln(emissions) x			
Maturity < 5 years	5.52**	7.90***	18.33***
	[2.35]	[2.90]	[3.59]
Maturity 5-10 years	7.56***	10.03***	20.07***
	[2.34]	[2.89]	[3.58]
Maturity 10-15 years	8.74***	11.30***	21.03***
	[2.34]	[2.89]	[3.58]
Maturity 15-20 years	9.82***	12.45***	22.02***
	[2.36]	[2.91]	[3.59]
Maturity 20-25 years	9.03***	11.73***	21.69***
	[2.32]	[2.88]	[3.58]
Maturity 25-30 years	8.01***	10.77***	20.94***
	[2.34]	[2.89]	[3.58]
Maturity > 30 years	9.28***	12.04***	22.11***
	[2.40]	[2.93]	[3.60]
Number of bonds	7599	7599	7599
Observations	263,682	263,797	263,797
R-squared	0.84	0.84	0.84

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in brackets; clustered at the security level. Coefficients on other variables omitted for brevity.

Sources: Bloomberg; Refinitiv; Trucost; authors' calculations.

Our key results of this section are models (1) and (2) of Table 3.4 and encapsulate our novel finding of a term structure or "curve" of credit risk-adjusted carbon premia, which is hump-shaped and depends on a given firm's sector.[37] While the jury is still out on the curve's shape, we offer two conjectures:

a) The first is the long-term nature of environmental risks, which, despite requiring critical action today, will become inevitable in only a few years.

[37] It is "credit risk-adjusted" because, in our analysis of the preference channel, we control for the probability of defaults.

Indeed, as the race to net zero continues, it is likely that companies that pollute the most will be the first to face dramatic investor dispreference (eg fire-sale risks) should transition risks which are seemingly far away appear more material to the public.[38]

b) The second is preferred habitat. The underlying assumption is that demand and supply forces play different roles across different sectors of the curve. It may well be that investors trading on information about the environmental impact of a company may operate more in particular maturities over others. In the light of our results, this may not be the very short term. For example, pension funds, which are increasingly aware of sustainable investing, tend to have a preference for longer-term bonds in order to match their liabilities; but they may not opt for the ultra-long segments due to liquidity and interest rate risk concerns.

The differentiated impact across industries probably reflects investors' greater scrutiny of firms viewed as brown – emitting much more significant amounts of GHGs into the atmosphere is highly penalised by the market. Put together, our findings support the presence of a so-called preference channel for carbon risk in the corporate bond market.

4. The risk channel

In this section, we explore how a firm's carbon footprint affects bond spreads through the credit risk channel. In other words, we test whether a firm's default probability – established as a key determinant of corporate bond spreads in Section 3 – reflects any exposure to transition risk as measured by carbon emissions. Our hypothesis is that firms with higher GHG emissions are more exposed to transition risk and are therefore more likely to default, all else equal. Our conjecture is in line with practices at banks and rating agencies who factor a firm's environmental impact into their credit risk assessments.

4.1 The model

To assess the impact of emissions on the probability of a given firm defaulting, we run the following panel regression:

$$\tilde{P}_{i,t} = \beta_{R,carbon} \times \ln(\text{Emissions}_{i,t-12}) + \delta' X_{i,t} + FE + \varepsilon_{i,t}. \quad (4)$$

where the left-hand variable is the five-year-ahead annualised probability of default for firm i at time t, gathered from Bloomberg. This probability differs from $P_{i,t}$ in equation (3) in that $\tilde{P}_{i,t}$ is its logit transformation or "log-odds of default", which we compute as:

$$\tilde{P}_{i,t} = \ln\left(\frac{P_{i,t}}{1 - P_{i,t}}\right) \quad (5)$$

[38] This particular result should be differentiated from that with a credit risk interpretation. The effect of carbon emissions on a firm's perception of default is explored in the following section, which covers the (credit) risk channel.

We apply this transformation so that our regressand is not bounded between 0 and 1. This way, our specification is also consistent with the commonly used logit models capturing default events (eg Duffie et al (2007)). In equation (4), $X_{i,t}$ is a set of firm-level control variables, including: size of assets, long-term debt-to-assets ratio, earnings-to-assets ratio, capital-to-assets ratio and return-on-assets (RoA). FE represents a vector of time and sector fixed effects and $\varepsilon_{i,t}$ is the residual. In this new specification, $\beta_{R,Carbon}$, the coefficient in front of the one-year lagged carbon emissions, corresponds to our estimate of the risk channel. And, if our hypothesis holds, $\beta_{R,Carbon}$ should be positive.

Similar to our analysis in the previous section, we estimate the model with monthly data starting in January 2017 and consider direct (scope 1) emissions and indirect emissions jointly (scopes 1+2 and scopes 1+2+3).

4.2 Full sample results

We first run the regression using the full sample and show our results in Table 4.1. As in Section 3, we offer four baseline specifications. Columns (2) to (4) correspond to the models capturing different carbon emission scopes. Column (1) is a benchmark model that leaves out carbon emissions but is otherwise identical to equation (4) above.

Effects of carbon emissions on firm PD

Table 4.1

	(1)	(2) Scope 1 emissions	(3) Scope 1+2 emissions	(4) Scope 1+2+3 emissions
ln(emissions)		0.06***	0.07***	0.06***
		[0.01]	[0.01]	[0.01]
ln(assets)	-0.14***	-0.19***	-0.21***	-0.20***
	[0.01]	[0.01]	[0.01]	[0.01]
Long-term debt/assets	1.53***	1.57***	1.56***	1.58***
	[0.07]	[0.07]	[0.07]	[0.07]
Earnings/assets	-0.13***	-0.13***	-0.13***	-0.13***
	[0.02]	[0.02]	[0.02]	[0.02]
Capital/assets	-0.43***	-0.49***	-0.53***	-0.51***
	[0.04]	[0.04]	[0.04]	[0.04]
Return on assets (%)	-0.04***	-0.04***	-0.04***	-0.04***
	[0.001]	[0.002]	[0.002]	[0.002]
Number of firms	2910	2831	2831	2831
Observations	150,176	140,774	140,858	140,858
R-squared	0.48	0.5	0.5	0.5

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in brackets; clustered at the firm level.

Sources: Bloomberg; Refinitiv; Trucost; authors' calculations.

The benchmark model yields a lower R^2 than models with carbon emissions, suggesting that firm-level greenhouse gases do indeed play a role in credit risk assessments. For all four models, the coefficients in front of the control variables bear the right signs and are statistically significant to the 1% level. For instance, as expected, a broader balance sheet, higher earnings, stronger capital and increased RoA are all related to lower probabilities of default. The opposite is true for the ratio of long-term debt to assets. Moreover, the magnitude of these coefficients is broadly similar across the different specifications, implying that our measure of carbon emissions represents information not contained in the set of other firm-specific characteristics considered.

Zooming in on models that consider carbon emissions, these baseline results confirm our hypothesis – a firm's emitted level of CO_2 and other greenhouse gases has an adverse impact on its probability of default. For all carbon emission measures, $\beta_{R,carbon}$ is estimated to be positive and statistically significant with at least 99% confidence. The magnitude of our estimate $\hat{\beta}_{R,carbon}$ is roughly similar regardless of which scopes are included in the computation. This finding suggests there may be limited differentiation between direct and indirect carbon emissions when it comes to assessing transition risk.

Given that our model is one of log-odds of default, how should these coefficients be interpreted in terms of default probability levels? Take scope 1 carbon emissions, for example. Halving emissions would translate to a 0.03 decrease in log-odds $\tilde{P}_{i,t}$. However, since the relationship between $\tilde{P}_{i,t}$ and $P_{i,t}$ is non-linear, the impact of carbon emissions on default probability (and thus, option-adjusted spreads) is not a constant. To compute the resulting effect on probability levels, we instead proceed as below.

First, we estimate the change on log-odds $\Delta\tilde{P}_i$ induced by a change in log-emissions ΔE. This is obviously a function of our coefficient $\hat{\beta}_{R,carbon}$:

$$\Delta\tilde{P}_i = \hat{\beta}_{R,carbon} \times \Delta E \tag{6}$$

Next, and for each probability level p_i, we compute its log-odds and then add the change in log-odds $\Delta\tilde{P}_i$ to obtain a new log-odds probability level $\tilde{P}_i'$. From equation (5), this is:

$$\tilde{P}_i' = \Delta\tilde{P}_i + \ln\left(\frac{p_i}{1-p_i}\right) \tag{7}$$

Finally, we apply the inverse logit transformation to our estimate $\tilde{P}_i'$ to derive the resulting default probability level p_i'. We lastly compare this with our original probability level p_i to get the estimated change in probability level Δp_i:

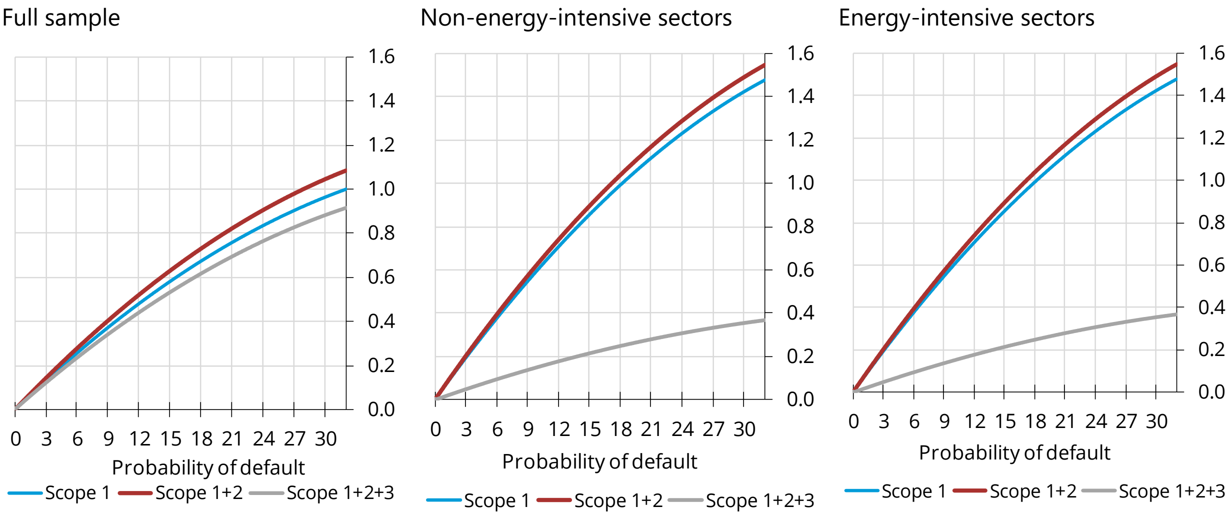
$$\Delta p_i = p_i' - p_i = \frac{1}{1+e^{\tilde{P}_i'}} - p_i \tag{8}$$

For our purposes, we define p_i continuously in the range [0, 0.32]. In other words, we only consider probabilities of default of up to 32% as it is the upper bound for this variable in our sample (see Section 2.1 for details).

The left-hand panel of Graph 4.1 plots our resulting estimates for each level of default probability when emissions are halved. Within the observed range, we notice that the impact increases with default likelihood. At the maximum probability of 32%, halving emissions – both direct and combined – translates to a 1 pp decrease in probability. These estimates can be used to derive the impact on option-adjusted spreads, which we will discuss in the next section.

Decline in five-year default probabilities induced by a 50% reduction in emissions

In percentage points | Graph 4.1



[1] The effect of a 50% decrease in carbon emissions is computed continuously for each probability level on the scale 0 to 32%.

Sources: Bloomberg; Refinitiv; Trucost; S&P Capital IQ; authors' calculations.

4.3 Energy-intensive vs non-energy-intensive sectors

As in the preference channel, we also wish to test whether the impact of emissions on corporate default is only viable in a few sectors, specifically the ones commonly viewed as "brown" industries with heavy GHG emissions. To this end, we consider two analyses, as before:

a. A subsample analysis, where we run regressions exclusively for firms belonging to energy-intensive sectors and for those in non-energy-intensive sectors, respectively.

b. A model which adds an interaction term between carbon emissions and whether a firm is from an energy-intensive sector or not. This adds a dummy to the regressor list in equation (4) above. Our classification of energy-intensive and non-energy-intensive sectors is identical to the previous section.

Table 4.2 shows the results. Columns (1), (4) and (7) are estimates based on the subsample of non-energy-intensive sectors; columns (2), (5) and (8) are estimates using the subsample of energy-intensive sectors; and columns (3), (6) and (9) are derived from models with the dummy interaction term.

Our results suggest that the impact of carbon emissions through the risk channel prevails across different sectors with a greater effect on energy-intensive sectors. In both subsamples, carbon emissions play a negative and statistically significant role in the probability of default. The results from regressions with interactions are consistent with the results from subsample analysis. For scope 1 and scope 1+2 emissions, their impact on the probability of default is larger in energy-intensive sectors. For scope 1 emissions, a 50% increase of the emissions would translate to a 0.04 increase in $\tilde{P}_{i,t}$ in non-energy-intensive sectors and a 0.07 increase in $\tilde{P}_{i,t}$ in energy-intensive sectors. These correspond to 0.9% and 1.5% increases in the probability of default at probability of 32% (centre and right-hand panels of Graph 4.1). The impact of scope 1+2 emissions is quite similar to that of scope 1 emissions: 1% and 1.6% respectively for non-energy–intensive and energy–intensive sectors. Interestingly, for scope 1+2+3 emissions, the risk channel impact on non-energy–intensive sectors is similar to that of other emission measures but the impact on energy-intensive sectors is relatively smaller. This could partly reflect data quality issues with scope 3 emissions.

Table 4.2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Scope 1 emissions			Scope 1+2 emissions			Scope 1+2+3 emissions		
ln(emissions)	0.06***	0.09***		0.06***	0.10***		0.07***	0.02***	
	[0.01]	[0.02]		[0.01]	[0.03]		[0.01]	[0.03]	
Non-energy-intensive x ln(ems.)			0.06***			0.06***			0.06***
			[0.01]			[0.01]			[0.01]
Energy-intensive x ln(emissions)			0.08***			0.09			0.05**
			[0.01]			[0.01]			[0.02]
ln(assets)	-0.19***	-0.24***	-0.19***	-0.20***	-0.25***	-0.21***	-0.21***	-0.17***	-0.20***
	[0.01]	[0.03]	[0.01]	[0.01]	[0.03]	[0.01]	[0.01]	[0.04]	[0.01]
Long-term debt/assets	1.57***	1.63***	1.57***	1.56***	1.64***	1.56***	1.55***	1.85***	1.58***
	[0.08]	[0.21]	[0.07]	[0.08]	[0.21]	[0.07]	[0.08]	[0.26]	[0.07]
Earnings/assets	-0.12***	-0.19***	-0.13***	-0.11***	-0.18***	-0.13***	-0.12***	-0.13***	-0.13***
	[0.02]	[0.07]	[0.02]	[0.02]	[0.07]	[0.02]	[0.02]	[0.08]	[0.02]
Capital/assets	-0.49***	-0.48*	-0.49***	-0.53***	-0.51*	-0.53***	-0.52***	-0.39***	-0.51***
	[0.04]	[0.27]	[0.04]	[0.04]	[0.27]	[0.04]	[0.04]	[0.25]	[0.04]
Return on assets (%)	-0.04***	-0.02***	-0.04***	-0.04***	-0.02***	-0.04***	-0.04***	-0.03***	-0.04***
	[0.002]	[0.003]	[0.002]	[0.002]	[0.003]	[0.002]	[0.002]	[0.004]	[0.002]
Number of firms	2,443	388	2,831	2,443	388	2,831	2,466	365	2,831
Observations	121,124	19,650	140,774	121,208	19,650	140,858	122,840	18,018	140,858
R-squared	0.49	0.54	0.5	0.49	0.54	0.5	0.48	0.56	0.5

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in brackets; clustered at the firm level.

Sources: Bloomberg; Refinitiv; Trucost; authors' calculations.

5. Combining the two channels

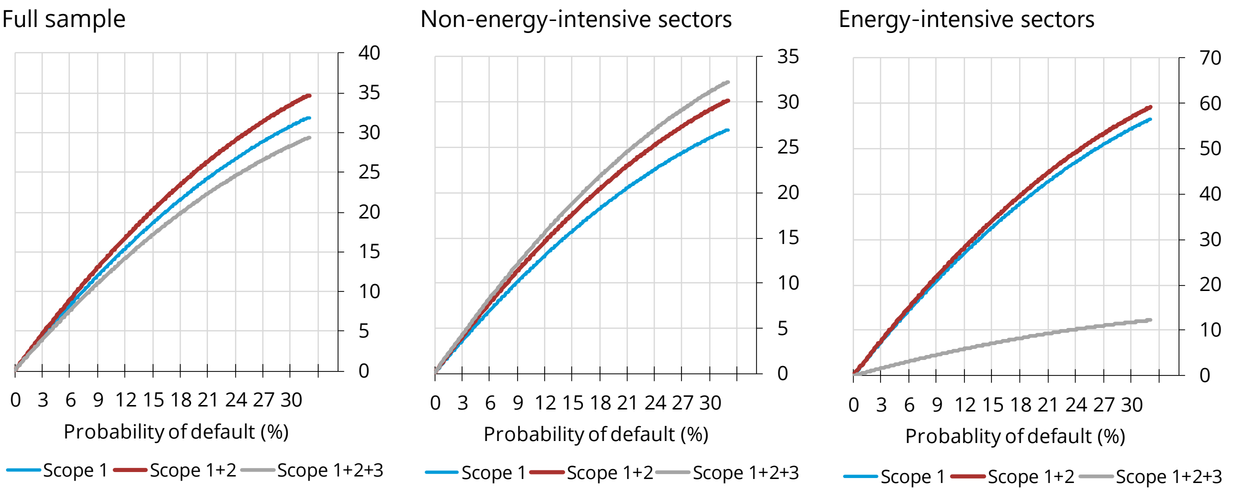
In our previous sections, we derived estimates for two types of carbon premium in corporate bonds: the credit risk-adjusted carbon premia and the credit risk carbon premia.

This section explores the *total* carbon premium that comes out of combining the effects above, which represent the preference and risk channels. To compute total premia, we first need to translate the impact of emissions on default probabilities (the risk channel) into an effect on option-adjusted spreads. This can be achieved by multiplying the estimated risk channel effect on default probabilities (Δp_i from equation (8)) by the effect of this probability on spreads ($\hat{\beta}_P$ from equation (3)).

Graph 5.1 showcases an example. For the full sample (left-hand panel) and for non-energy-intensive sectors (centre panel), a halving of carbon emissions would narrow corporate spreads by as much as 30 basis points for both direct and combined emissions (several scopes together). For energy-intensive firms, the impact is much larger. For scope 1 and scopes 1+2, a halving of carbon emissions could translate to a near 60 basis points decline in spreads.

Decline in OAS induced by a 50% reduction in carbon emissions

In basis points — Graph 5.1



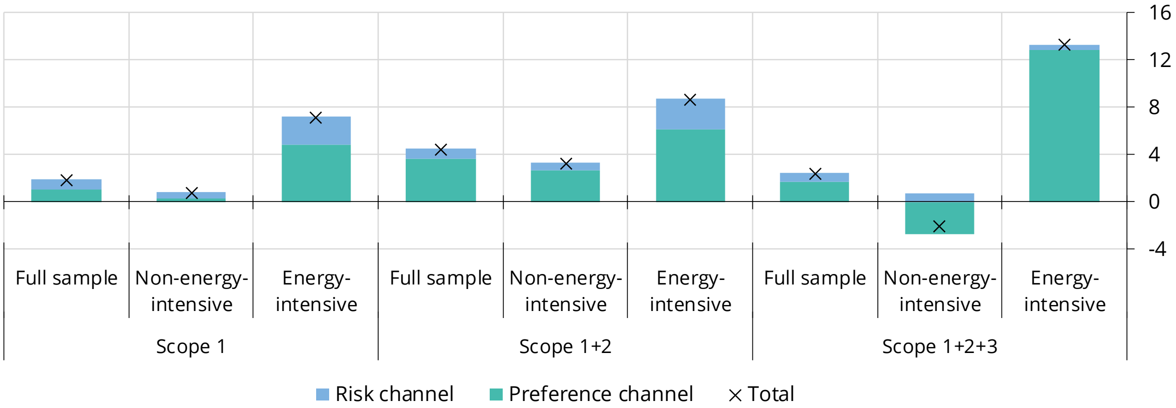
Sources: Bloomberg; Refinitiv; Trucost; S&P Capital IQ; authors' calculations.

With the effect of the risk channel expressed in terms of the spread, we can now combine it with the impact of the preference channel. To this end, we consider a typical firm, whose probability of default equals the sample average, which is 0.56%. Based on this, Graph 5.2 plots the total impact on spreads of a halving in firm-level GHG across different emission measures and sectors. For the full sample (first bar of each graph section), the total impact ranges from 2 to 4.5 basis points, depending on the emission scopes considered. Looking at the bar colours, we note that total premia are mainly explained by the preference channel (green bars) rather than the risk channel (blue bars). Of course, this is a function of the probability of default *level* of the company in question. For the average firm, the PD is low (recall it is approximately 0.56%). This attribution of the total premia changes when we look at a different PD level. For instance, when the PD is one standard deviation above the average (1.54%), the contributions from the two channels are on more equal footing.

Comparing across different sectors, the total impact appears larger for energy-intensive firms. Concretely: for a typical firm in an energy-intensive sector, the impact is around 8 basis points for scope 1 and scope 1+2 emissions and more than 13 basis points for scope 1+2+3 emissions combined. In comparison, for a typical firm in the non-energy-intensive category, the impact is at most 3 basis points for scope 1 and 2 combined.

Combined impact on spreads of a 50% reduction in carbon emissions

In basis points Graph 5.2



Sources: Bloomberg; Refinitiv; Trucost; S&P Capital IQ; authors' calculations.

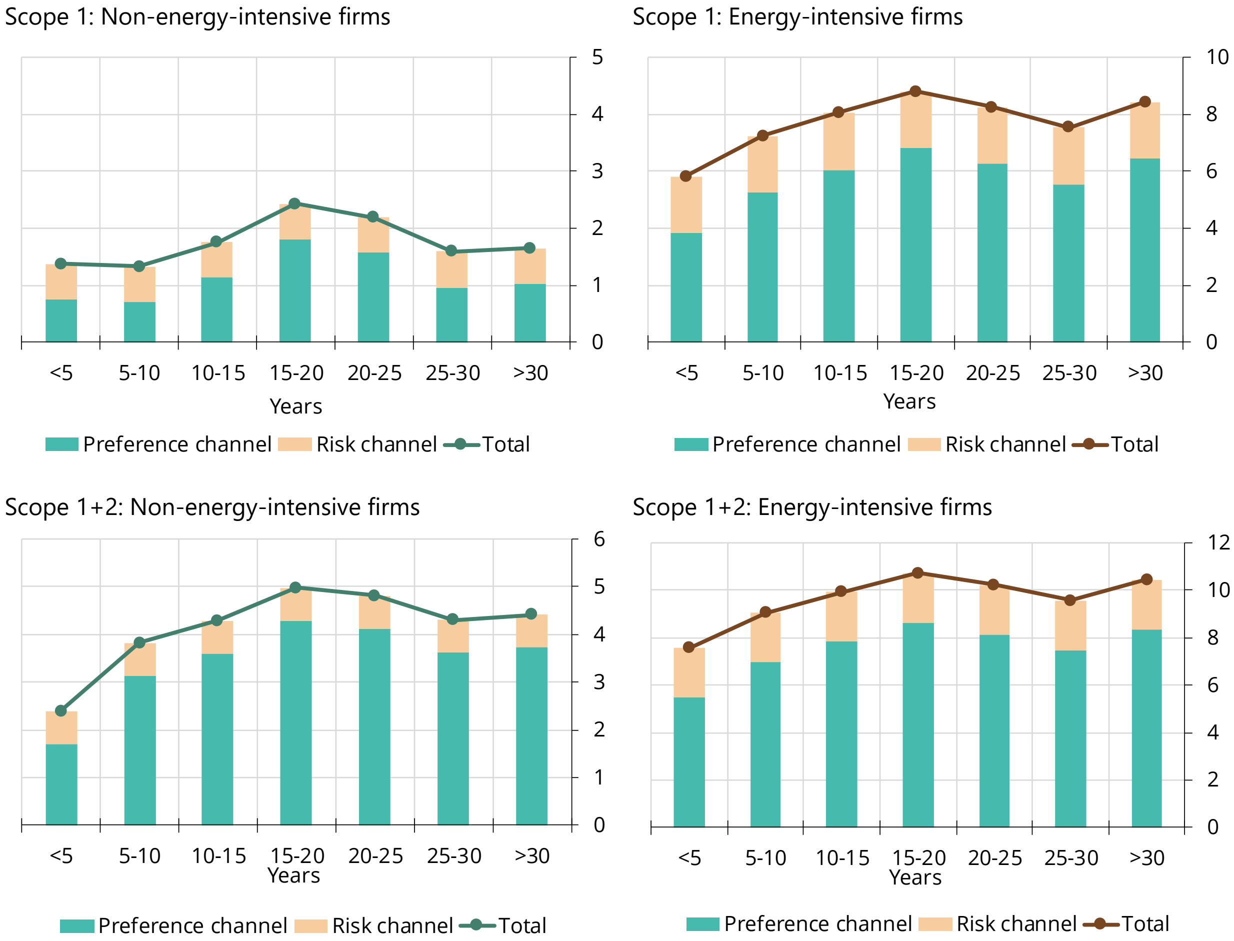
We then look into the term structure of total carbon premia. Recall that the term structure coming from the preference channel is hump-shaped. The addition of the risk channel largely preserves this hump shape (results available upon request). This is because the impact of carbon emissions on spreads via credit risk depends on the impact of emissions on a firm's default probability and on the mapping of a firm's default probability to the spreads of bonds issued by the company. The former is independent of maturity, as default probability is gauged at the firm level and not the bond level, while the latter is broadly identical across different maturities.[39] With the upward shift induced by the addition of the risk channel, the total carbon premium in the belly (the maturity bucket being 15–20 years) is in the range of 3.5 to 6 basis points, depending on emission measures.

Last, we separately contrast the term structures of total carbon premia for non-energy-intensive and energy-intensive sectors. Graph 5.3 presents our results for scope 1 (top panels) and scope 1+2 (bottom panels). For brevity, we focus on the results of scope 1+2 emissions – the measure yielding the largest total carbon premia. As can be seen in Graph 5.3 (bottom two panels), the term structure is hump-shaped for both non-energy-intensive and energy-intensive categories. Across the maturity spectrum, total carbon premia are almost twice as large for energy-intensive firms than for non-energy-intensive ones. For bonds maturing in 15–20 years, a halving in scope 1+2 emissions would narrow their spreads by more than 10 basis points. For "greener" firms, the effect is around 5 basis points.

39 We have also estimated $\hat{\beta}_P$ from a more flexible model in which we interact carbon emissions with the maturity bucket indicator. The results are very close. For brevity, they are not presented here but are available upon request.

Term structure of total carbon premia: by sector[1]

Spread decrease induced by a 50% reduction in emissions, in basis points | Graph 5.3



[1] Effect of a 50% reduction in the respective emission scope totals.

Source: Bloomberg; Refinitiv; Trucost; S&P Capital IQ; authors' calculations.

6. Robustness checks

In Sections 3 and 4, we explored whether our findings hold when excluding securities from the most carbon-intensive sectors. This led us to conclude that our results are robust to these formulations and that we can in fact compute two different term structures of total carbon premia. To further validate our findings, we have conducted other robustness checks, which we present in this subsection.

6.1 Robustness checks for the preference channel

We consider two robustness checks for the preference channel, with alternative measures of default probability and liquidity, respectively.

The first check consists of swapping Bloomberg's measure of probability of default for our own computations in the preference channel model. As detailed in Section 2, our own computations are based on the work of Merton (1974).[40] This procedure returns an alternative metric of five-year-ahead default probability, which we use as regressor in place of the Bloomberg ready-made ones.

The regression models using our computed default probabilities are in Table 6.1. Before looking at the carbon emission results, we review any changes to the model without carbon emissions (column 1). What are the most important changes? In particular, a rise of 1 percentage point in the risk-neutral default probability can be translated to an increase of 3 basis points in bid-ask spreads. This is about a 10th of the effect found in physical default probabilities. The result is intuitive, since in the risk-neutral world – and because investors are in aggregate risk-averse – prices imply higher probabilities to negative scenarios than they do to positive ones. Another important change is in equity return volatility, which now has a coefficient 1.7 times the original. The rest of the coefficients show similiar magnitudes across the board. The impact of duration, age, coupons, outstanding amounts and liquidity appear close to our baseline estimates (see Table 3.1, for example).

We move on to models (columns 2–4) that capture estimates for our term structure of credit risk-adjusted carbon premia with risk-neutral default probabilities. All models show that maturity, sector and carbon emissions are statistically significant, helping to explain corporate spreads. The correlation between carbon emissions and corporate spreads is positive, as in the core results. In terms of magnitude, the coefficients for bonds from non-energy-intensive companies are in the 1 to 3 range – close to our original results. For energy-intensive companies, the effect appears somewhat higher, with coefficients reaching a level of up to above 13 (10 in the original model). The twin term structures appear hump-shaped, nonetheless. In summary, our choice of default probability does not drive our results, which appear to hold in both the physical and risk-neutral worlds.

The second robustness check involves varying our liquidity measure. Chen et al (2007) use several measures of liquidity to show that the notion is priced in the cross section of corporate bond spreads. In our paper, we have chosen the absolute measure of Roll (see Appendix 2) as our preferred liquidity variable, given the simplicity of its computation and availability of the required data. In these alternative specifications, we explore whether using *observed* bid-ask spreads (as opposed to *synthetic* ones) affects our results. To this end, we gather close bid and ask yields for the corporate bonds in our sample from Bloomberg, and compute bid-ask spreads to re-run our key models. Table 6.2 summarises our estimates.

40 The theory behind the approach is that the equity of a firm can be viewed as a call option on the underlying value of the firm, with a strike price equal to the face value of the firm's debt. In brief: given a time series for the value of equity and liabilities for a particular firm, we can calibrate its corresponding asset values, the volatility of assets, and the probability of default. See Appendix 1 for details.

Preference channel models using risk-neutral PD

Table 6.1

	No emissions	Scope 1 emissions	Scope 1+2 emissions	Scope 1+2+3 emissions
Risk-neutral default probability (%)	3.09***	3.28***	3.31***	3.14***
	[0.39]	[0.38]	[0.38]	[0.37]
Duration	4.94***	2.45***	1.93***	1.48***
	[0.12]	[0.25]	[0.27]	[0.28]
Age	0.40***	0.36***	0.37***	0.37***
	[0.14]	[0.13]	[0.13]	[0.13]
Coupon	11.05***	9.28***	9.09***	8.91***
	[0.48]	[0.44]	[0.43]	[0.44]
ln(amount outstanding)	−2.83***	−2.91***	−2.92***	−2.94***
	[0.31]	[0.29]	[0.29]	[0.29]
Equity return volatility	31.08***	31.01***	30.99***	31.18***
	[0.92]	[0.91]	[0.91]	[0.90]
Liquidity	0.50***	0.47***	0.47***	0.47***
	[0.02]	[0.02]	[0.02]	[0.02]
Callable	−7.97***	−6.98***	−6.78***	−6.99***
	[1.00]	[0.95]	[0.94]	[0.94]
Non-energy-intensive x ln(emissions) x				
Maturity < 5 years		−0.57	2.66**	−10.21***
		[0.75]	[1.14]	[1.94]
Maturity 5-10 years		1.56**	4.76***	−8.10***
		[0.75]	[1.13]	[1.94]
Maturity 10-15 years		2.22***	5.48***	−7.31***
		[0.75]	[1.14]	[1.95]
Maturity 15-20 years		3.25***	6.53***	−6.25***
		[0.76]	[1.14]	[1.94]
Maturity 20-25 years		2.90***	6.29***	−6.46***
		[0.76]	[1.14]	[1.94]
Maturity 25-30 years		2.07***	5.61***	−7.00***
		[0.77]	[1.14]	[1.95]
Maturity > 30 years		2.24***	5.83***	−6.77***
		[0.79]	[1.16]	[1.96]
Energy-intensive x ln(emissions) x				
Maturity < 5 years		8.96***	12.60***	24.08***
		[2.58]	[3.18]	[3.95]
Maturity 5-10 years		11.00***	14.73***	25.84***
		[2.57]	[3.17]	[3.94]
Maturity 10-15 years		12.20***	16.03***	26.83***
		[2.56]	[3.17]	[3.93]
Maturity 15-20 years		13.33***	17.23***	27.88***
		[2.59]	[3.19]	[3.95]
Maturity 20-25 years		12.57***	16.55***	27.56***
		[2.55]	[3.16]	[3.94]
Maturity 25-30 years		11.52***	15.57***	26.81***
		[2.57]	[3.17]	[3.94]
Maturity > 30 years		12.85***	16.89***	28.02***
		[2.62]	[3.21]	[3.96]
Number of bonds	7,633	7596	7,596	7,596
Observations	266,133	263,771	263,886	263,886
R-squared	0.83	0.83	0.83	0.83

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in brackets; clustered at the security level.

Sources: Bloomberg; Refinitiv; Trucost; authors' calculations.

Preference channel models using an alternative liquidity measure

Table 6.2

	No emissions	Scope 1 emissions	Scope 1+2 emissions	Scope 1+2+3 emissions
Default probability (%)	31.65***	31.91***	31.86***	31.59***
	[1.274]	[1.249]	[1.250]	[1.245]
Duration	5.236***	2.867***	2.375***	1.939***
	[0.116]	[0.249]	[0.267]	[0.277]
Age	0.569***	0.531***	0.535***	0.532***
	[0.134]	[0.126]	[0.125]	[0.126]
Coupon	10.51***	8.760***	8.585***	8.418***
	[0.470]	[0.420]	[0.417]	[0.420]
ln(amount outstanding)	-2.884***	-2.960***	-2.971***	-2.978***
	[0.313]	[0.288]	[0.286]	[0.289]
Equity return volatility	17.90***	17.87***	17.90***	18.05***
	[0.928]	[0.927]	[0.924]	[0.922]
Bid-ask spread	0.434***	0.407***	0.405***	0.405***
	[0.0169]	[0.0160]	[0.0159]	[0.0158]
Callable	-7.999***	-7.148***	-6.954***	-7.144***
	[0.987]	[0.933]	[0.926]	[0.927]
Non-energy-intensive x ln(emissions) x				
Maturity < 5 years		-1.088	2.449**	-5.760***
		[0.713]	[1.137]	[1.830]
Maturity 5-10 years		1.015	4.511***	-3.685**
		[0.706]	[1.129]	[1.825]
Maturity 10-15 years		1.637**	5.192***	-2.933
		[0.716]	[1.132]	[1.830]
Maturity 15-20 years		2.610***	6.187***	-1.926
		[0.719]	[1.135]	[1.826]
Maturity 20-25 years		2.263***	5.944***	-2.141
		[0.727]	[1.137]	[1.824]
Maturity 25-30 years		1.397*	5.219***	-2.727
		[0.729]	[1.139]	[1.826]
Maturity > 30 years		1.490**	5.368***	-2.567
		[0.758]	[1.153]	[1.835]
Energy-intensive x ln(emissions) x				
Maturity < 5 years		5.519**	7.900***	18.33***
		[2.349]	[2.901]	[3.591]
Maturity 5-10 years		7.557***	10.03***	20.07***
		[2.341]	[2.893]	[3.583]
Maturity 10-15 years		8.740***	11.30***	21.03***
		[2.337]	[2.891]	[3.580]
Maturity 15-20 years		9.818***	12.45***	22.02***
		[2.360]	[2.910]	[3.592]
Maturity 20-25 years		9.031***	11.73***	21.69***
		[2.321]	[2.878]	[3.582]
Maturity 25-30 years		8.007***	10.77***	20.94***
		[2.338]	[2.891]	[3.582]
Maturity > 30 years		9.279***	12.04***	22.11***
		[2.395]	[2.927]	[3.604]
Number of bonds	7,642	7599	7,599	7,599
Observations	266,241	263,682	263,797	263,797
R-squared	0.837	0.844	0.844	0.844

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in brackets; clustered at the security level.

Sources: Bloomberg; Refinitiv; Trucost; authors' calculations.

This shows that: (1) the coefficient on observed bid-asks is statistically significant across all specifications, (2) it is positive, as expected; and (3) the order of magnitude is very close to that of our synthetic bid-ask measure. This is an encouraging outcome for the absolute measure of Roll as a liquidity proxy. Next, we focus on scope 1 and 2 emissions. With regard to their effect on spreads, the liquidity variable change induces minimal changes in the statistical power and magnitude of the tests. Our carbon premia results appear consistent with our choice of bid-ask spread.

6.2 Robustness checks for the risk channel

In testing the risk channel, we also consider the risk-neutral default probabilities that we compute on our own, to measure credit risk. The results are shown in Table 6.3.

The result showing that the risk channel is at work in both energy-intensive and non-energy-intensive sectors is robust to this alternative measure of default risk. The coefficients in front of scope 1 and scope 1+2 emissions for both energy and non-energy firms are positive and statistically significant. The coefficients in front of scope 1+2+3 emissions, however, lost significance. Comparing different sectors, coefficients in energy-intensive sectors are greater those in non-energy-intensive sectors, consistent with our main result in Section 4. Unsurprisingly, as risk-neutral default probabilities reflect both the probability of default *and* the default risk premium, the coefficients in front of both carbon emissions and control variables are larger.

Risk channel models using risk-neutral PD

Table 6.3

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Scope 1 emissions			Scope 1+2 emissions			Scope 1+2+3 emissions		
ln(emissions)	0.162	0.701***		0.210*	0.790***		-0.00716	0.141	
	[0.103]	[0.187]		[0.108]	[0.213]		[0.114]	[0.331]	
Non-energy-intensive x ln(emissions)			0.170*			0.215**			-0.0163
			[0.0973]			[0.101]			[0.110]
Energy-intensive x ln(emissions)			0.538***			0.585***			-0.0348
			[0.143]			[0.155]			[0.225]
ln(assets)	-0.882***	-1.31***	-0.91***	-0.94***	-1.41***	-0.97***	-0.73***	-0.652	-0.68***
	[0.124]	[0.278]	[0.113]	[0.135]	[0.301]	[0.122]	[0.136]	[0.424]	[0.130]
long-term debt/assets	9.072***	8.957***	8.902***	9.052***	9.070***	8.892***	9.241***	7.597***	8.926***
	[0.629]	[1.568]	[0.581]	[0.628]	[1.566]	[0.581]	[0.628]	[1.708]	[0.583]
Earnings/assets	-0.31**	-0.449	-0.37***	-0.303**	-0.421	-0.36***	-0.337**	-0.755	-0.35***
	[0.148]	[0.363]	[0.135]	[0.147]	[0.361]	[0.134]	[0.146]	[0.539]	[0.137]
Capital/assets	-2.341***	-3.629**	-2.39***	-2.46***	-3.859**	-2.51***	-2.24***	-4.041*	-2.29***
	[0.410]	[1.727]	[0.389]	[0.417]	[1.714]	[0.394]	[0.405]	[2.101]	[0.389]
Return on assets	-0.181***	-0.074**	-0.15***	-0.18***	-0.076**	-0.16***	-0.16***	-0.17***	-0.15***
	[0.0152]	[0.0296]	[0.0136]	[0.0153]	[0.0296]	[0.0136]	[0.0151]	[0.0420]	[0.0140]
Number of firms	2,295	374	2,670	2,295	374	2,670	2,324	345	2670
Observations	57,788	10,832	68,620	57,832	10,832	68,664	58,923	9,741	68664
R-squared	0.444	0.437	0.441	0.444	0.437	0.441	0.458	0.352	0.438

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in brackets; clustered at the firm level.

Sources: Bloomberg; Refinitiv; Trucost; authors' calculations.

6.3 Robustness checks for total carbon premia

Our main analysis calculates total carbon premia in two steps. Indeed, premia through the preference and risk channels are computed, separately, in a first step, and then combined in a second step. However, as a robustness check, we can also estimate total carbon premia in one go. To do this, we simply need to swap our measure of default for a series of firm-level variables. This takes our regression model closer to those exploring the determinants of corporate spreads without isolating the risk channel. In other words, when doing this, our model looks less like Gilchrist and

Zakrajšek's (2012) and more like those found in Elton et al (2001), Campbell and Taksler (2003), and Chen et al (2007).

Concretely, we replace firm default probability by the natural logarithm of the firm's assets, its ratio of long-term debt to assets, its ratio of earnings to assets, its ratio of capital to assets and its return-on-assets. Table 6.4 shows our results. Overall, forecasting power does not suffer and the statistical significance on the triple interaction (sector, maturity, emissions) is preserved. Also, the coefficients on energy-intensive bonds appear slightly higher. Qualitatively, this alternative set of specifications does not alter our findings. Quantitatively, our estimated total impact is somewhat larger. For example, according to the estimates on Table 6.4, a halving in scope 1+2 carbon emissions would narrow spreads by around 8 and 18 basis points for non-energy-intensive firms and energy-intensive firms, respectively, at the belly of the curve. In contrast, our estimates in Section 5 suggest total effects of 5 and 10 basis points, respectively.

A simple model to compute total carbon premia (continues on next page) Table 6.4

	(1) No emissions	(2) Scope 1 emissions	(3) Scope 1+2 emissions	(4) Scope 1+2+3 emissions
ln(assets)	-26.44	-49.37**	-58.22***	-44.57**
	[20.44]	[19.73]	[19.58]	[20.25]
Long-term debt/assets	115.2***	117.9***	118.2***	116.5***
	[6.946]	[6.767]	[6.746]	[6.688]
Return on assets (%)	-1.739***	-1.737***	-1.761***	-1.700***
	[0.108]	[0.107]	[0.107]	[0.107]
Earnings/assets	-24.26***	-26.61***	-25.50***	-24.94***
	[3.514]	[3.286]	[3.237]	[3.265]
Capital/assets	12.31**	13.48***	13.14***	13.03***
	[4.974]	[4.914]	[4.919]	[4.917]
Duration	5.009***	2.404***	1.859***	1.381***
	[0.118]	[0.250]	[0.268]	[0.278]
Age	0.507***	0.475***	0.475***	0.474***
	[0.137]	[0.128]	[0.128]	[0.129]
Coupon	11.01***	9.187***	8.998***	8.827***
	[0.484]	[0.435]	[0.433]	[0.436]
ln(amount outstanding)	-2.959***	-3.044***	-3.055***	-3.062***
	[0.315]	[0.291]	[0.289]	[0.292]
Equity return volatility	29.09***	29.13***	29.12***	29.27***
	[0.874]	[0.857]	[0.853]	[0.850]
Liquidity	0.473***	0.445***	0.442***	0.442***
	[0.0178]	[0.0168]	[0.0167]	[0.0167]
Callable	-7.939***	-6.854***	-6.662***	-6.847***
	[1.008]	[0.955]	[0.947]	[0.949]

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in brackets; clustered at the firm level. Sources: Bloomberg; Refinitiv; Trucost; authors' calculations. Table continues on next page.

A simple model to compute total carbon premia (continued)

Table 6.4

	(1) No emissions	(2) Scope 1 emissions	(3) Scope 1+2 emissions	(4) Scope 1+2+3 emissions
Non-energy-intensive x ln(emissions) x				
Maturity < 5 years		0.212	3.947***	-3.606*
		[0.849]	[1.252]	[1.977]
Maturity 5-10 years		2.366***	6.073***	-1.456
		[0.843]	[1.244]	[1.975]
Maturity 10-15 years		3.054***	6.829***	-0.628
		[0.851]	[1.246]	[1.980]
Maturity 15-20 years		4.090***	7.889***	0.449
		[0.854]	[1.248]	[1.978]
Maturity 20-25 years		3.806***	7.714***	0.281
		[0.861]	[1.250]	[1.979]
Maturity 25-30 years		3.000***	7.048***	-0.246
		[0.863]	[1.251]	[1.982]
Maturity > 30 years		3.131***	7.250***	0.00527
		[0.887]	[1.263]	[1.990]
Energy-intensive x ln(emissions) x				
Maturity < 5 years		9.830***	13.24***	21.55***
		[2.372]	[2.971]	[3.870]
Maturity 5-10 years		11.92***	15.43***	23.35***
		[2.364]	[2.964]	[3.862]
Maturity 10-15 years		13.11***	16.72***	24.34***
		[2.360]	[2.961]	[3.860]
Maturity 15-20 years		14.22***	17.91***	25.38***
		[2.378]	[2.975]	[3.871]
Maturity 20-25 years		13.48***	17.25***	25.16***
		[2.358]	[2.960]	[3.864]
Maturity 25-30 years		12.53***	16.37***	24.47***
		[2.364]	[2.964]	[3.862]
Maturity > 30 years		13.83***	17.66***	25.54***
		[2.418]	[2.996]	[3.884]
Number of bonds	7,396	7359	7,359	7,359
Observations	257,092	254,735	254,850	254,850
R-squared	0.83	0.838	0.838	0.838

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in brackets; clustered at the firm level. Sources: Bloomberg; Refinitiv; Trucost; authors' calculations. The first part of this table is on the previous page.

7. Conclusions

In theory, corporate bond spreads represent compensation for bearing credit risk. In practice, they represent compensation for much more. As discussed, they encapsulate, among other things, compensation for the probability of default, lack of liquidity, higher volatility in firm value and, in the light of our results, increased firm-level pollution, as captured by greenhouse gas emissions.[41] The effect of firm-level

[41] Our result covers publicly traded companies exclusively.

emissions on corporate bond pricing has two aspects: firstly, regarding investor preferences; and secondly, regarding credit risk. We call these the preference and risk channels, respectively.

From a qualitative standpoint, the two channels arise for different reasons. First, investors may prefer holding debt issued by firms that are more environmentally friendly (vis-à-vis that of those that are not). This phenomenon evokes that of the liquidity premium, where on-the-run securities may be preferred to off-the-run ones, and compensation is due. Second, regardless of whether a firm is more favoured than another, some companies may be more exposed to risks during the transition to a low-carbon world. Carbon taxes, consumer preferences and technological change are only some of the factors that, if not planned for, could affect corporate financial health and therefore firm-level default risk.

From a quantitative standpoint, we find statistically significant evidence of both phenomena. In terms of economic significance, the impact is larger for energy-intensive firms in both channels. In addition, we find that the term structure of carbon premia – encapsulating both the preference channel and risk channels – is hump-shaped, with the largest premia at the belly of the curve (15–20 years). For a bond in this maturity bucket, which is issued by an energy-intensive firm, a halving of firm-level GHG emissions can reduce its spread by over 10 basis points.

Our results highlight the role of capital markets in the transition to net zero. Documenting the existence of a carbon premium provides evidence that investors differentiate between firms based on their carbon footprints. Such differentiation could incentivise firms to either reduce their GHG emissions or, less preferable from both the investor's and society's perspectives, to make it look as if they are doing so. Of course, any possibility of the latter underscores the need for strict disclosure standards or similar measures to protect investors and stakeholders from deception. In any case, whether the current size of the carbon premium can lead to a meaningful economic impact is a question yet to be investigated.

Our results also shed light on the financial stability implications of the decarbonisation transition. While we can take some comfort in the result that transition risks have been priced into corporate bonds, it remains to be seen if they have been priced in to a sufficient degree.

References

Agrawal, D, E Elton, M Gruber and D Mann (2001): "Explaining the rate spread on corporate bonds", *Journal of Finance*, vol 56, pp 247–77.

Aswani, J, A Raghunandan, and S Rajgopal (forthcoming). "Are carbon emissions associated with stock returns? Reply", *Review of Finance*.

Amato, J and E Remolona (2003): "The credit spread puzzle", *BIS Quarterly Review,* December 2003.

Amihud, Y and H Mendelson (1986): "Asset pricing and the bid-ask spread", *Journal of Financial Economics*, vol 17, pp 223–49.

Bai, J, T Bali and Q Wen (2019): "Common risk factors in the cross-section of corporate bond returns", *Journal of Financial Economics*, vol 131, pp 616–42.

Bai, J, R Golstein and F Yang (2020): "Is the credit spread puzzle a myth?", *Journal of Financial Economics*, vol 137, pp 297–319.

Bharath, S, and T Shumway (2008): "Forecasting default with the Merton distance to default model", *The Review of Financial Studies,* vol 21, no 3, pp 339–69.

Bolton, P, M Despres, L Pereira da Silva, F Samama and R Svartzman (2020): *The green swan: central banking and financial stability in the age of climate change,* Bank for International Settlements.

Bolton, P, and M Kacperczyk (2021a): "Do investors care about carbon risk?", *Journal of Financial Economics*, vol 142, pp 517–49.

——— (2021b): "Global pricing of carbon-transition risk", *NEBR Working Papers*, no 28510.

——— (forthcoming): "Are carbon emissions associated with stock returns? Comment", *Review of Finance*.

Bondioli, M, M Golberg, N Hu, C Li, O Maalaoui and H Stein (2021): *The Bloomberg Corporate Default Risk Model (DRSK) for Public Firms*, Bloomberg Finance LP.

Boss, M and M Scheicher (2002): "The determinants of credit spread changes in the euro area", in *Market functioning and central bank policy*, *BIS Papers*, no 12, August, pp 181–99.

Briere, M and S Ramelli (2021): "Green sentiment, stock returns, and corporate behavior", mimeo.

Busch, T, M Johnson and T Pioch (2018): "Consistency of corporate carbon emission data", *University of Hamburg Report*, WWF Deutschland, Hamburg.

——— (2020): "Corporate carbon performance data: Quo vadis?", *Journal of Industrial Ecology*, vol 26, no 1, pp 350–63.

Campbell, J and G Taksler (2003): "Equity volatility and corporate bond yields", *Journal of Finance,* vol 58, pp 2321–49.

Carbone, S, M Giuzio, S Kapadia, J Krämer, K Nyholm and K Vozian (2021): "The low-carbon transition, climate commitments and firm credit risk", *ECB Working Papers*, no 2631.

Chava, S (2014): "Environmental externalities and cost of capital", *Management Science*, vol 60, pp 2223–47.

Chen, L, D Lesmond and J Wei (2007): "Corporate yield spreads and bond liquidity" *Journal of Finance,* vol 62, pp119–49.

Christopoulos, A (2020): "The Absolute Roll Measure".

Collin-Dufresne, P, R Goldstein and S Martin (2001): "The determinants of credit spread changes", *Journal of Finance*, vol 56, pp 2177–07.

Duffee, G (1998): "The relation between Treasury yields and corporate bond yield spreads", *Journal of Finance,* vol 53, pp 2225–41.

Duffie, D, L Saita and K Wang (2007): "Multi-period corporate default prediction with stochastic covariates", *Journal of Financial Economics*, vol 83, no 3, pp 635–65.

Duan, T, F Li and Q Wen (forthcoming). "Is carbon risk priced in the cross-section of corporate bond returns?", *Journal of Financial and Quantitative Analysis*.

Ehlers, T, K Greiff and F Packer (2022): "The pricing of carbon risk in syndicated loans: which risks are priced and why?", *Journal of Banking & Finance*, vol 136, pp 106–80.

Elsenhuber, U and A Skenderasi (2019): "ESG investing: practical considerations for public investors", in *Evolving Practices in Public Investment Management: Proceedings of the Seventh Public Investors Conference*.

Felhütter, P and S Schaefer (2018): "The myth of the credit spread puzzle", *Review of Financial Studies,* vol 31, pp 351–67.

Ge, W and M Liu (2015): "Corporate social responsibility and the cost of corporate bonds", *Journal of Accounting and Public Policy*, vol 34, pp 597–624.

Giglio, S, B Kelly and J Stroebel (2021): "Climate finance", *Annual Review of Financial Economics*, vol 13, pp 15–36.

Gilchrist, S and E Zakrajšek (2012): "Credit spreads and business cycle fluctuations", *American Economic Review*, vol 102, no 4, pp 1692–720.

Goss, A and G Roberts (2011): "The impact of corporate social responsibility on the cost of bank loans", *Journal of Banking and Finance*, vol 35, pp 1794–810.

Greenhouse Gas Protocol (GGP)(2004): "Corporate Accounting and Reporting Standard, Revised Edition".

Gueant, O (2019): "Measuring liquidity on the corporate bond market", *Scientific Working Papers of the Autorité des Marchés Financiers* (AMF), March.

He, Z and K Milbradt (2014): "Endogenous liquidity and defaultable bonds", *Econometrica*, vol 82, no 4, pp 1443–508.

Heck, S (2021): "Corporate bond yields and returns: a survey", *Financial Markets and Portfolio Management*, Springer; Swiss Society for Financial Market Research, vol 36, no 2, pp 179–201.

Huynh, TD and Y Xia (2021): "Climate change news risk and corporate bond returns", *Journal of Financial and Quantitative Analysis*, vol 56, no 6, pp 1985-2009.

Intergovernmental Panel on Climate Change (IPCC)(2022): "Summary for policymakers", in P Shukla, J Skea, R Slade, A Al Khourdajie, R van Diemen, D McCollum, M Pathak, S Some, P Vyas, R Fradera, M Belkacemi, A Hasija, G Lisboa, S Luz and J Malley (eds), Climate change 2022: mitigation of climate change, Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York.

Merton, R (1974): "On the pricing of corporate debt: the risk structure of interest rates", *Journal of Finance*, vol 29, no 2, pp 449–70.

Painter, M (2020): "An inconvenient cost: The effects of climate change on municipal bonds", *Journal of Financial Economics*, vol 135, pp 468–82.

Polbennikov, S, A Desclee, L Dynkin and A Maitra (2016): "ESG ratings and performance of corporate bonds", *The Journal of Fixed Income,* vol 26, no 1, pp 21–41.

Roll, R (1984): "A simple implicit measure of the effective bid-ask spread in an efficient market", *Journal of Finance*, vol 39, no 4, pp 1127–39.

Rossi, M (2014): "Realized volatility, liquidity and corporate yield spreads", *Quarterly Journal of Finance,* vol 4, no 1.

Seltzer, L, L Starks and Q Zhu (2022): "Climate regulatory risks and corporate bonds", Federal Reserve Bank of New York, *Staff Reports*, no 1014, April.

Appendix 1. Estimating risk-neutral default probabilities

It is possible to estimate default probabilities using Merton's structural model (1974). We start from the assumption that the total value of the firm V (its assets) follows a geometric Brownian motion:

$$dV = \mu_V V dt + \sigma_V V dW$$

where μ_v is the expected return on the value of the firm, σ_V is the volatility of the firm's value and dW is an increment of the standard Weiner process. Next, we must make an assumption about the firm's capital structure. It is assumed that the firm has issued D amount of a single zero-coupon bond of T years maturity.

These assumptions imply that the value of the firm's equity, which we denote E, can be viewed as a call option on the underlying value of the firm V, with a strike price equal to the face value of the firm's debt D and a time to maturity of T. According to the Black-Scholes pricing formula, the value of the firm's equity (the "put option") is given by:

$$E = V\Phi(d_1) - e^{-rT} D\Phi(d_2)$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution function and r the risk-free rate, which is used to continuously discount the value of the debt. Furthermore:

$$d_1 = \frac{ln(V/D) + (r + 0.5\sigma_V^2)T}{\sigma_V^2\sqrt{T}}, d_2 = \delta_1 - \sigma_V\sqrt{T}$$

This way, the value of the firm's equity depends on the total value of the firm and time, which allows us to relate the volatility of the firm's value σ_V to the volatility of its equity σ_E. From Ito's Lemma, and given that under this option pricing framework $\delta E/\delta V = \Phi(d_1)$, we can derive this relationship as:

$$\sigma_E = \left(\frac{V}{E}\right)\Phi(d_1)\sigma_V$$

The inputs to the Merton model are therefore the value of equity, the value of debt and the volatility of equity. Naturally, because a company's debt structure is more complex than the aforementioned zero-coupon bond, we assume that the debt threshold is somewhere between the face value of the short-term debt (D_{ST}) and long-term debt (D_{LT}). Concretely:

$$D = D_{ST} + 0.5D_{LT}$$

In addition, we use a horizon T of five years in total, which matches the longest horizon available for Bloomberg's default probabilities. Firm data for the model is collected from S&P Capital IQ as discussed in section 2.2. For the implementation and, as in Gilchrist and Zakrajšek (2012), we use an iterative procedure proposed by Bharath and Sumway (2008), which addresses large swings in estimated volatility for the firm's value σ_E. For a time series of inputs, the outputs of the model are a time series of σ_V and V. We can then use these to compute firm-specific default probability PD as:

$$PD = \Phi\left(-\frac{ln(V/D) + (\mu_V - 0.5\sigma_V^2)}{\sigma_V}\right)$$

Appendix 2. The absolute Roll measure of illiquidity

The general theory behind the liquidity premium originated with Amihud and Mendelson (1986) and, since at least Chen et al (2007), liquidity has been documented as an important determinant of credit spreads. In their original work, realised bid–ask spreads appear as one of the central measures for gauging bond market illiquidity. However, the absence of *observable* (as opposed to *quoted*) bid-ask spreads for corporate bonds is an issue for the data gathering process (see Gueant (2019) for example).

In its stead, interesting alternatives have been proposed. A theoretically attractive one is the Roll (1984) measure, which allows us to compute a theoretical or *effective* bid-ask, based solely on daily closing price data. In brief, if p_t is the end-of-day price for a bond, we can compute its effective bid-ask spread λ as:

$$\lambda = 2\sqrt{-Cov(\Delta P_t, \Delta P_{t+1})}$$

where $Cov(\Delta P_t, \Delta P_{t+1})$ is the autocovariance of price changes. From this expression, it is easy to see that a complex number is derived when $Cov(\Delta P_t, \Delta P_{t+1}) > 0$. Alternative versions of the measure have been proposed to address the issue – for example, by dropping those observations where the bid-ask spread could potentially be negative. However, these methods lead to gaps in the data. An approach which seeks to preserve the amount of input data available is proposed by Christopoulos (2020) and dubbed the *absolute* Roll measure.

The absolute Roll measure $\hat{\lambda}$ is given by:

$$\hat{\lambda} = 2\sqrt{|-Cov(\Delta P_t, \Delta P_{t+1})|}$$

which leads to a strictly non-negative bid-ask spread applicable to all traded securities that are limited to closing price information. Given the availability of closing price data for our bond sample, we favour the use of this measure in our model of corporate spreads.

To procure monthly data (as our panel regression requires), we follow these steps:

1. For each bond j, we gather all daily closing price data available for month t. Assuming 20 trading days per month, this is a time series of daily prices $\{p^j_{t/20}, p^j_{2t/20}, \ldots, p^j_t\}$.

2. We compute the absolute Roll measure as the autocovariance of this process.

3. We store this computation as the effective bid-ask for month t, $\hat{\lambda}_t$ for each bond j.

4. We repeat this process for the following month, across all bonds.

Appendix 3. Regression analysis with carbon intensities

Our main results are based on analysing a firm's carbon footprint using carbon emissions, which is consistent with Bolton and Kacperczyk (2021a), who link the carbon premium in stock returns to carbon emissions. This choice is also aligned with regulatory frameworks, such as climate stress tests, which tend to focus on activities with high levels of emissions.

To further test the robustness of our findings, we have repeated our core analysis using another common measure of a firm's carbon footprint: carbon emission intensity, which is defined as the ratio of carbon emissions to revenue. This metric has been used in prior studies, such as Ehlers et al (2022) and Duan et al (forthcoming), who investigate the relationships between syndicated loan spreads and corporate bond returns, respectively, and carbon emission intensities.[42]

Tables A3.1 and A3.2 show the results of our analyses using carbon emission intensities in place of carbon emissions. They serve as analogues for Tables 3.4 and 4.2, respectively, in the body of the paper. Our findings suggest that carbon emission intensities are priced into corporate bond spreads through both preference and risk channels. However, we observe that the maturity component (and thereby, the term structure of carbon premia) is statistically significant only in non-energy-intensive sectors.

The preference channel: results with carbon intensities

Table A3.1

	(1) Scope 1 emissions	(2) Scope 1+2 emissions	(3) Scope 1+2+3 emissions
Non-energy-intensive x ln(intensity) x			
Maturity < 5 years	0.143	3.460***	1.209
	[0.820]	[1.236]	[2.306]
Maturity 5-10 years	4.412***	9.274***	7.096***
	[0.808]	[1.208]	[2.286]
Maturity 10-15 years	5.235***	10.57***	9.226***
	[0.913]	[1.252]	[2.308]
Maturity 15-20 years	6.309***	13.11***	12.27***
	[0.922]	[1.279]	[2.311]
Maturity 20-25 years	4.602***	11.36***	11.34***
	[0.911]	[1.301]	[2.318]
Maturity 25-30 years	0.99	7.599***	9.271***
	[0.871]	[1.281]	[2.317]
Maturity > 30 years	0.642	7.430***	9.608***
	[1.031]	[1.366]	[2.369]
Energy-intensive x ln(intensity) x			
Maturity < 5 years	-3.457*	-5.065**	-3.848
	[2.024]	[2.389]	[4.041]
Maturity 5-10 years	0.316	-0.871	0.274

[42] See Bolton and Kacperczyk (forthcoming) and Aswani et al (forthcoming) for a debate regarding which metric better measures a firm’s carbon footprint: carbon emissions or carbon intensities.

	[2.004]	[2.367]	[4.019]
Maturity 10-15 years	1.953	1.225	2.612
	[2.025]	[2.390]	[4.020]
Maturity 15-20 years	3.782*	3.436	5.083
	[2.097]	[2.441]	[4.045]
Maturity 20-25 years	1.315	1.255	3.878
	[1.992]	[2.368]	[4.019]
Maturity 25-30 years	-1.803	-1.545	1.749
	[2.010]	[2.382]	[4.032]
Maturity > 30 years	1.18	1.432	4.578
	[2.397]	[2.671]	[4.132]
Number of bonds	7599	7599	7599
Observations	263,682	263,682	263,682
R-squared	0.84	0.843	0.844

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in brackets; clustered at the security level. Coefficients on other variables omitted for brevity. Sources: Bloomberg; Refinitiv; Trucost; authors' calculations.

The risk channel: results with carbon intensities

Table A3.2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Scope 1 emissions			Scope 1+2 emissions			Scope 1+2+3 emissions		
ln(inten.)	0.07***	0.11***		0.09***	0.12***		0.00	0.14***	
	[0.01]	[0.02]		[0.02]	[0.03]		[0.04]	[0.03]	
Non-energy-intensive x ln(inten.)			0.07***			0.09***			0.14***
			[0.01]			[0.02]			[0.03]
Energy-intensive x ln(inten.)			0.11***			0.12***			0.00
			[0.02]			[0.03]			[0.04]
ln(assets)	-0.14***	-0.15***	-0.14***	-0.14***	-0.15***	-0.14***	-0.14***	-0.14***	-0.14***
	[0.01]	[0.02]	[0.01]	[0.01]	[0.02]	[0.01]	[0.03]	[0.01]	[0.01]
long-term debt/assets	1.56***	1.62***	1.55***	1.54***	1.63***	1.54***	1.84***	1.54***	1.57***
	[0.09]	[0.22]	[0.08]	[0.08]	[0.22]	[0.08]	[0.26]	[0.08]	[0.08]
Earnings/assets	-0.12***	-0.19***	-0.14***	-0.12***	-0.18**	-0.13***	-0.13	-0.13***	-0.13***
	[0.02]	[0.07]	[0.02]	[0.02]	[0.07]	[0.02]	[0.09]	[0.02]	[0.02]
Capital/assets	-0.45***	-0.31	-0.44***	-0.50***	-0.33	-0.49***	-0.38	-0.50***	-0.50***
	[0.05]	[0.28]	[0.04]	[0.05]	[0.28]	[0.05]	[0.25]	[0.05]	[0.04]
Return on assets	-0.05***	-0.03***	-0.04***	-0.05***	-0.03***	-0.04***	-0.04***	-0.05***	-0.04***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Number of firms	2,443	388	2,831	2,443	388	2,831	365	2466	2831
Observations	121,112	19,650	140,762	121,208	19,650	140,858	18,018	122,840	140858
R-squared	0.49	0.55	0.5	0.49	0.55	0.5	0.57	0.49	0.5

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in brackets; clustered at the firm level. Sources: Bloomberg; Refinitiv; Trucost; authors' calculations.

Climate scenarios for fixed income investors

Eric Bouyé, Carmen Herrero Montes and Daniel Vela Barón[1]

Abstract

This paper offers a framework for building capital market assumptions for fixed income portfolios, including expected returns, standard deviations and conditional value-at-risk (CVaR), under diverse climate scenarios. First, it introduces climate risks and reviews the macro-financial models used to design climate scenarios, specifically by showcasing the use of Network for Greening the Financial System and National Institute Global Econometric Model (NGFS NiGEM) macroeconomic data to build financial scenarios for analysing asset classes with interest rates as underlying risk factors. Second, this paper presents a simple method to infer missing data in term structure models, with an application to the Dynamic Nelson-Siegel (DNS) model for both univariate and multivariate cases. Third, it applies the proposed framework to USD currency under six scenarios: Net Zero 2050, Below 2°C, Divergent Net Zero, Delayed Transition, Nationally Determined Contributions, and Current Policies. Expected returns and risk measures are computed for fixed income asset classes (government, corporate and inflation-linked bonds) for three different investment horizons.

Keywords: climate risks, climate conditional value-at-risk, capital market assumptions, asset allocation, scenario analysis, Nelson-Siegel.

JEL classification: G11, D81, C60.

[1] World Bank: 1818 H Street, NW Washington, DC 20433; emails: ebouye@worldbank.org, cherreromontes@worldbank.org, dvelabaron@worldbank.org. The findings, interpretations and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organisations or those of the Executive Directors of the World Bank or the governments they represent.

1. Introduction

Assessing the impact of climate change has become a key consideration for financial institutions, as climate change poses significant risks to asset values and investment returns. At the same time, governments and regulators are implementing new regulations and guidelines to encourage financial institutions to address the risks and opportunities posed by climate change. In 2019, the European Union introduced the Sustainable Finance Disclosure Regulation (SFDR), which requires financial market participants and advisers to disclose information on the integration of sustainability risks into their investment decision-making process.

In March 2022, the Securities and Exchange Commission proposed rule amendments requiring a domestic or foreign registrant to include certain climate-related information in its registration statements and periodic reports. This would include a "[. . .] *proposed definition of scenario analysis [that] both states that (i) when applied to climate-related assessments, scenario analysis is a tool used to consider how, under various possible future climate scenarios, climate related risks may impact a registrant's operations, business strategy, and consolidated financial statements over time; and that (ii) registrants might use scenario analysis to test the resilience of their strategies under future climate scenarios, including scenarios that assume different global temperature increases, such as, for example 3°C, 2°C, and 1.5°C above pre-industrial levels."*

The *Recommendations of the Task Force on Climate-Related Financial Disclosures* (TCFD (2017)) gives five reasons for using scenario analysis: (i) to help organisations consider issues, such as climate change, that are highly uncertain, with medium- to long-term outcomes and potentially disruptive effects; (ii) to enhance strategic conversations about the future and broaden decision-makers' thinking across a range of plausible scenarios; (iii) to help organisations frame and assess the potential range of plausible business, strategic, and financial impacts from climate change and associated management actions; (iv) to help organisations identify indicators for monitoring the external environment and adjust their strategies and financial plans accordingly; and (v) to assist investors in understanding the robustness of organisations' strategies and financial plans and in comparing risks and opportunities across organisations.

As defined in Batten (2018), physical risks arise from the increased vulnerability of current or unadaptable human and natural systems to climate-related hazards, such as gradual global warming and extreme weather events. In contrast, transition risks result from the transition to a low-carbon economy. From an economic perspective, physical risks can lead to unanticipated demand or supply shocks in the short or medium term while significantly affecting productivity and economic growth in the long term. Transition risks are shorter term and depend on the capacity to adapt; they can affect demand, supply and medium-term economic growth.

In recent decades, integrated assessment models (IAMs) have become the standard models for simulating the interactions between the economy and the environment. IAMs were first developed in the 1970s as a response to the growing concern over the environmental impacts of economic growth. They have since evolved to include various economic and environmental variables. Early models focused on the relationship between economic growth and resource depletion, but later models incorporate factors such as population growth, energy use, land use and climate change.

One of the most widely used IAMs is the Dynamic Integrated Climate-Economy (DICE) model developed by William Nordhaus in the 1990s. The model combines a neoclassical economic model with a climate model to simulate the impacts of climate change on economic growth. The model has been used to estimate the optimal carbon tax rate needed to limit global warming to 2°C above pre-industrial levels. Nordhaus developed a regional version, the Regional Integrated Model of Climate and the Economy (RICE), in the 2000s. RICE includes regional differences in both climate impacts and economic development. Since then, a myriad of models has been developed; Weyant (2017) reviews the use of IAMs over the previous 30 years. As pointed out by Pindyck (2013) and Hourcade et al (2021), IAMs present some pitfalls; one main challenge is uncertainty in the underlying data and assumptions.

Determining the impact of physical and transition risks on the performance of different asset classes has become a topic of interest. For example, Bowman et al (2022) study the climate change impact on sovereign bonds. They use forward-looking climate forecasts based on models reviewed by the Intergovernmental Panel for Climate Change (IPCC) and other published sources. Tokat-Acikel et al (2021) explore the implications of climate change for expected returns and strategic asset allocation of public asset classes. The International Actuarial Association (IAA) (2022) suggests using the scenarios for climate-related risk assessment devised by the Network for Greening the Financial System (NGFS)[2] to measure market risk vulnerabilities from both climate change and transitional effects on the economy. This scenario approach can be used with a bottom-up approach that considers the exposure of individual investments to climate-related risks and a top-down approach that estimates the impact of climate scenarios on macroeconomic parameters.

For instance, the European Central Bank (2021) evaluated the increasing credit and market risk for banks by assessing the impact on the probability of default, the loss-given-default, the revaluation of the trading book, and capital shortfalls. The authors use NGFS scenarios to identify that the Disorderly Transition and Hot House World scenarios imply higher loan defaults and asset valuation losses. Furthermore, Allen et al (2020) suggest a bottom-up climate-related risk assessment involving transition risks by examining the impact of increases in carbon prices and productivity shocks from the NGFS scenarios.

In particular, the effects of climate change on fixed income assets can be measured from market, credit and liquidity perspectives, accounting for the macroeconomic impacts of physical and transition risks. The contribution of this paper is twofold. First, it provides a methodology to simulate capital market assumptions for fixed income assets using term structure factors. Second, it proposes a simple imputation method to infer missing data, a problem arising with limited climate scenario data.

This paper's next section presents a detailed exposition of the three overarching classifications employed by NGFS to construct climate scenarios and their impact on key macroeconomic indicators. The third section then expands on the model employed in this paper to derive primary risk factors. The fourth section presents ways to translate the risk factors into capital market assumptions for three distinct fixed income assets, namely US Treasuries, US Treasury Inflation-Protected Securities (TIPS), and BBB-rated corporate

[2] The Network for Greening the Financial System (NGFS), established in 2017, is a group of central banks and supervisors working to support the transition to a sustainable economy.

bonds denominated in US dollars. Finally, the concluding section summarises the main findings.

2. The data and scenarios

Defining relevant scenarios that represent physical and transition risks is pertinent to creating a robust narrative for assessing the many potential paths for fixed income returns. The Network for Greening the Financial System (NGFS)[3] has developed several climate scenarios to help financial institutions evaluate the potential impacts of climate change on their investments and develop strategies for managing climate-related risks. Using two dimensions – the amount of physical risk and amount of transition risk – NGFS establishes three categories: Orderly Transition, Disorderly Transition and Hot House World.

1. The **Orderly Transition** category assumes a smooth, gradual transition to a low-carbon economy with limited financial and economic disruptions. This scenario envisions policy actions that align with the goals of the Paris Agreement, including a significant increase in renewable energy investments and a phased-out use of fossil fuels. Thus, it represents a relatively low exposure to physical and transition risks. Within this category, two main scenarios are identified:
    - <u>Net Zero 2050:</u> This is the most optimistic scenario, with limited climate-related impact on GDP growth. The central assumption is that the world will achieve net-zero carbon emissions by 2050. The global temperature increases to 1.5°C above pre-industrial levels, as outlined in the Paris Agreement.
    - <u>Below 2°C:</u> This scenario assumes that the increase in global temperature is limited to below 2°C above pre-industrial levels without explicitly considering the achievement of net-zero emissions by 2050. This scenario would have additional physical risk but a slightly lower transition risk than the previous scenario.
2. The **Disorderly Transition** category assumes a sudden transition with severe financial and economic disruptions. This represents significant transition risk but low physical risk in the long term. The scenarios in this category envision a situation in which insufficient climate policies lead to abrupt and widespread changes in market conditions. These changes could lead to a sharp repricing of carbon-intensive assets, triggering widespread defaults and financial instability.
    - <u>Divergent Net Zero:</u> Similar to the Net Zero 2050 scenario, this scenario considers a world where net-zero emissions are achieved by 2050. Here, however, the change to a low-carbon economy is uneven and fragmented across regions and sectors. This scenario results in a significant divergence in the pace and nature of the changes, leading to substantial shocks in some macro variables.

[3] The NGFS scenarios can be found at www.ngfs.net/ngfs-scenarios-portal/. This paper uses the third set of NGFS scenarios, published in September 2022.

- Delayed 2°C: This scenario assumes that the world fails to limit the global temperature increase to below 2°C above pre-industrial levels and instead experiences a delay in the transition to a low-carbon economy. This delay might occur for various reasons, such as policy inaction, technological hurdles or political barriers. Contrary to the previous scenario, the Delayed 2°C scenario has higher physical risks, which lead to more significant economic and financial losses, but it presents a slightly lower transition risk.

3. The **Hot House World** category assumes a failure to limit global warming, resulting in severe physical impacts, such as more frequent and intense heat waves, droughts and storms. The scenarios in this category envision a world in which global warming exceeds 4°C above pre-industrial levels, causing significant damage to the environment, infrastructure and human health.

    - Nationally Determined Contributions (NDCs): This scenario considers a world in which countries fully implement their Nationally Determined Contributions (NDCs) under the Paris Agreement.[4] The NDCs submitted by governments, however, are insufficient to limit the global temperature increase to below 2°C or 1.5°C over pre-industrial levels. The physical risk is therefore significantly higher than in the previous scenarios. Still, given that the action to move to a lower-carbon economy is not sufficient, the transition risk is not very significant.
    - Current Policies: This scenario is the most adverse one for the environment, as it assumes that countries continue to implement their current policies and do not adopt any additional climate action measures beyond those currently in place. This scenario significantly increases greenhouse gas emissions, leading to severe physical risks and economic and financial losses. The transition risk is null, while the physical risk is substantial.

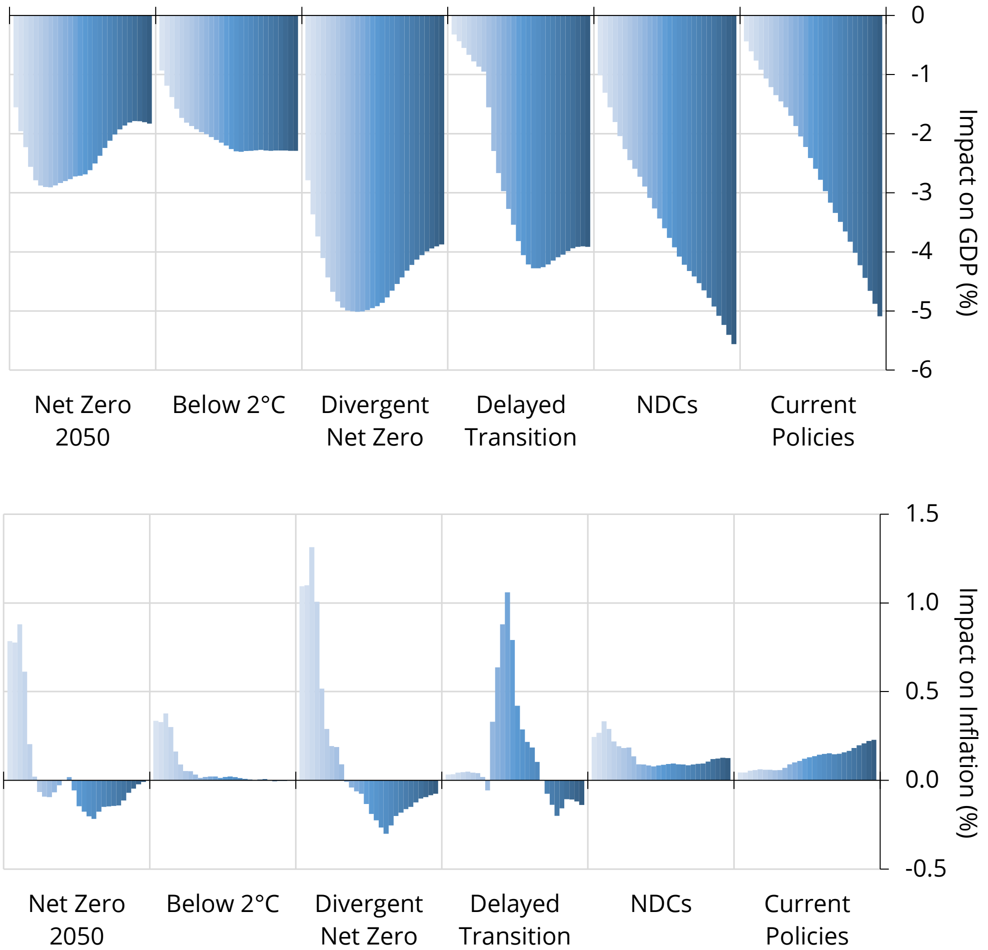
This paper will also focus on the six NGFS scenarios to define their final impact on US fixed income assets in the medium and long term, mainly by stressing the US yield curve. Several macroeconomic factors, such as inflation, GDP growth, unemployment and equity prices, will impact the yield curve. Graph 2.1 displays the evolution of GDP and inflation under different scenarios.

In the Orderly Transition category, inflation remains stable. In the Net Zero 2050 scenario, inflation is even lower in the medium and long term due to investments in energy efficiency, renewable energy and other low-carbon technologies. In the short term, some minor inflationary pressures are created, as transitioning to a low-carbon economy requires upfront investments and changes to production processes. In terms of GDP, the gradual transition to a low-carbon economy makes GDP growth relatively stable. Compared with the other scenarios, in the Net Zero 2050 and the Below 2°C scenarios, the transition to a low-carbon economy drives innovation and productivity gains,

[4] NDCs are pledges made by countries to reduce their greenhouse gas emissions and take other climate action measures. The Paris Agreement (Article 4, paragraph 2) requires each party to prepare, communicate and maintain the successive Nationally Determined Contributions (NDCs) that it intends to achieve.

contributing to GDP growth. Between the two, the Below 2°C scenario performs better in the short and medium term as it is less affected by transition risks. Still, long-term GDP growth is more robust for the Net Zero 2050 scenario as it is less vulnerable to future climate-related hazards.

Impact on GDP and inflation vs baseline scenario (%) — Graph 2.1



Source: The Network for Greening the Financial System.

By contrast, in the Disorderly Transition category, inflation increases significantly in the short and medium term, with the Divergent Net Zero scenario being more severely impacted. This high inflation results from supply chain disruptions and increasing energy prices. Moreover, the higher risk premiums caused by financial instability also push up inflation. In the long term, inflationary pressures ease as the economy adapts to the low-carbon shift – but the same adverse factors GDP cause growth to contract as both production and consumption decrease. There is a slight recovery in the long term, as some physical risks are averted.

Finally, in the Hot House World category, the more frequent and intense extreme weather events disrupt agricultural production and other economic activities in the long term, leading to higher food prices and supply chain disruptions. The Current Policies and the NDCs scenarios will have the highest inflation by 2050. However, in the short and medium term, the inflationary pressures are not as extreme as those in the Disorderly Transition category. For GDP, these two scenarios present a gradual decline in GDP growth due to disruptions of economic activity, destruction of physical infrastructure, and adverse health impacts. GDP growth is significantly lower in the long term for these two scenarios as compared with the other four scenarios, which are much less vulnerable to physical risks.

3. The model

The expected returns and the standard deviations of the fixed income instruments are obtained through three steps.

First, we build the forward-looking term structure of nominal and real interest rates and corporate spreads through time (until the desired investment horizon) for each scenario using the NGFS macroeconomic variables. We then map the newly built variables onto risk factors using a Nelson-Siegel representation. Second, we simulate the risk factors using a constrained vector autoregressive (VAR) model. Third, we obtain the expected returns and standard deviations for US Treasuries, US Treasury Inflation-Protected Securities (TIPS), and BBB corporate bonds.

We initially select the financial variables of interest[5] – short-term and long-term rates – and map them on a Nelson-Siegel functional form.

By using the following Nelson and Siegel (1987) formula,[6] the level β_1 and the slope β_2 can be estimated with the two mentioned variables,

$$y(\tau) \quad = \beta_1 + \beta_2 \left[\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right] + \beta_3 \left[\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right]$$

with τ the maturity and λ a fixed parameter. However, we have three parameters to estimate using two observed variables. This issue can be solved by modelling the distribution of the curvature β_3 conditionally on β_1 and β_2 based on historical data. With z a 3-dimensional vector, β_3 conditional on β_1 and β_2 is defined as follows:

5 The NGFS scenarios include estimates of the central bank intervention rate and the long-term interest rate. However, in the long term, the NGFS intervention and long-term rates converge, leading to a flattening of the yield curve. For this reason, we developed our own estimates of the short-term rate using a Taylor rule model. The Taylor rule, a monetary policy rule first proposed in 1993 by John Taylor to guide central banks in setting short-term interest rates, is based on the state of the economy. The formula is driven by the neutral interest rate (the rate appropriate in the absence of any inflation or output gap), the difference between the current inflation rate and the target inflation rate, and the output gap (the difference between actual and potential GDP). Likewise, the long-term rate is also driven by the potential output as an indicator of economic growth prospects, the inflation rate and inflation volatility to approximate inflation expectations, and the central bank policy rate.

6 Developed by Nelson and Siegel (1987) and further characterised by Diebold and Li (2006).

$$z = \begin{bmatrix} y \\ x \end{bmatrix}, \text{ where } y = \beta_3, \text{ and } x = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

characterised as:

$$\mu = \begin{bmatrix} \mu_y \\ \mu_x \end{bmatrix}, \text{ where } \mu_y = \mu_3, \text{ and } \mu_x = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

and:

$$\Sigma = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix}.$$

The distribution of y conditional on $x = \hat{x}$, is a multivariate normal:

$$(y|x = \hat{x}) \sim N(\bar{\mu}, \bar{\Sigma})$$

where $\hat{x}$ refers to the projected climate scenario betas and:

$$\bar{\mu} = \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(\hat{x} - \mu_x)$$

$$\bar{\Sigma} = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}.$$

The other risk factors, ie credit spreads for corporate bonds and real yields for TIPS, are estimated as follows. The interest rate of corporate bonds is a function of the current government rates, the slope of the yield curve, the equity price index and its volatility. Real yields are obtained by assuming that the most immediate inflation expectations follow the current inflation rate, while in the long term, they are aligned with the Fed's inflation target, with linear interpolation for intermediate values. These other risk factors are then estimated using a Nelson-Siegel model, ensuring a similar modelling framework as nominal interest rates.

Second, we simulate the risk factors using a constrained vector autoregressive (VAR) model, defined as follows:

$$\mathrm{B}_t = A + \Theta \mathrm{B}_{t-1} + e_t$$

with B_t the vector of risk factors: the three Nelson-Siegel factors (level, slope and curvature) for the nominal or real curves, plus the credit spread, e_t a random variable, and A and Θ the constant vector and the matrix of parameters to be estimated, respectively. The risk factors at time t are a function of the risk factors at time $(t-1)$ plus a constant and a random variable. The parameters are obtained through maximum likelihood estimation using 20 years of monthly historical data, assuming e_t follows a multivariate Gaussian distribution.[7] Then, the risk factors are simulated over the investment horizon (Monte Carlo) conditionally to the values of the risk factors under each NGFS scenario.[8]

Once the risk factors are simulated, we obtain the components of asset returns at each point in time, which allows us to build the distributions of asset returns for a given investment horizon. For US Treasuries and corporate bonds, the return is calculated by adding the price and coupon returns. The price return for each period (PR_t) is estimated by pricing a set of N par bonds with different maturities that represent the asset universe.

7 See Diebold and Li (2006) and Diebold and Rudebusch (2013) for a description of the methodology to estimate Dynamic Nelson-Siegel (DNS) models.

8 The simulation of the VAR model is constrained to ensure that the average of the risk factors equals the unconditional expectation at each scenario point in time. One possible criticism of this approach is that the VAR parameters and covariance structure are the same across different scenarios. Another option would be to consider a VAR with switching regimes or a Bayesian VAR for the estimation step.

This pricing (P) is a function[9] of the coupon (c), which corresponds to the respective tenor's yield at the beginning of the period, the maturity (m), and the yield (y), which is derived from the projected Nelson and Siegel factors. The price return is defined as follows:

$$PR_t = \sum_{i=1}^{N} \omega_i \left(\frac{P_i(y_{t,i}, c_{t-1,i}, m_i)}{100} - 1 \right)$$

with ω the percentage weight for each par bond that represents a segment of the fixed income asset class.

Similarly, the coupon return for each period (CR_t) is estimated as the weighted average of the coupon for a specific par bond:

$$CR_t = \sum_{i=1}^{N} \omega_i \left(\frac{c_{t-1}}{\Delta t} \right)$$

with Δt the chosen time period (one year for the empirical results).

The return of inflation-linked bonds is estimated similarly to the US Treasuries, with price and coupon returns, but using the real yield curve instead of the nominal yield curve. Additionally, there is a third component, the realised inflation return (IR_t) to calculate the total return:

$$R_t = (1 + RR_t)(1 + IR_t) - 1$$

with RR_t being the price plus coupon return estimated with real yields.

4. Capital market assumptions

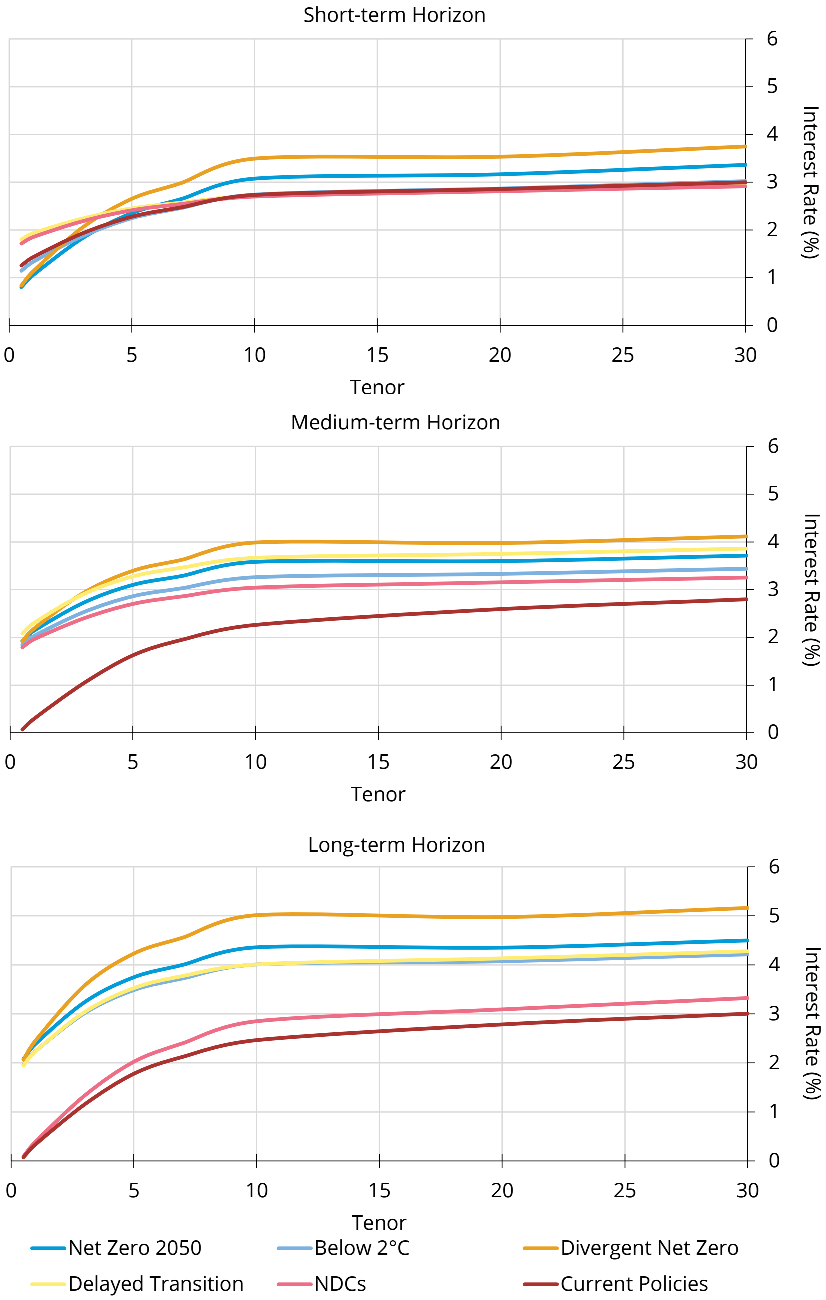
This section provides capital market assumptions based on fixed income risk factors derived from the NGFS climate scenarios. The fixed income risk factors consist of yield curves and credit spreads defined according to the methodology described in the previous section. The capital market assumptions are presented for three indices: US Treasury Bonds, US Treasury Inflation-Protected Securities (TIPS), and BBB corporate bonds.

The US Treasury Bonds index represents the issuance of US Treasuries with an average maturity of 8.3 years. The yield curve is the primary risk factor that will impact the return distribution of this asset class. The US TIPS index represents the issuance of US inflation-linked bonds with an average maturity of 7.8 years. The primary risk factors are the real yield curve and inflation. The BBB corporate bonds index represents corporate bonds issued in US dollars, with a credit rating between BBB– and BBB+ and an average maturity of 10.6 years. In addition to the yield curve, the credit spread is also a relevant risk factor. For the three indices, the asset price's sensitivity to changes in the yield curve, or credit spread if relevant, is assumed to remain constant over time. It should be noted that this analysis considers only market risk and does not take into account the impact of potential defaults.

9 For corporates, the pricing function takes into account the credit spread over the government yield curve.

To capture the impact of transition risk in the short and medium term and to consider the effect of physical risk in the long term, three investment horizons are analysed: a short-term horizon of five years, a medium-term horizon of 14 years, and a long-term horizon of 28 years. Graph 4.1 presents the nominal yield curve by the end of the three horizons for the six NGFS scenarios.

Forecasted nominal yield curves | Graph 4.1



Source: Authors' calculations.

The impact of high inflation expectations in the Net Zero 2050 and Divergent Net Zero scenarios causes the yield curve to steepen more on the short-term horizon than in the other four scenarios. On the medium-term horizon, the effect of falling GDP growth in the yield curve of the Current Policies scenario is notable, resulting in significantly lower interest rates for all tenors than in the other five scenarios. Additionally, the Disorderly Transition scenarios show higher interest rates, given the increasing inflation in these scenarios. On the long-term horizon, the NDCs scenario is similar to the Current Policies scenario, with a low-yield and low GDP environment due to the additional physical risk of the Hot House World category.

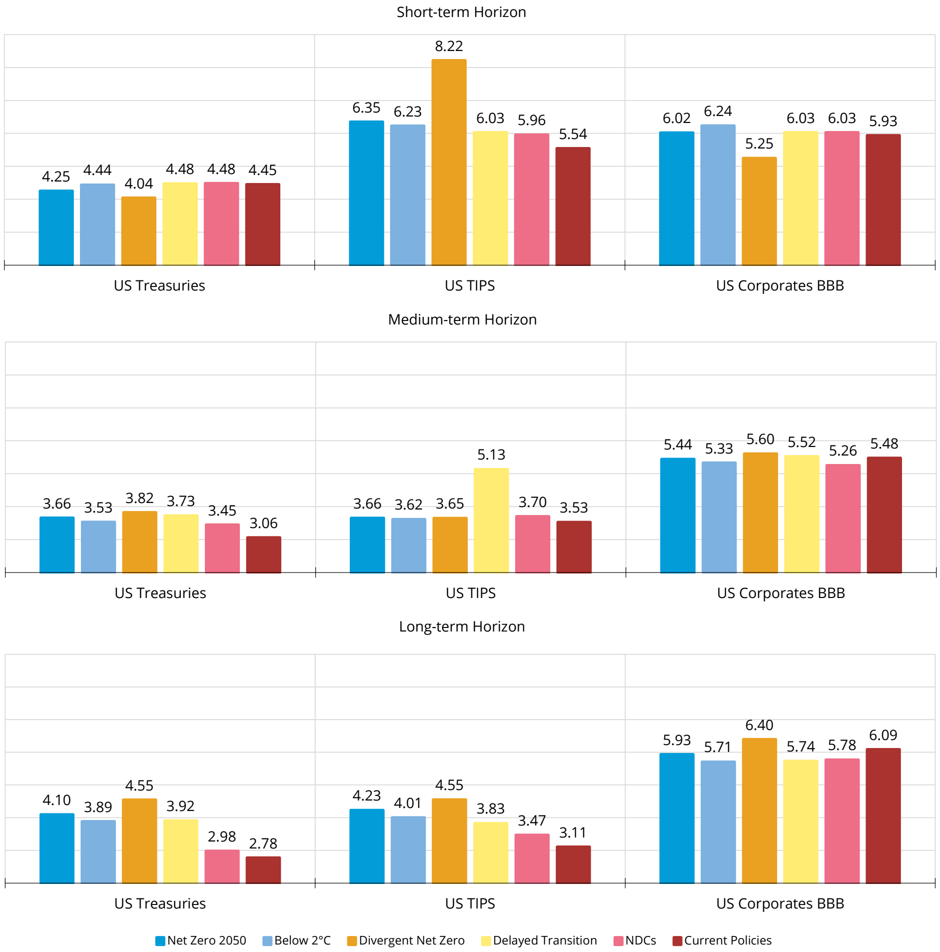
With the specific yield curves for each period, a return distribution is built for the three asset classes with a VAR model adjusting the long-term mean of the risk factors to those expected under the six NGFS scenarios. The yield curves are represented by three parameters, namely the level, the slope and the curvature, following the Nelson and Siegel formula to facilitate the VAR model implementation.

The expected returns for the fixed income asset classes are reported in Graph 4.2. The impact of rising interest rates on the Net Zero 2050 and Divergent Net Zero scenarios, will influence the returns of US Treasuries and BBB corporate bonds in the short term. Nonetheless, in the medium and long term, with rates remaining high and stable, these scenarios demonstrate consistent and elevated returns for the aforementioned asset classes. Under the Disorderly Transition category, high inflationary periods positively impact nominal returns. For the Divergent Net Zero scenario, this occurs in the short term, thereby rendering the returns of TIPS particularly attractive. Conversely, in the Delayed Transition scenario, the high inflation period occurs in the medium term, influencing the return of TIPS. Notably, the Hot House World scenarios exhibit a long-term impact on the assets' returns, where low interest rates weaken the return of US Treasuries and TIPS compared with other scenarios. However, the heightened uncertainty associated with these scenarios increases the credit spreads, making the returns of corporates comparable with those in the other scenarios over the long term. It is pertinent to mention that the additional uncertainty may result in more defaults, a possible outcome that falls outside the purview of this analysis.

Expected returns

Annualised returns (%)

Graph 4.2



Source: Authors' calculations.

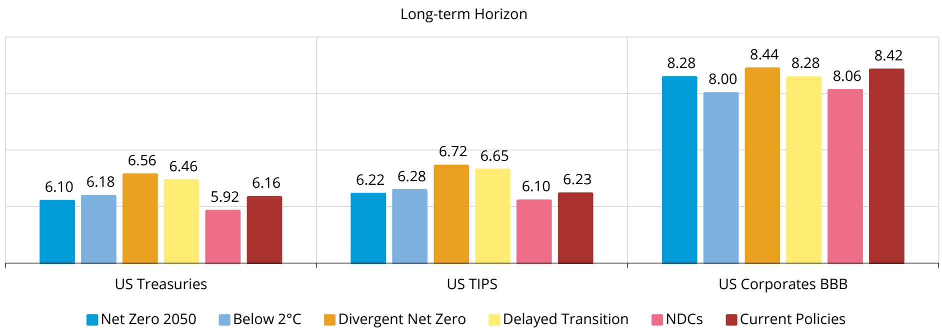
The present study employs a long-term analysis, thereby introducing significant uncertainty into the findings. To illustrate this point, Graph 4.3 showcases the annual volatility of the returns, which proves substantial for all scenarios and asset classes considered.[10] Among the various categories, the Disorderly Transition scenarios exhibit the highest volatility, due to pervasive transition risks that trigger sweeping changes in yield curves and credit spreads across the investment horizon. Conversely, the scenarios categorised under Hot House World display relatively lower volatilities due to the persistently low interest rates, which continue for most of the investment horizons. Nevertheless, the substantially low expected returns associated with low interest rates

10 Figure 4.3 reports the time series volatility. We also computed the cross-sectional volatility for different horizons. We find that the cross-sectional volatility is lower than the time series volatility.

reflect a more negative Climate VaR[11] (with a confidence level of 95%) against the scenarios of the other categories (see Graph 4.4).

Expected volatility (%)

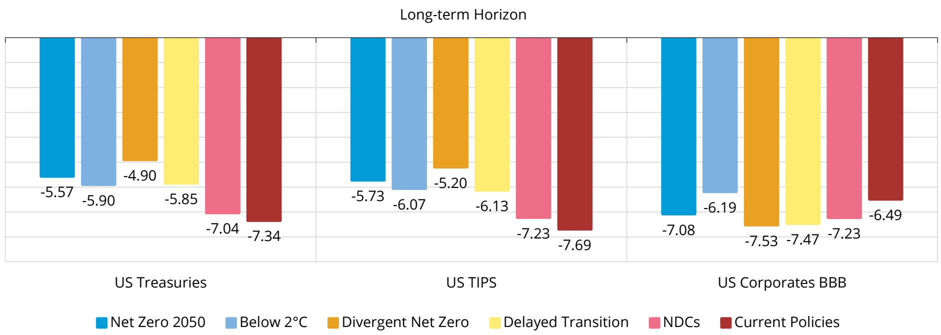
Graph 4.3



Source: Authors' calculations.

Climate VaR (95%) (%)

Graph 4.4



Source: Authors' calculations.

Concluding remarks

The paper offers a framework for building capital market assumptions, including expected returns and standard deviations, for fixed income portfolios under varied climate scenarios. It provides an empirical application for US Treasuries, US TIPS, and US BBB

[11] The Climate VaR represents the worst 5% return outcome for all the simulations performed through the investment horizon.

corporate bonds for three investment horizons (short term, medium term, and long term). A possible extension would consist in building capital market assumptions for an expanded universe of asset classes, eg public equities, infrastructure, real estate, private equity and commodities.

References

Allen, T, S Dees, J Boissinot, C Caicedo Graciano, V Chouard, L Clerc, A de Gaye, A Devulder, S Diot, N Lisack, F Pegoraro, M Rabaté, R Svartzman and L Vernet (2020): "Climate-related scenarios for financial stability assessment: An application to France", *Banque de France Working Papers*, https://doi.org/10.2139/ssrn.3653131.

Antolín-Díaz, J, I Petrella and J F Rubio-Ramírez (2021): "Structural scenario analysis with SVARs," *Journal of Monetary Economics*, vol 117, pp 798–815.

Batten, S (2018): "Climate change and the macro-economy: a critical review", Bank of England, *Staff Working Papers,* no 706.

Bowman, L, D Hu, M Hu, A Madaan and A Baldaque da Silva (2022): "Assessing climate change impact on sovereign bonds", *The Journal of Portfolio Management*, *Novel Risks 2022*, vol 48, no 10, pp 98–118.

Cevik, S and J Tovar Jalles (2020): "This changes everything: climate shocks and sovereign bonds", *IMF Working Papers*, WP/20/79.

Diebold, F and C Li (2006): "Forecasting the term structure of government bond yields", *Journal of Econometrics*, vol 130, pp 337–64.

Diebold, F and G Rudebusch (2013): *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach*, Princeton University Press.

European Central Bank, European Systemic Risk Board (2021): *ESRB climate-related risk and financial stability*, European Systemic Risk Board.

Hourcade J, F Ghersi, R Grandjean, J Lefèvre, P Tankov and S Voisin (2021): "Climate-economic scenarios and models: A reading guide for sustainable finance", *Opinions & Débats*, vol 23.

International Actuarial Association (2022): "Application of climate-related risk scenarios to asset portfolios".

Koberle, A, G Ganguly and A Ostrovnaya (2021): "A guide to building climate-financial scenarios for financial institutions", *Grantham Institute Briefing paper*, vol 35, https://doi.org/10.25561/85497.

Nelson, C and A Siegel (1987): "Parsimonious modeling of yield curves", *Journal of Business*, vol 60, pp 473–89.

Network for Greening the Financial System (2020): "Guide to climate scenario analysis for central banks and supervisors", *NGFS Technical Document*.

Pindyck, R (2013): "Climate change policy: what do the models tell us?" *Journal of Economic Literature*, vol 51, no 3, pp 860–72.

——— (2017): "The use and misuse of models for climate policy". *Review of Environmental Economics and Policy*, vol 11, pp 100–14.

Securities and Exchange Commission (2022): *The Enhancement and Standardization of Climate-Related Disclosures for Investors*, www.sec.gov/rules/proposed/2022/33-11042.pdf.

Task Force on Climate-related Financial Disclosures (2017): *Recommendations of the Task Force on Climate-related Financial Disclosures.*

Tokat-Acikel, Y, M Aiolfi, L Johnson, J Hall and J (Yiwen) Jin (2021): "Top-down portfolio implications of climate change", *The Journal of Portfolio Management*, *Novel Risks 2021*, vol 47, no 9, pp 69–91.

Weyant, J (2017): "Some contributions of integrated assessment models of global climate change", *Review of Environmental Economics and Policy,* vol 11, no 1.

Net zero sovereign bond portfolios

Katharina Schwaiger,[1] Riyadh Ali[2] and Andrew Ang[3]

Abstract

We seek to construct fixed income portfolios for sovereign bonds with climate insights. Our climate methodology for sovereign bonds can be applied as an overlay on any benchmark and tilts towards sovereigns more prepared for the transition to a low-carbon economy and away from those which are less prepared. The tilts seek to reduce sovereign carbon emissions in line with the Paris Agreement. Carbon emissions are represented by historical emissions, and we add forward-looking transition metrics to represent a country's future preparedness. We report findings for both developed markets and emerging markets sovereign portfolios and show that improved climate metrics may be achieved while retaining a similar risk-return profile vis-à-vis the benchmark.

JEL classification: G11, G28, Q54, Q56.

[1] BlackRock, katharina.schwaiger@blackrock.com.

[2] BlackRock, riyadh.ali@blackrock.com.

[3] BlackRock, andrew.ang@blackrock.com.

1. Introduction

Sovereign bonds can be important for many investors as a potential source of safety, liquidity, income and diversification, and as a source of returns. In addition, many investors are mandated to hold government bonds. Despite the central role of sovereign bonds in investor portfolios, there are few studies that integrate environmental, social and governance (ESG) considerations or seek to reduce carbon emissions in this asset class – in contrast to the now large literature integrating ESG and decarbonising portfolios in equities. In this paper, we seek to show how investors can potentially meet climate objectives in sovereign bond portfolios for both developed markets and emerging markets.

We show how to take into account environmental considerations in sovereign allocations, with a particular focus in tilting sovereign portfolios in efforts to reduce carbon emissions. Our framework can be interpreted as tilts to E (the environmental component of ESG) for sovereigns better prepared for the transition to a low-carbon economy. The tilts seek to satisfy the requirements to be a Paris-Aligned Benchmark (PAB), in line with the recommendations set by the European Union Technical Expert Group on Sustainable Finance (EU TEG (2019)) and the guidelines issued by the Institutional Investors Group on Climate Change (IIGCC (2021)) to try to limit average global temperature increases to well below 2°C by the end of the 21st century as specified by the 2015 Paris Agreement. The sovereign climate tilts can be dialled up or down, according to investors' risk preferences, and overlaid on any sovereign benchmark.

As an example of our sovereign climate framework, we assume a sovereign benchmark that obtains a long-run, diversified sovereign exposure. We compute the alphas implied by the benchmark weights and use those in an optimisation, imposing constraints that lower carbon emissions and uplift sovereign environmental (E) characteristics. Thus, we obtain a maximal risk-return portfolio with improved sovereign E criteria. The flexibility of incorporating any sovereign benchmark is important given the large sovereign bond allocations by many investors and the risk-reducing role of sovereign bonds during stress periods (see recently, for example, Jacobsen and Lee (2020); Ren et al (2020)). Additionally, governments have an essential role in reducing global warming by setting frameworks and incentives to reduce carbon emissions. Therefore, the inclusion of E data, such as carbon emissions, in sovereign allocations recognises this impact.

The literature investigating ESG and climate influences on sovereign and corporate bonds is still relatively small. Cevik and Jalles (2020) find that climate change has impacted sovereign bond yields: climate change affects the resilience and vulnerability of economies and government budgets, in excess of traditional determinants of sovereign risk. This makes it more urgent to show how to take into account sustainable information into sovereign bond allocations – especially as potential predictors of excess returns. An older literature has examined how country ratings or political risk is priced in markets, with early papers being Howell and Chaddick (1994) and Erb et al (1996), but these studies concentrate mainly on equity markets. Our results are consistent with Martinelli and Vallee (2021) and Rahman et al (2021) who show that taking into account ESG considerations in sovereign bonds does not detract from returns, but our focus is on meeting the requirements of a Paris-aligned sovereign bond portfolio with climate-related sovereign metrics. Our approach is most similar to Kaul et al (2021) in specifying a tracking error optimisation with constraints incorporating lower carbon emissions and improved climate criteria, but with further analysis, and we also examine climate-aware sovereign EM portfolios. In particular, we highlight the relationship between tracking error, or active risk, relative to exogenous sovereign benchmarks that are needed to implement the climate-aware tilts.[4]

The rest of this paper is organised as follows. In Section 2, we describe climate data on a country level – both current but also forward-looking. In Section 3, we describe a proposed optimisation framework to take into account climate considerations in sovereign bonds, focusing on constructing climate-aware sovereign bond portfolios. We present the empirical hypothetical results in Section 4. The final section concludes.

2. Data

Sovereign debt has been an important part of investor portfolios since the development of international sovereign debt markets in the 1820s (see Flandreau and Flores (2009)), but only recently have sovereign sustainability priorities become important considerations for investors. In this section, we focus on integrating climate (ie Environmental, or E) aspects of sovereigns in an investment strategy. Governments have perhaps the most important role in reducing climate externalities (see, for example, Nordhaus (2021)). Providers of sovereign ESG scores also place a significant weight on E characteristics, such as those published by the World Bank (Gratcheva et al (2020)).

While our framework is relevant for any sovereign bond portfolio, in our empirical work we specify a sovereign benchmark for developed markets and for emerging markets. For developed markets the benchmark consists of equal-weighted 10-year bond futures in Australia, Canada, Germany, Japan, the United Kingdom and the United States. For emerging market economies (EMEs), the benchmark consists of equal-weighted 10-year bond futures in India, Korea, Mexico, Poland, South Africa, Singapore and Thailand. These represent some of the most liquid and frequently traded developed and EME sovereign bonds as of March 2023. China may be another country to include and, despite being the world's largest absolute emitter of greenhouse gases, it ranks well in the cross section using the greenhouse gas emissions per capita metric and as a country it has committed to carbon neutrality by 2060.[5]

We use two data sets that measure a country's emission as well as climate profile: carbon dioxide (CO_2) and other greenhouse gas (GHG) emissions as well as the Germanwatch Climate Change Performance Index (CCPI) for each country. GHG emissions other than CO_2 include methane, nitrous oxide and hydrofluorocarbons and other fluorinated gases (perfluorocarbons and sulphur hexafluoride).[6] The CO_2 data contain emissions related to fossil fuel use and industrial processes such as cement production. Our data on CO_2 and GHG emissions are obtained from MSCI

4 Diversification and asset allocation may not fully protect you from market risk.

5 See for example: China – Climate Performance Ranking 2023 | Climate Change Performance Index (ccpi.org).

6 Non-CO2 emissions defined under the Kyoto protocol are found in Annex A of https://unfccc.int/sites/default/files/resource/docs/cop3/l07a01.pdf.

and are stated in emissions per capita per year in terms of tonnes per capita. We have CO_2 and GHG emissions at the country level from 2019.

The Germanwatch CCPI incorporates forward-looking metrics and policy assessment relevant for a country's alignment with the Paris Agreement. The data set covers 57 countries and the EU; these countries account for more than 90% of global GHG emitters and data are available since 2005.[6] The CCPI assesses each country's performance in four categories: GHG Emissions (40% of the overall ranking), Renewable Energy (20%), Energy Use (20%) and Climate Policy (20%).[7] We take CCPI data from 2017 since the underlying methodology of the CCPI has been revised and adapted to the new climate policy landscape of the Paris Agreement since that date.[8] Countries can achieve a rating of up 100 and can further be classified into a Very High, High, Medium, Low and Very Low rating. Only a Very High rating would suggest that the country is aligned with the Paris agreement, but as of 2023 no country has achieved that rating yet.

Graph 2.1 graphs the latest GHG and CO_2 emissions per capita, and Graph 2.2 shows the latest CCPI profiles for the countries in our investment universe consisting of Australia, Canada, Germany, Japan, the United Kingdom and the United States, as well as India, Korea, Mexico, Poland, South Africa and Thailand as of the last update in January 2023. The emissions data are reported as April 2021. Since $CO_2$is the major component of GHG emissions, both CO_2 and GHG are highly correlated, at around 95% across the developed markets countries in the cross section. Australia has amongst the highest emission intensity of both CO_2 and GHG, at 14.31 and 21.1 tonnes per capita, respectively, followed by Canada and the United States, while the United Kingdom has the lowest within the group, at 4.95 and 6.24 tonnes per capita, respectively. For EMEs, Korea has the highest per capita emissions at 12.13 (CO_2) and 13.74 (GHG), while India has the lowest within the group, at 1.9 and 2.66 tonnes per capita, respectively. Looking at the CCPI scores at 2023 (Graph 2.2), the United Kingdom is ranked the highest by CCPI (with a score of 63.07) in our developed market universe, followed by Germany and Japan. Canada is the worst with a score of 26.47. For our EME universe, India has the highest score at 67.35, followed by Mexico, while Korea has the lowest with 24.91.

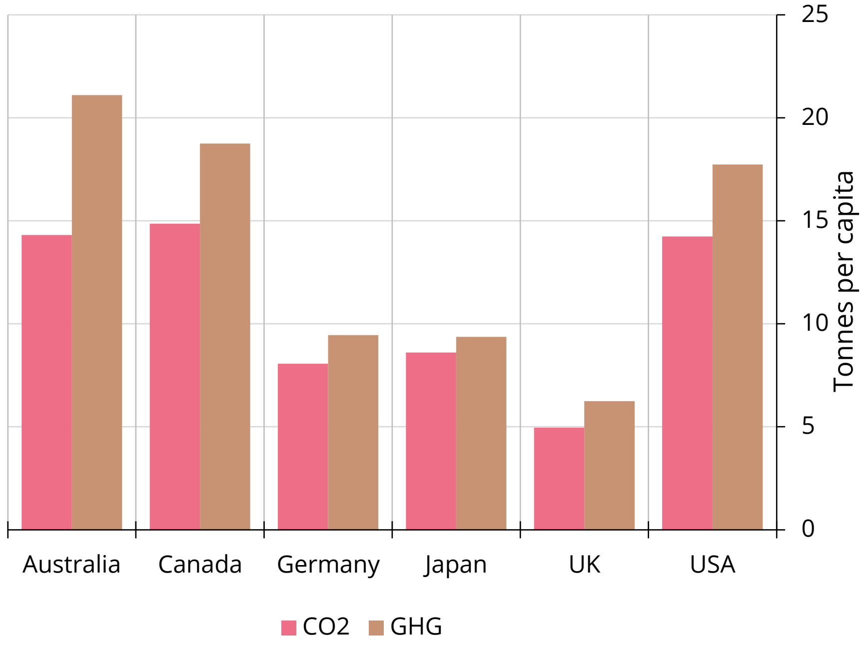
7 The full set of metrics included in the CCPI are described at ccpi.org/methodology.

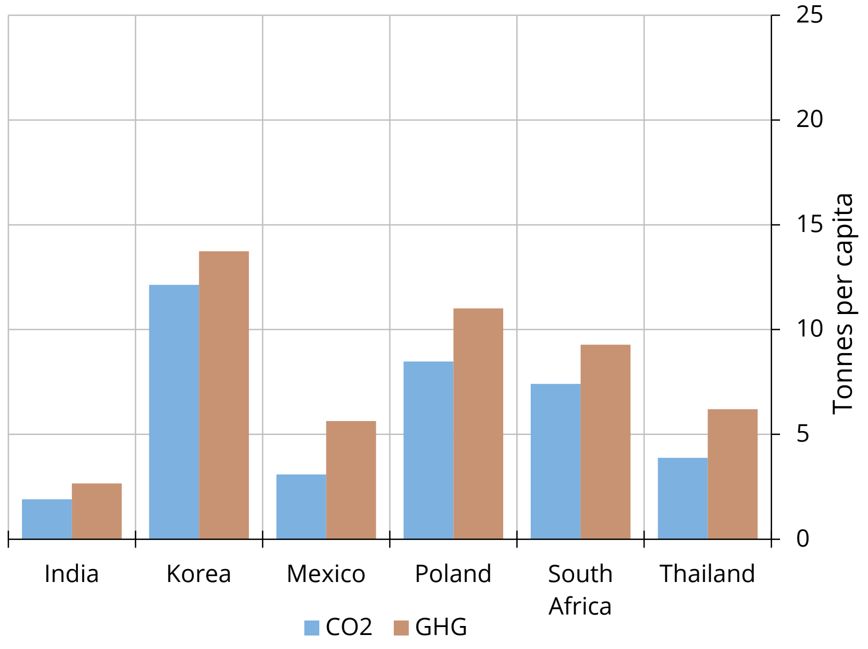
8 The methodology was revised to include emissions from deforestation by CCPI in 2013. Some other sectors, such as agriculture, were not included until 2017 due to data issues. In 2018 the methodology changed to include all GHG emissions (from only energy-related CO2) and Germanwatch began to started to check whether countries set their targets correctly and are fulfilling their promise made in 2015 at the climate conference in Paris.

GHG and CO_2 emissions per capita by country

Graph 2.1

Tonnes per capita





Source: MSCI. Data as of January 2023.

Climate Change Performance Index (CCPI) score for DM and EM countries | Graph 2.2

Score

Source: CCPI. Data as of January 2023.

3. Methodology

Our approach is motivated by the need to limit global temperature warming to well below 2°C by the end of the 21st century, which was the aim of the Paris Agreement adopted by 186 countries in 2015. According to the most recent Sixth Assessment of the Intergovernmental Panel on Climate Change (IPCC (2021)), this goal requires significant and ongoing decreases in greenhouse gas (GHG) emissions to approach the same level of carbon emissions as in 1850–1900 by 2050, which is the net zero transition.[9] In order to help investors align their portfolios with these climate goals, the European Union has adopted a set of standards developed by the Technical Expert Group on Sustainable Finance (EU TEG (2019)). These were adopted into law by the EU in 2020.[10] In addition, we also follow the guidelines issued by the Institutional Investors Group on Climate Change (IIGCC (2021)), which were also formulated to help investors meet the Paris Agreement goals.

We take as given a well diversified sovereign benchmark. Relative to that benchmark, we specify a series of tilts which overweight the countries best prepared for the transition and tilt to those countries with lower carbon emissions. Conversely, we reduce the positioning in countries with poor alignment to the Paris Agreement and countries with higher carbon emissions. This seeks to align with the recommendations outlined by the EU TEG and IIGCC, which specify to "tilt portfolios towards higher performing issuers… to the maximum extent possible, exceeding the average benchmark score".[11] The benchmark is defined by both the EU TEG and IIGCC as the Climate Change Performance Index (CCPI), which is published by Germanwatch.[12]

Since we specify tilts relative to an exogenously specified benchmark, our approach corresponds to setting a sovereign bond climate overlay. This has several potential advantages. First, the benchmark builds in the different motivations for investors to hold sovereign bonds, which include diversification to risky assets like equities (Campbell et al (2020)), liquidity and safety (Brunnermeier (2009)), collateral requirements (Gorton and Laarits (2018)), to seek excess returns in sovereign bonds by harvesting macro factor premia (Ang and Piazzesi (2003); Pauksta et al 2022)), style factor premia (Fama and Bliss (1987); Campbell and Shiller (1981); Ilmanen (2011)), the regulatory treatment of sovereign issues for insurance companies, pensions, and other institutions (BIS (2017)), and other reasons. The benchmark captures the primary reasons for holding a given sovereign bond portfolio and the climate overlay then adjusts those weights in line with those countries most aligned with the Paris Agreement. Second, our methodology can be applied on any sovereign portfolio. Finally, we can dial up or down the tilts of the climate overlay to attempt to trade off

9 See www.ipcc.ch/site/assets/uploads/2021/08/IPCC_WGI-AR6-Press-Release_en.pdf. The only one of the climate shared socioeconomic pathways (SSP) considered by IPCC (2021) that meets the criteria of the Paris Agreement is SSP1-1.9. The recommendations by the EU TEG and IIGCC are specifically intended to help investors create portfolios that seek to attain SSP1-1.9.

10 See COMMISSION DELEGATED REGULATION (EU) of 17.7.2020 supplementing Regulation (EU) 2016/1011 of the European Parliament and of the Council as regards minimum standards for EU Climate Transition Benchmarks and EU Paris-aligned Benchmarks.

11 See https://ec.europa.eu/info/sites/default/files/business_economy_euro/banking_and_finance/documents/190930-sustainable-finance-teg-final-report-climate-benchmarks-and-disclosures_en.pdf.

12 https://germanwatch.org/en/CCPI.

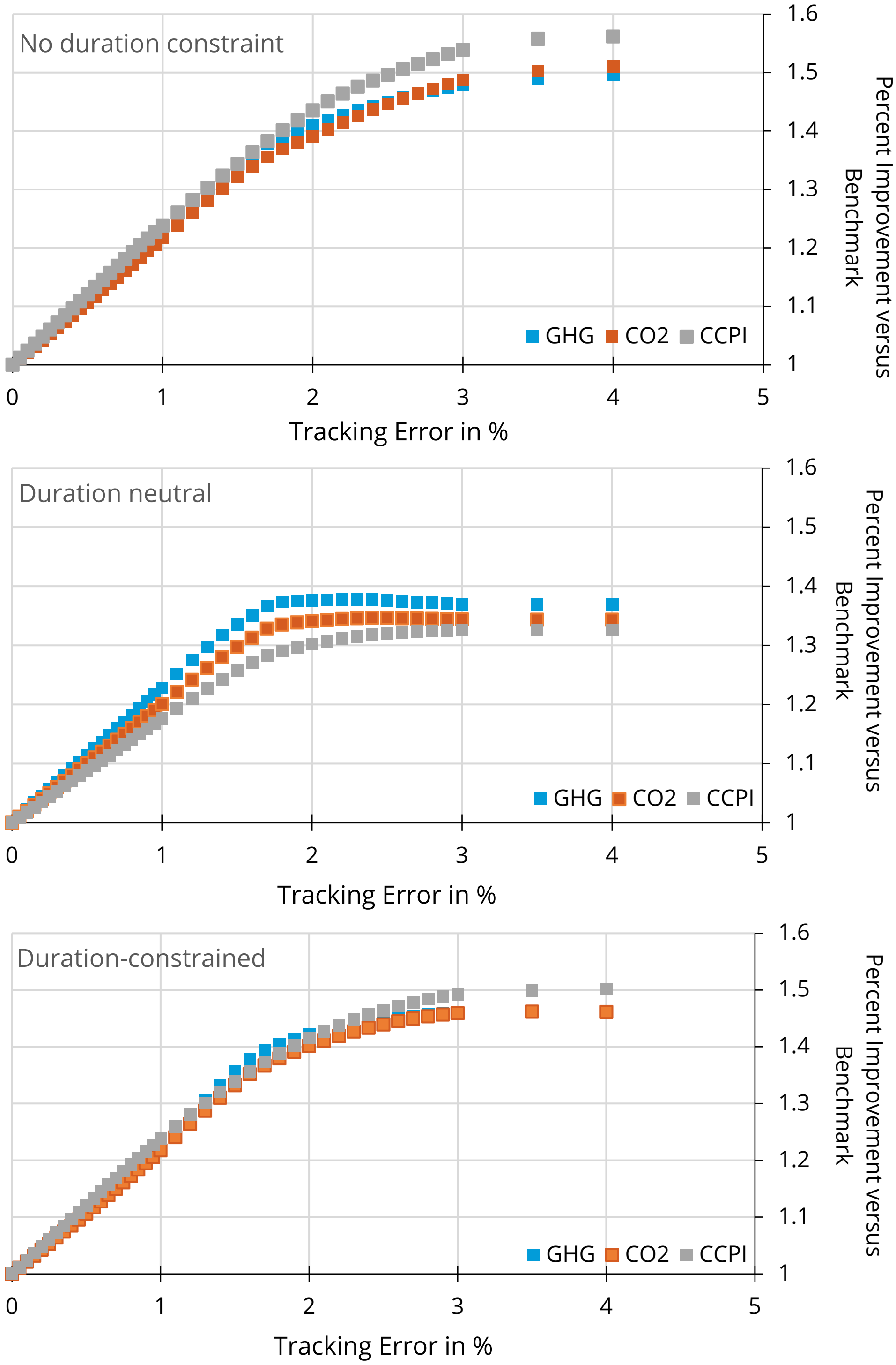
climate vis-à-vis other investment considerations in the investor's sovereign bond allocation.

Our first analysis is maximising the portfolio's CCPI score $A_{ccpi}h$, while fixing the tracking error to the benchmark index with weights h_{bmk} where h are the holdings of the environmental-optimised portfolio and A_{ccpi} is a vector of CCPI climate scores. Given the negative correlation between CPPI and the carbon emission metrics of over 70% (Graphs 2.1 and 2.2), we can observe the resulting portfolio's carbon emissions of both GHG $A_{GHG}h$ and CO2 $A_{CO2}h$, where A_{GHG} and A_{CO2} are the GHG and CO_2 emissions for each country. Relevant for fixed income instruments, we also observe the duration profile and investigate the effects of no duration constraints, duration neutrality (portfolio duration = benchmark duration) and duration bounds of ± 0.5 years. In Graph 3.1 and 3.2, we show the results for the developed market portfolio and the EME portfolio, respectively. We can observe that the higher the tracking error budget, the higher the improvement of the portfolio level environmental characteristics: with a tracking error of 1%, the developed market portfolio has an improvement of around 20% in all environmental characteristics, while for the EME portfolio the improvement is 20% for CPPI and almost double for GHG and CO_2. When doubling the tracking error to 2%, the climate characteristics also improve by almost double, but the effect is not linear and the improvement slows down afterwards. When duration constraints are added, the improvement effect is even further limited. In the case of the EME portfolio, the CCPI improvement is additionally limited by the holdings constraint, which limits the potential allocation into high-scoring countries – and in fact only India has a much higher CCPI score than the average, while in the developed market portfolio both Germany and the United Kingdom are much higher than the average.

Trade-off between tracking error and climate metrics improvement for developed markets portfolio

Graph 3.1

Percent

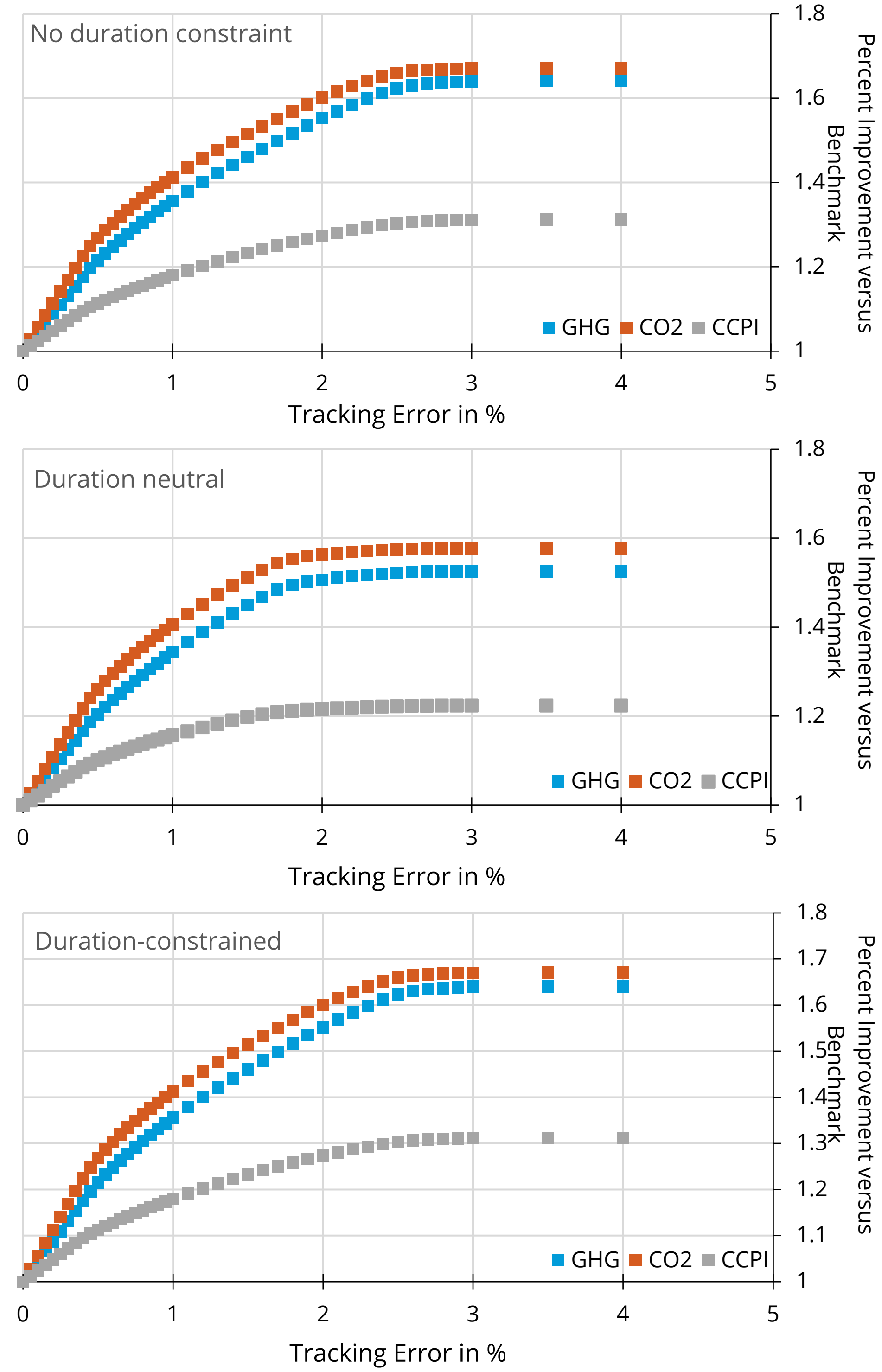


Source: CCPI, MSCI, BlackRock Calculations. Data as of March 2023.

Trade-off between tracking error and climate metrics improvement for EME portfolio

Graph 3.2

Percent



Source: CCPI, MSCI, BlackRock Calculations. Data as of March 2023.

For the second analysis we follow the approach of Kaul et al (2022) and we denote the sovereign weights of the benchmark index as h_{bmk}. We first infer implied alphas, $\alpha_{implied}$, from the benchmark weights. We assume a mean-variance representative agent, so we can write, following Black and Litterman (1991) and others:

$$\alpha_{implied} \propto V h_{bmk}, \tag{1}$$

where V is the covariance matrix of the benchmark sovereign returns.

To incorporate our environmental targets, we specify a new optimisation taking the implied alphas from the benchmark and additional constraints to upweight the CCPI rating and reduce carbon emissions:

$$\max_{h} \alpha_{implied}^{\top} h - \lambda h^{\top} V\, h, \tag{2}$$

such that

$$\begin{aligned} A_{ccpi} h &\geq LB_{ccpi} \times A_{ccpi} h_{bmk} \\ A_{GHG} h &\leq UB_{GHG} \times A_{GHG} h_{bmk} \\ A_{CO2} h &\leq UB_{CO2} \times A_{CO2} h_{bmk} \end{aligned} \tag{3}$$

where h are the holdings of the E-optimised portfolio such that $h - h_{bmk}$ reflects the active Environmental tilts relative to the sovereign benchmark. Equation (2) is a standard mean-variance optimisation with risk aversion λ taking the implied alphas from the sovereign benchmark (see equation (1)).

The climate constraints in equation (3) can be interpreted as follows. First, we seek to target an increase of 10% or more in the CCPI score. Second, we specify a reduction of 14% or more reduction in GHG and CO_2 emissions intensity relative to the benchmark. These increases in the CCPI score and decreases in GHG and CO_2 emissions can be changed for different investors placing more or less importance in the E considerations. We calibrate the risk aversion coefficient, λ, such that without additional constraints (ie using only equation (3)), $h = h_{bmk}$.

In addition, we further specify other investment constraints:

$$\sum h = \sum h_{bmk} \tag{4}$$

$$h \geq 0.05 \times h_{bmk}$$
$$Duration \cdot h_{bmk} - years \leq Duration \cdot h \leq Duration \cdot h_{bmk} + years$$

The first constraint is that the active weights, $h - h_{bmk}$, sum to zero, which reflects the active tilts relative to the benchmark. The second constraint places a lower bound of 5% below benchmark weights. This also ensures that we take no leveraged sovereign weights relative to the benchmark. Finally, we specify the duration exposure to be similar to the current sovereign benchmark at within $\pm$0.5 years of deviation from benchmark duration.

We apply the methodology to a developed market sovereign benchmark of equal-weighted 10-year bond futures in Australia, Canada, Germany, Japan, the United Kingdom and the United States and to an EME sovereign benchmark of equal weighted bond futures in India, Korea, Mexico, Poland, South Africa and Thailand.

4. Empirical results

Graph 4.1 presents the hypothetical results of the optimisation in equations (2)–(4) in this universe. Panel A reports the holdings of the E-optimised portfolio in brown relative to the equal-weighted benchmark in red (each with a weight of 16.7%) as of March 2023. The climate optimisation reduces the positioning in Canada, at 8.1%, because the CCPI score of Canada is significantly low at 26.47 relative to the CCPI of the portfolio of 44.38 (see Graph 2.1). Conversely, the weight of Japan is 24.3%: the large overweight is due to the GHG and CO_2 emissions of Japan, which are low relative to the portfolio. Australia and the United States have high carbon emission intensities, at 21.1 for Australia and 17.73 for the United States for GHG per capita, compared with 13.77 for the benchmark portfolio. This explains the underweight positions to Australia and the United States, and the optimiser consequently overweights the United Kingdom, Germany, and Japan. While the United Kingdom has the best environmental characteristics, it also has the highest duration, which results in a limited overweight position compared with Germany and Japan, which have a duration closer to the benchmark. This has not always been the case, as Panel B shows how the United Kingdom used to be the largest position, followed by Japan and Germany.

In Panel C of Graph 4.1, we report the cumulative hypothetical returns of the climate overlay. The optimisation is run at the daily frequency. The CCPI climate ratings are updated annually in December and we use the same ratings information for the following year. MSCI GHG and CO_2 emissions per capita are available at the monthly frequency, and we forward-fill the data over the next month. We show the raw cumulative returns of the sovereign portfolio and the optimised portfolio with climate tilts. The raw returns of the benchmark and climate portfolio are very close. Over the sample from 2017 to 2023, the climate overlay strategy tracks the benchmark closely. At the beginning of the sample, there is some outperformance, whereas from November 2018 to April 2020 the excess returns detract. Nevertheless, there is a close correspondence. We do not expect the climate overlay to itself have alpha, rather Panel C shows that the climate portfolio tracks the sovereign benchmark closely with an average tracking error of 3%. The benchmark has an IR of –0.09, while the climate portfolio has an IR of –0.1.

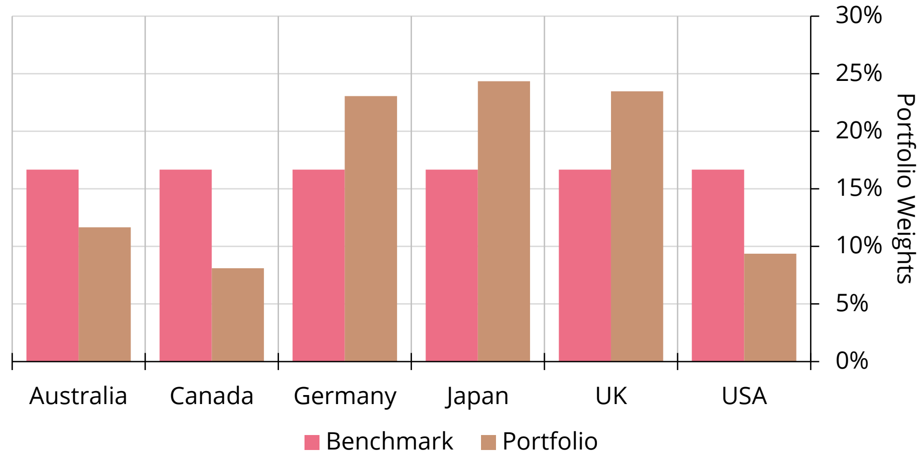
We repeat the analysis for the EM portfolio and Graph 4.2 presents the hypothetical results of the optimisation in equations (2)–(4) in this universe. Panel A reports the holdings of the E-optimised portfolio in brown relative to the equal-weighted benchmark in red (each with a weight of 16.7%) as of March 2023. The climate optimisation reduces the positioning in Korea, at 6.43%, because the CCPI score of Korea is significantly low at 24.91 relative to the CCPI of the portfolio of 45.82 (see Graph 1). Conversely, the weight of India is 27.19%: the large overweight is due to the high CPPI score, and GHG and CO2 emissions of India which are low relative to the portfolio. For the remaining countries the portfolio weight is similar to the benchmark weight as they in particular have similar CCPI scores ranging from 37.94 for Poland to 51.77 for Mexico. Panel B plots the historical holdings: India is mainly the largest position with Thailand being favoured only in 2021.

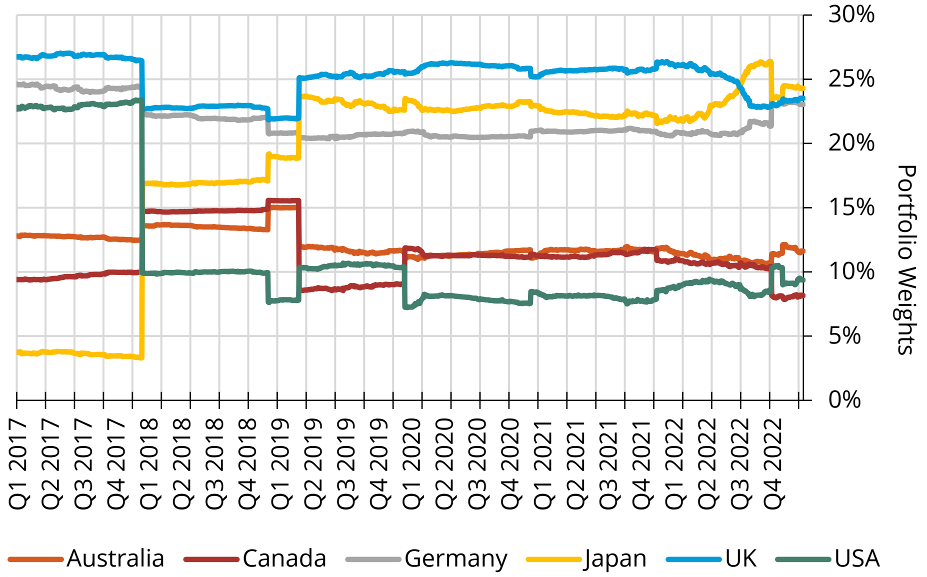
In Panel C of Graph 4.2, we report cumulative hypothetical returns of the climate overlay. We show the raw cumulative returns of the sovereign portfolio and the optimised portfolio with climate. The raw returns of the benchmark and climate portfolio are very close. Similar to the developed market portfolio over the sample from 2017 to 2023, the EME climate overlay strategy tracks the benchmark closely with an average tracking error of 3%. At the beginning of the sample, there is some outperformance, whereas from November 2018 to April 2020 the excess returns detract. Nevertheless, there is a close correspondence. We do not expect the climate overlay to itself have alpha, rather Panel C shows that the climate portfolio tracks the sovereign benchmark closely. The benchmark has an IR of 0.15, while the climate portfolio has an IR of 0.21. Compared with the developed market portfolio, the IRs are higher, and the performance is more driven by larger overweights and underweights – overweight India and underweight Korea.

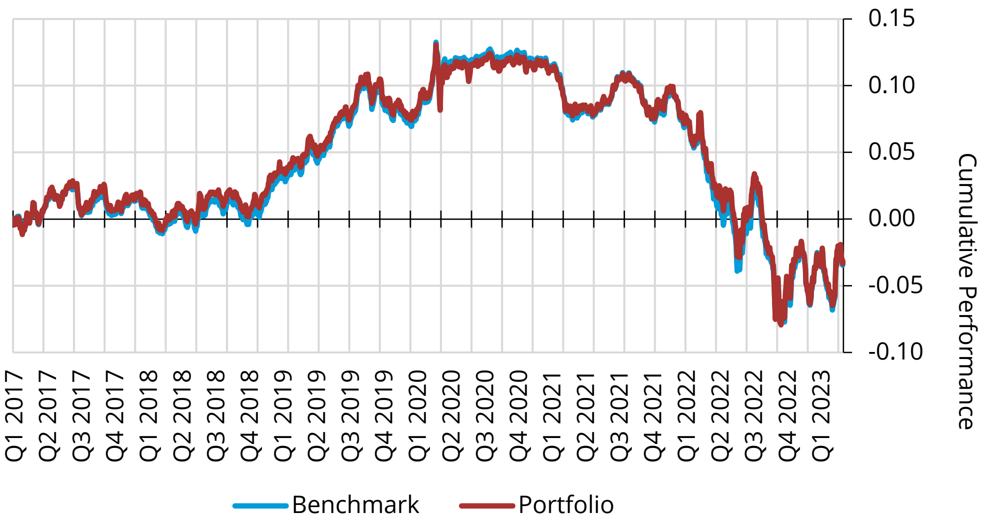
Climate overlay strategy: holdings and cumulative performance for developed markets portfolio

Graph 4.1

In per cent as of March 2023







Source: CCPI, MSCI, BlackRock Calculations. Data as of March 2023.

The hypothetical performance returns are provided for illustrative purposes only and are not meant to be representative of actual performance returns of, or to project or predict returns for, any account, portfolio, strategy or asset allocation. The hypothetical performance period is from January 2017 to March 2023.

*The displayed hypothetical returns are subject to a number of significant limitations. They are illustrative of a product or strategy that does not exist, and therefore do not reflect the deduction of any fees or expenses, including advisory, management and performance fees, as well as brokerage fees, commissions and other expenses that might normally apply. In addition, the allocation decisions reflected in the hypothetical returns were not made under actual market conditions and cannot completely account for the impact of financial risk in actual portfolio management.

The performance shown above is hypothetical and does not represent the investment performance or the actual accounts of any investor(s) or any fund(s). The securities with the hypothetical performance were selected with the full benefit of hindsight, after their performance over the period shown was known. Past hypothetical performance results are not indicative of future returns.

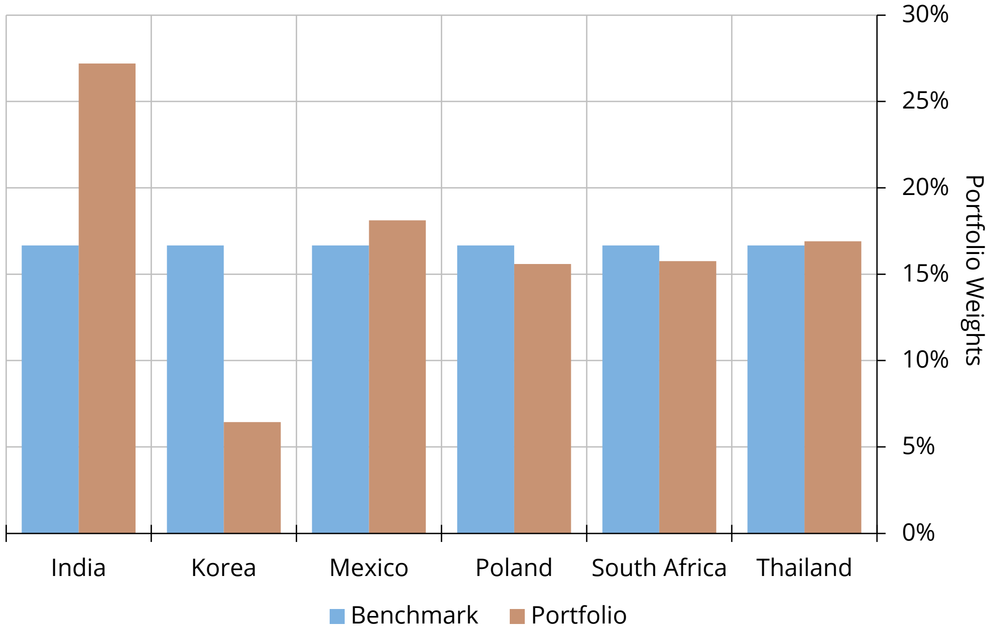
The performance shown does not represent any existing portfolio, and as such, is not an investible product. This represents the model-driven allocations, explained in Section 3 (Methodology), to the underlying well-diversified sovereign benchmark. The underlying performance is based on actual historical performance. The aggregate performance of the model is hypothetical and the model is formulated with benefit of hindsight.

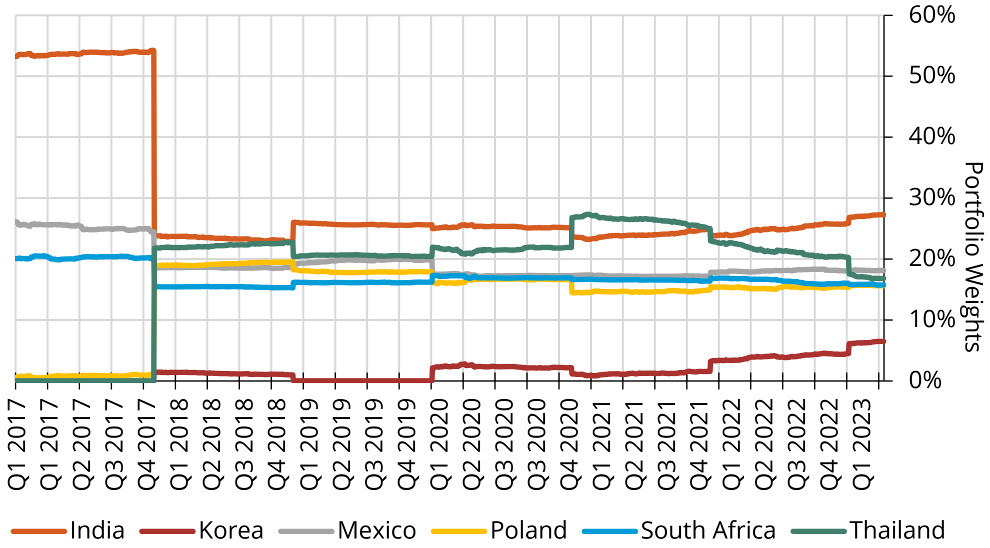
There are frequently sharp differences between a hypothetical performance record and the actual record subsequently achieved. Therefore, hypothetical performance records invariably show positive rates of return. Another inherent limitation of these results is that the allocation decisions reflected in the performance record were not made under actual market conditions and, therefore, cannot completely account for the impact of financial risk in actual portfolio management.

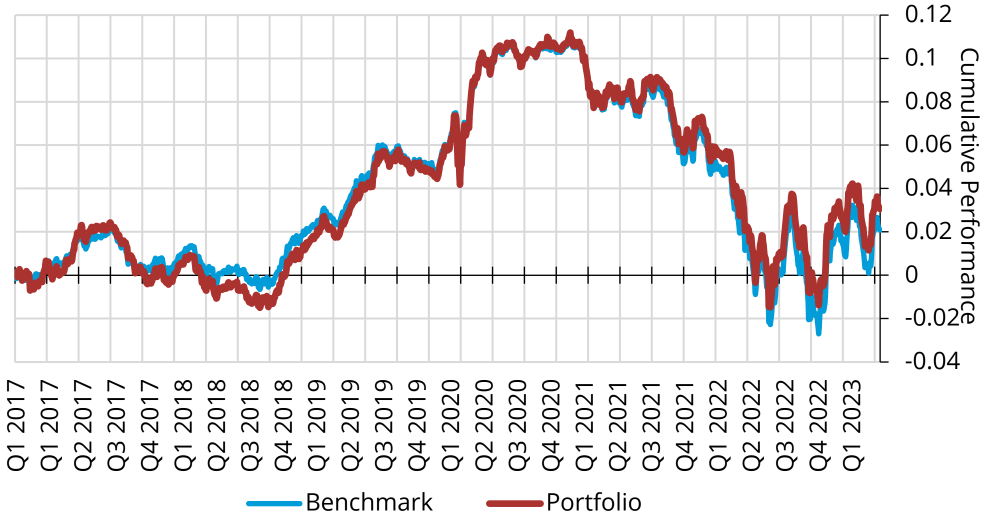
Climate overlay strategy: holdings and cumulative performance for EME portfolio

Graph 4.2

In percent as of March 2023







Source: CCPI, MSCI, BlackRock Calculations. Data as of March 2023.

The hypothetical performance returns are provided for illustrative purposes only and are not meant to be representative of actual performance returns of, or to project or predict returns for, any account, portfolio, strategy or asset allocation. The hypothetical performance period is from January 2017 to March 2023.

*The displayed hypothetical returns are subject to a number of significant limitations. They are illustrative of a product or strategy that does not exist, and therefore do not reflect the deduction of any fees or expenses, including advisory, management and performance fees, as well as brokerage fees, commissions and other expenses that might normally apply. In addition, the allocation decisions reflected in the hypothetical returns were not made under actual market conditions and cannot completely account for the impact of financial risk in actual portfolio management.

The performance shown above is hypothetical and does not represent the investment performance or the actual accounts of any investor(s) or any fund(s). The securities with the hypothetical performance were selected with the full benefit of hindsight, after their performance over the period shown was known. Past hypothetical performance results are not indicative of future returns.

The performance shown does not represent any existing portfolio, and as such, is not an investible product. This represents the model-driven allocations, explained in Section 3 (Methodology), to the underlying well-diversified sovereign benchmark. The underlying performance is based on actual historical performance. The aggregate performance of the model is hypothetical and the model is formulated with benefit of hindsight.

There are frequently sharp differences between a hypothetical performance record and the actual record subsequently achieved. Therefore, hypothetical performance records invariably show positive rates of return. Another inherent limitation of these results is that the allocation decisions reflected in the performance record were not made under actual market conditions and, therefore, cannot completely account for the impact of financial risk in actual portfolio management.

Conclusion

We have sought to show how to incorporate sustainability considerations in sovereign bond portfolios for both developed and EME countries. We show how to incorporate positive tilts for countries that are more prepared for the transition to a low-carbon economy and negative tilts for countries that are less prepared. These tilts use information from the Climate Change Performance Index along with explicit reductions in carbon dioxide and greenhouse gases, which follow the recommendations laid out by the EU's Technical Expert Group on Sustainable Finance (EU TEG) and the Institutional Investors Group on Climate Change (IIGCC) for Paris-

Aligned Benchmarks. We show that with a low tracking error of up to 1%, both developed market and EME portfolios can have an improved climate profile with an increase of around 20% in both backward-looking (CO_2 and GHG) as well as forward-looking (CCPI) metrics – and, in the case of EMEs, an improvement of as much as 40% in the backward-looking metrics. The methodology of the sovereign climate overlay can be applied to any sovereign benchmark. We believe our results have the potential to be important for policymakers, as well as asset owners and asset managers, who may need to hold sovereign bonds for investment or regulatory purposes and would like to apply sustainability considerations in their sovereign bond portfolios.

References

Ang, A and M Piazzesi (2003): "A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables", *Journal of Monetary Economics,* vol 50, pp 745–87.

BIS (2017): *The regulatory treatment of sovereign exposure*, Discussion paper.

Black, F and R Litterman (1991): "Asset allocation: combining investor views with market equilibrium", *Journal of Fixed Income,* vol 1, pp 7–18.

Brunnermeier, M (2009): "Deciphering the liquidity and credit crunch 2007–2008", *Journal of Economic Perspectives*, vol 23, pp 77–100.

Campbell, J, C Pflueger and L Viceira (2020): "Macroeconomic drivers of bond and equity risks", *Journal of Political Economy,* vol 128, pp 3148–85.

Campbell, J and R Shiller (1991): "Yield spreads and interest rate movements: a bird's eye view", *Review of Economic Studies,* vol 58, pp 495–514.

Cevik, S and J Jalles (2020): "This changes everything: climate shocks and sovereign bonds", *IMF Working Papers*, no 2020-079.

Chan, Y, K Hogan, K Schwaiger and A Ang (2020): "ESG in factors", *Journal of Impact and ESG Investing,* vol 1, pp 25–45.

Erb, C, C Harvey and T Viskanta (1996): "Political risk, economic risk, and financial risk", *Financial Analysts Journal,* vol 52, pp 29–46.

EU Technical Expert Group on Sustainable Finance (2019): *TEG Final Report on Climate Benchmarks and Benchmarks' ESG Disclosures*, https://ec.europa.eu/info/sites/info/files/business_economy_euro/banking_and_finance/documents/190930-sustainable-finance-teg-final-report-climate-benchmarks-and-disclosures_en.pdf.

Fama, E and R Bliss (1987): "The information in long-maturity forward rates", *American Economic Review,* vol 77, pp 680–92.

Flandreau, M and J Flores (2009): "Bonds and brands: foundations of sovereign debt markets: 1820–1830", *Journal of Economic History,* vol 69, pp 646–84.

Gorton, G and T Laarits (2018): "Collateral damage", *NBER Working Papers*, no 24298.

Gratcheva, E, T Emery and D Wang (2020): *Demystifying Sovereign ESG*. World Bank Group.

Howell, L and B Chaddick (1994): "Models of political risk for foreign investment and trade: an assessment of three approaches", *Columbia Journal of World Business,* vol 29, pp 70–91.

Ilmanen, A (2011): *Expected Returns: An Investor's Guide to Harvesting Market Rewards*, Wiley & Sons.

IIGCC (2021): *Net Zero Investment Framework: Implementation Guide*, www.parisalignedinvestment.org/media/2021/03/PAII-Net-Zero-Investment-Framework_Implementation-Guide.pdf.

IPCC (2021): *Climate Change 2021: The Physical Science Basis*, Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, in press.

Jacobsen, B, and W Lee (2020): "Risk parity optimality even with negative Sharpe ratio assets", *Journal of Portfolio Management,* vol 47, pp 110–19.

Kaul, K, K Schwaiger, M Si and A Ang (2022): "Sustainable alpha in sovereign and corporate bonds", *Journal of Investment Management*, vol 19, pp 1–22.

Martinelli, L and L-S Vallee (2021): "Measuring and managing ESG risks in sovereign bond portfolios and implications for sovereign debt investing", *Journal of Portfolio Management*, vol 47, pp 198–223.

Nordhaus, W (2021): *The Spirit of Green*, Princeton University Press.

Pauksta, E, K Kaul, T Parker, S Radell and A Ang (2022): "Investing in US core fixed income with macro and style factors," *Journal of Portfolio Management*, vol 38, pp 45–65.

Rahman, L, J Rosten, P Monroy and S Huang (2021): "Does ESG matter for sovereign debt investing?", *Journal of Fixed Income,* vol 31, pp 51–64.

Ren, H, S Siwinski, C Yu and A Ang (2020): "Public pension portfolios in a world of low rates and low risk premiums", *Journal of Investment Consulting*, vol 21, pp 30–44.

Part 3

Quantitative portfolio management

Systematic investment strategies for sovereign fixed income portfolios

Mike McMorrow[1]

Abstract

This paper constructs and analyses active systematic investment strategies for sovereign fixed income investors based on two signals: carry and term premium. Using generic, fitted zero coupon yields across the five SDR yield curves, we identify sets of monthly time-varying portfolio weights based on each of our signals and an equally weighted combination of the two. The baseline portfolio construction setup is tailored to investors with a relatively low risk tolerance, such as foreign exchange reserve managers, by modelling all assets on an FX-hedged basis and requiring neutral duration at the portfolio level. The backtesting exercise is conducted with realised returns from actual securities, accounts for transaction costs, and uses a systematic rebalancing rule. The results suggest that strategies based on the term premium signal can deliver appealing excess returns, while strategies based on carry are more questionable in terms of return, at least over the limited sample used for this paper. Combining the signals, however, may lead to some diversification benefits.

JEL classification: G10, G11, G12.

[1] Bank for International Settlements. The author (mike.mcmorrow@bis.org) is grateful to Ingo Fender for helpful comments on this draft and to various members of the Asset Management and Advisory team at the BIS for many suggestions on earlier iterations of this work. He would also like to recognise the late Vahe Sahakyan for his insights, contributions and warm collaboration. Discussions with participants of the Sixth Public Investors Conference in Ottawa are also gratefully acknowledged. Any remaining errors are the responsibility of the author. The views expressed are those of the author and do not necessarily reflect those of the BIS.

1. Introduction

This paper explores systematic tactical investment strategies for sovereign fixed income portfolios and takes an in-depth look at implementation issues. Systematic investment strategies are part of a rich set of research on empirical factors dating back to the 1980s. This paper contributes to this expansive literature by focusing on sovereign fixed income strategies designed for investors with limited risk tolerance, such as foreign exchange reserve managers, and utilises two signals that are widely followed in the fixed income community: carry and term premium.[2]

The baseline setup reflects an environment in which investors such as reserve managers are likely to operate.[3] Due to the usual reserve management objectives of safety, liquidity and return, such investors may be more comfortable with interest rate risk than with currency risk, and may even be more comfortable with curve positions rather than outright bets on duration. They may also be constrained in their ability to trade certain derivatives for speculative purposes (eg futures), and/or to go outright short a security, ie to sell beyond what is held in the benchmark.

With this environment in mind, we construct portfolio strategies using carry, term premium and an equally weighted combination of the two by finding optimal portfolio weights using a set of generic zero coupon securities across the five SDR currencies. We allow durations from three months to eight years, and constrain the weights to sum to zero, giving rise to a set of fully funded overweight and underweight positions that exploit cross-market and cross-curve dynamics. We design a hypothetical sovereign SDR benchmark to identify constraints on underweight positions at a given yield curve point, given the inability to go short. We model our generic fixed income instruments on a hedged basis, thereby removing currency risk from the strategies, and constrain the optimisation problem such that overall portfolio duration is zero. Our "baseline" setup will thus consist of a set of cross-market, cross-curve positions that incur no net FX or duration risk at the portfolio level.

We build a time series of monthly optimal portfolio weights beginning at the end of September 2016 – which represents the first end-month observation in which the CNY was a constituent of the SDR – through March 2023.[4] We re-optimise on a monthly basis, thereby allowing evolutions in the macro and market environment to be reflected in portfolio weights at a frequency appropriate for tactical asset allocation.

We backtest our strategies with a paper portfolio that uses returns from actual securities and accounts for transaction costs. While futures could be an ideal outlet for implementing such strategies, we avoid any reliance on derivatives given that certain types of investors may be limited in their ability to use such instruments – be it an outright prohibition or a restriction for hedging purposes only. We experiment with different rebalancing rules to account for the trade-off between rebalancing precision and transaction costs, specifically between minimising the duration drift at

[2] Carry is defined as the return on a fixed-income security when the yield remains unchanged. The term premium is the excess yield on a term fixed income security over and above the average short rate expected over the life of the security.

[3] See Fender et al(2022) for a description of a conventional reserve management framework.

[4] See www.imf.org/en/News/Articles/2016/09/29/AM16-NA093016IMF-Adds-Chinese-Renminbi-to-Special-Drawing-Rights-Basket.

the portfolio level due to imperfect rebalancing, and the higher transaction costs that result from frequently switching in and out of securities.

Finally, we experiment with portfolio construction setups that are somewhat less constrained, and may therefore be appropriate for more risk-tolerant investors. Specifically, we explore a version that allows for modestly positive or negative duration positions at the overall portfolio level. In another formulation, we allow FX risk from exposure to CNY, due to the fact that, among other reasons, hedging the CNY gives rise to volatility in portfolio weights over time, thereby generating heavier transaction costs.

We find that the baseline term premium strategy generates encouraging risk-adjusted returns over our sample, while the baseline carry strategy generates slightly negative returns. Combining the two signals can give rise to some diversification benefits, although over the whole sample a combined strategy underperforms one based on the term premium alone. We find that leaving the CNY unhedged generates superior risk-adjusted returns to the fully hedged setup, although with greater maximum drawdown and with minimal diversification between the signals. Finally, we find little evidence that allowing flexible portfolio duration leads to better risk-adjusted performance.

The broader literature on factor-based investing is expansive in the dimensions of both factor types and asset classes. Most of the focus has been on the equity market, where factors such as price-to-book ratios and measures of the equity risk premium have been explored. Exploration of the fixed income market has been a more recent phenomenon, with contributions such as Asness et al (2013) assessing factors such as momentum and value. Value is of particular interest, especially given the lack of clarity of how to define it in a fixed income context. For example, Asness et al (2013) use the five-year change in the yield on 10-year bonds. We find the term premium a more conceptually appealing measure, given that its definition – the premium that can be harvested by holding a term fixed income security in excess of what would be earned through rolling over very short-term assets – contains a forward-looking element. Carry has also been explored extensively in the literature. In addition to Coche et al (2018), Koijen et al (2018), Ahmerkamp and Grant (2013) and Baz et al (2015) all show that carry can be predictive of excess returns in various contexts, including across different asset classes, and may be combined with other signals to exploit diversification benefits.

This paper is closely related to Bjorheim et al (2018) and Coche et al (2018) – both of which were featured at the Sixth Public Investors Conference – and extends their results by offering answers to several portfolio design and implementation questions.

Bjorheim et al (2018) employ a novel macro-based process for modelling yield curves to extract term premium estimates. Specifically, they devise a shadow rate model (see, for example, Lombardi and Zhu (2018)) based on a modified Nelson-Siegel equation to model yield curves at their effective lower bounds – a key feature of monetary policy evolution since the Great Financial Crisis (GFC). Their shadow rate approach allows for the continual use of a modified Taylor Rule, in which macroeconomic variables, such as the output gap and inflation expectations, condition the evolution of the short rate (the first yield curve factor in their approach). This evolution is used to identify term premium estimates, which are subsequently used as signals to predict excess returns across the four currencies that constituted the SDR prior to October 2016. They find that, on an out-of-sample basis, the term premium outperforms expected returns and carry as a predictive signal, whether one assumes perfect foresight of macroeconomic developments or employs a "mean-reverting macro" forecasting approach. Their results showing the superior excess return predictability of the term premium signal are consistent with those found in this study, and their use of a more extended sample – spanning decades rather than years – may help to reassure that the results presented herein are not the spurious product of a limited performance history.

Coche et al (2018) conduct a similar, detailed study of the excess return predictability of a carry signal. Using a history beginning in the 1970s, they backtest strategies from three perspectives: cross-curve, where positions are permitted within a given yield curve; cross-market, where positions at a specific maturity point are taken across yield curves; and cross-curve, cross-market, where flexibility is provided across yield curves and duration points. They find that the cross-market flexibility is key to generating meaningful excess returns. They further find evidence that excess returns may be time-varying and regime-dependent. We make use of the carry signal with these findings in mind.

The remaining sections of this paper are organised as follows:

2. **Identifying signals:** estimating carry and term premium.
3. **Building portfolios**: identifying weights based on the signals.
4. **Evaluating with a paper portfolio:** simulating a realistic setting to measure historical performance.
5. **Assessing performance:** how did the strategies fare historically?
6. **Considering alternative setups:** do changes in the portfolio construction setup lead to meaningfully different results?

We then conclude with a few reflections going forward.

2. Identifying signals

2.1 Fitting yield curves

To identify signals, we used fitted yields obtained from a modified Nelson-Siegel (NS) model, based on Nelson and Siegel (1987), Nyholm (2015) and Bjorheim et al (2018), and used by Coche et al (2018) and Fender et al (2020), among others. The modelling framework relies on shadow yield curve factors and allows us to retain the linkage between macroeconomic variables and yield curve factors at the effective lower bound. Once projected forward, shadow rates are subsequently transformed into projected yield distributions that – due to the design of the shadow rate procedure – are truncated at the effective lower bound as observed historically.

We rely on spliced data histories for several of the SDR yield curves, allowing us to estimate yield curve factors with long data histories. These long histories will help add richness to the forward-looking simulations that will be used to estimate the distributions that we will associate with our signals (explained in more detail in section 3). Specific data sources and their relevant histories are displayed in Table 1.

Data sources for yield curve estimation

Table 1

Yield curve	Source	Period of use
US	Federal Reserve Board (H.15)	03/1953 - 04/1989
	Bloomberg	05/1989 - 02/2023
Germany	Bundesbank	08/1974 - 12/1994
	Bloomberg	01/1995 - 02/2023
UK	Bank of England	01/1970 - 12/1994
	Bloomberg	01/1995 - 02/2023
Japan	Japan Ministry of Finance	09/1974 - 03/1989
	Bloomberg	04/1989 - 02/2023
China	Bloomberg	03/2003 - 02/2023

Source: Author.

2.2 Estimating carry

Carry is calculated as the return an investor receives over holding period k in an unchanged yield curve environment. Formally, it is described as follows, where P_t is the spot price of a sovereign bond at time *t*, $F_{t,t+1}$ is the price at time *t* of a futures contract expiring in period $t+1$, and X_t is the amount of capital that finances the investment into the futures contract.

$$r^C_{t\to t+1} = \frac{P_t - F_{t,t+1}}{X_t} \quad (1)$$

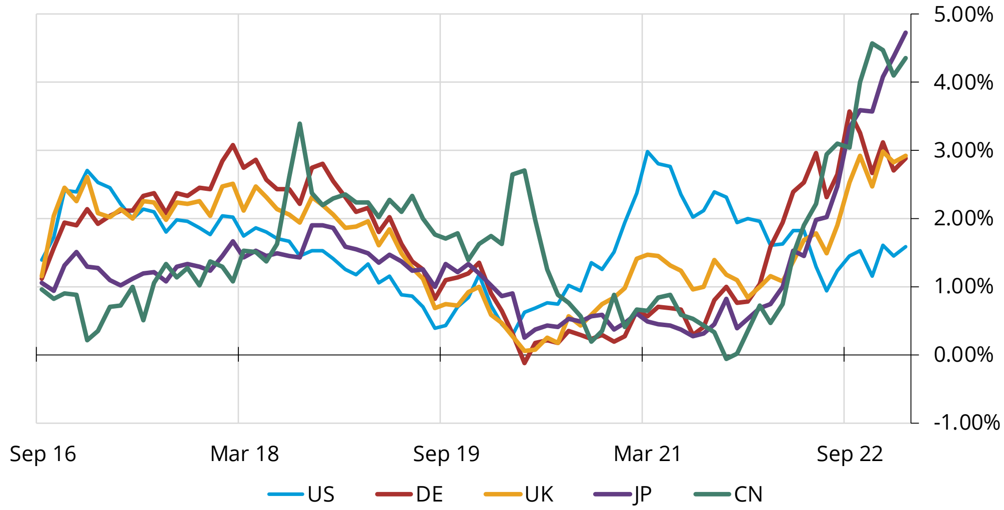
Note that this description is implicitly in local currency terms. Given that our strategies will involve investment into multiple yield curves, we incorporate an FX dimension to account for hedging. Our carry signal, in final form, can be described with $F^f_{t,t+1}$, an FX forward contract on the foreign currency expiring at $t+1$, and S^f_t, the spot FX rate. Both exchange rates are expressed as the domestic currency per unit of foreign currency.

$$CarrySig^j_t = (1 + r^C_{t\to t+1})\left(\frac{F^f_{t,t+1}}{S^f_t}\right) - 1 \quad (2)$$

We apply these formulas to fitted zero coupon yields to construct our metrics of carry. Given that covered interest parity relationships embedded in FX forward pricing will generally offset short rate differentials in local currency terms, our carry signals will essentially capture the *relative slope and curvature* of yield curves, rather than differences in the levels. Nonetheless, any deviations from covered interest rate parity will also be manifest in the signals. Graph 1 displays the evolution of the carry signal on a 10-year government bond for the five SDR yield curves – the United States (“US”), Germany (“DE”, our proxy for the euro area), the United Kingdom (“UK”), Japan (“JP”) and China (“CN”) – at a monthly frequency over our sample.

10-year carry signals across SDR yield curves Graph 1

Percentage



Sources: Bloomberg, author's calculations.

2.2 Estimating term premium

The term premium is the excess yield an investor expects to earn on a fixed income security (with no credit risk) of maturity τ over and above the average short rate (SR) expected to prevail over the life of the instrument.

$$TP_t(\tau) = y_t(\tau) - \frac{1}{\tau}\sum_{i=1}^{\tau} \mathrm{E}(SR_i) \tag{3}$$

Key to measuring the term premium is the estimation of the expected short rate path. With a few extensions, we employ the macro-based approach described in Bjorheim et al (2018), where the first yield curve factor (the shadow short rate) is projected with an autoregressive term along with the output gap and inflation expectations as exogenous variables. Note that the incorporation of the shadow rate approach allows for a coherent estimation with macroeconomic variables at the effective lower bound.

$$\beta_{1,t} = \alpha + \theta\beta_{1,t-1} + \gamma OG_t + \delta\pi_t^e + \varepsilon_t \tag{4}$$

The output gap is calculated as the ratio of GDP to potential GDP, where potential GDP growth is computed recursively as an exponentially smoothed average of the previous period's realised GDP growth rate and the estimated rate of potential growth; the previous period's estimated output gap is also allowed to influence its evolution.

We extend the approach in a few ways. First, we utilise the filtering approach put forth in Stock and Watson (2007) that incorporates unobserved components with stochastic volatility (UCSV) to estimate inflation expectations. Given that central banks are widely believed to respond to persistent movements in inflation – rather than those perceived to be temporary – this unobserved trend component can help strip out unwanted noise.[5]

We further extend the model setup by using Consensus Forecasts from Consensus Economics as assumptions for future GDP growth rates and CPI inflation. These assumptions are fed into the equations for projecting the output gap and inflation expectations, thereby giving rise to the explanatory variables featured in equation (4).[6]

We estimate Taylor Rules individually for each yield curve with maximum likelihood on the available data histories.[7] Given the limited experience with macroeconomic performance in China, especially under the current – and somewhat more flexible – exchange rate regime, as well as the limited availability of seasonally adjusted macroeconomic data, we treat the projection of the CNY short rate differently. Instead of estimation with a Taylor Rule, we simply take the path implied by mean economist forecasts of the short rate as reported by Bloomberg.[8]

A key step in the shadow rate approach of Bjorheim et al (2018) is the transformation of shadow factors back to yields, with yields constrained by their effective lower bounds. The short rate path used in the calculation of the term premium will embody this feature, ensuring that it never falls below the effective lower bound (which is either 0% or the lowest negative yield observed in the historical data for a given yield curve).

Given that we will be using the term premium signal in a multicurrency, hedged context, we incorporate a foreign exchange component into the signal. Given that the term premium represents the yield *in excess* of the expected short rate path, we judge that the most relevant return is the excess hedged FX return, ie the return harvested when hedging a foreign currency investment with FX forwards above and beyond that expected by covered interest rate parity. Formally, this can be described with $F_{t,t+k}^{f}$, a forward contract on the foreign (ie non-numeraire) currency of expiry $t+k$, S_t^f, the spot FX rate, $y_t^d(k)$, an annualised short-term rate with maturity k on the domestic yield curve, and $y_t^f(k)$, an annualised short-term rate with maturity k on the foreign yield curve. The annualised hedged excess return from the perspective of the domestic currency can be described as follows.

5 Rather than conducting the filtering exercise on each simulation, we take a more computationally efficient approach by first conducting the UCSV procedure on the historical data sample, and then regressing the stochastic trend component on realised inflation. The parameters from this regression are then used to forecast inflation expectations using Consensus Forecasts for headline inflation.

6 As in Bjorheim et al (2018), we transform quarterly GDP growth into a monthly figure by applying the pattern of monthly growth in industrial production to quarterly GDP, while ensuring the average monthly GDP growth rate equals the originally reported quarterly GDP growth rate.

7 When doing so, we also periodically adjust the intercept term to force the long-term equilibrium rate implied by the model to be consistent with long-term expectations for the short rate as reported by Consensus Forecasts.

8 Bloomberg consensus forecasts are available for SHIBOR, the interbank rate in Shanghai. Given that we are interested in a projection of the three-month government bond yield, forecasts are adjusted with the most recently observed difference between SHIBOR and the fitted three-month government bond yield.

$$r^d_{hFX,t} = \frac{1}{k}\left(\frac{F^f_{t,k}}{S^f_t} - 1\right) - \left(\frac{1 + y^d_t(k)}{1 + y^f_t(k)} - 1\right) \tag{5}$$

We note that this expression does not precisely yield the FX basis as conventionally defined, given that the interest rate differential is calculated using government yields rather than LIBOR rates (off of which FX forwards have traditionally been priced). As a result, this will be an imperfect hedge in that it results in exposure to Libor-government spreads.

We project this excess hedged FX return over the life of the relevant instrument with a simple AR(1) process.

$$r^d_{hFX,t} = \varphi + \psi r^d_{hFX,t-1} + \varepsilon_t \tag{6}$$

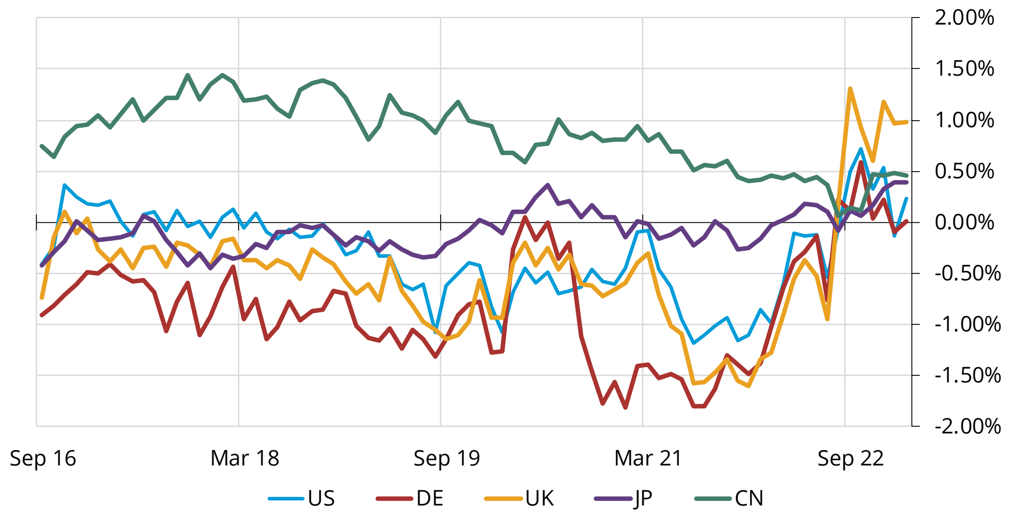
With this FX component, our term premium signal, $TPSig^j_t(\tau)$ at time t for country j and maturity τ is expressed as follows:

$$TPSig^j_t(\tau) = y_t(\tau) - \frac{1}{\tau}\sum_{i=1}^{\tau} \mathrm{E}(SR_i) + \frac{1}{\tau}\sum_{i=1}^{\tau} \mathrm{E}(r^d_{hFX,i}) \tag{7}$$

With this enhancement, we have greater confidence that any deviations from covered interest rate parity (CIRP) are captured by our signal, thereby avoiding – in expectation – bias from the hedging activity. Graph 2 shows the evolution of the 10-year term premium signal in each of the five SDR yield curves over the sample.[9]

10-year term premium signals across SDR yield curves | Graph 2

Percentage



Sources: Bloomberg, author's calculations.

[9] The 10-year term premium in the US is positively correlated with other estimates, including those put forth by Kim and Wright (2005) and Adrian et al (2013).

3. Building portfolios

3.1 Designing a strategy

Equipped with our signals, we design portfolio strategies tailored to our hypothetical reserve manager with a low risk appetite. Given that signals, in this application, are intended to support active management in a tactical asset allocation framework, we construct our portfolios as a set of overweight and underweight positions versus a hypothetical benchmark that sum to zero by construction. The strategies are reoptimised on a monthly basis to ensure that market movements and changes in the outlook are routinely captured.

In defining our universe of eligible assets, we consider five generic securities representing yield curve points across each of the five SDR currencies – specifically, the three-month, two-year, four-year, six-year and eight-year zero coupon yields. For the JPY curve, we restrict the eligible universe to the three-month point in the light of the Bank of Japan's policy of yield curve control over the sample. For EUR-denominated assets, we restrict our interest to the German yield curve, given that the strategy is focused on exploiting signals relating to interest rate risk, not sovereign credit risk. In total, this results in an eligible universe of 21 assets. Each generic instrument is permitted a maximum weight of 50%.

We assume our hypothetical reserve manager holds the five SDR currencies in proportion to their SDR weights.[10] Furthermore, we presume they manage each currency portfolio to a benchmark as determined by the maturity distributions in relevant sovereign bond indices from ICE out to five years – generally equating to currency durations between two and three years (JPY excepted). We further assume that our reserve manager cannot go outright short any sector, and is thus constrained by the weights implied in this benchmark. Table 2 shows the resulting lower-bound constraints that will be applied to our portfolio optimisation problem as implied by the SDR weights and index compositions at the start of our sample (end-September 2016). Given that we expect our reserve manager to hold each currency portfolio for policy purposes (with a small budget for deviation for active management), we presume they select the SDR as the numeraire, thereby viewing currency risk as anything that deviates from these weights.

We design a "baseline" setup that reflects this risk aversion: generally greater comfort with interest rate risk than currency risk, and generally greater comfort with curve risk than with outright duration exposure. We thus require all currency risk to be hedged back to an SDR-neutral position, and that the net duration across active positions is zero.

10 SDR weights are approximately 41% (USD), 32% (EUR), 8% (GBP), 8% (JPY) and 11% (CNY) and are held constant over the sample for simplicity.

Upper- and lower-bound individual asset constraints Table 2
In per cent

	US		DE		UK		JP		CN	
	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
0.25Y	–10%	50%	–7%	50%	–1%	50%	–8%	50%	–2%	50%
2Y	–19%	50%	–14%	50%	–3%	50%	0%	0%	–5%	50%
4Y	–13%	50%	–11%	50%	–4%	50%	0%	0%	–4%	50%
6Y	0%	50%	0%	50%	0%	50%	0%	0%	0%	50%
8Y	0%	50%	0%	50%	0%	50%	0%	0%	0%	50%
Total	–41%		–32%		–8%		–8%		–11%	

Sources: IMF, ICE, author's calculations.

For a given period, we identify separate sets of portfolio weights for each signal by conducting optimisation exercises in the mean-variance space. In the most conventional mean-variance application (Markowitz (1952)), the practitioner would estimate the expected return and portfolio volatility from the observed, historical multivariable return distribution, and minimise volatility subject to a given return objective (among other possible constraints). In our application, the optimisation problem takes a modified set of inputs, in that we minimise portfolio volatility while imposing that the first moment is equal to our signal of interest, $CarrySig_t^j$ or $TPSig_t^j$. For our estimate of the covariance matrix, we do not use the historical return distribution per se; instead, we simulate 1,000 prospective yield curve paths over a horizon of 12-months, and compute monthly returns across simulations for each asset in our universe.[11] We then aggregate monthly returns into a single annual figure for each simulation/asset combination. This leaves us with a 1,000 by 21 matrix of returns from which our covariance matrix is estimated.

We identify optimal portfolio weights for three strategies: carry, term premium and a combination of the two, where the "combined strategy" takes the simple average of the carry and term premium strategy weights. As has been observed in the literature, combining different signals may help enhance risk-adjusted returns with diversification effects. One might also argue that it is another way to reduce exposure to the model risks described above.

We identify portfolio weights for each of the three portfolios at a monthly frequency from a sample beginning end-September 2016 and ending end-February 2023. We are thus left with a 78 by 21 matrix of weights for each of the three portfolios, reflecting our sample size and asset class universe, respectively.

3.2 Historical portfolio weights

Let us now examine the results of these signals-based optimisation exercises. We summarise them using two perspectives, each broken down by the individual currency components: net positions and duration contributions. Net positions reflect the sum of weights within a given currency, while duration contributions reflect the

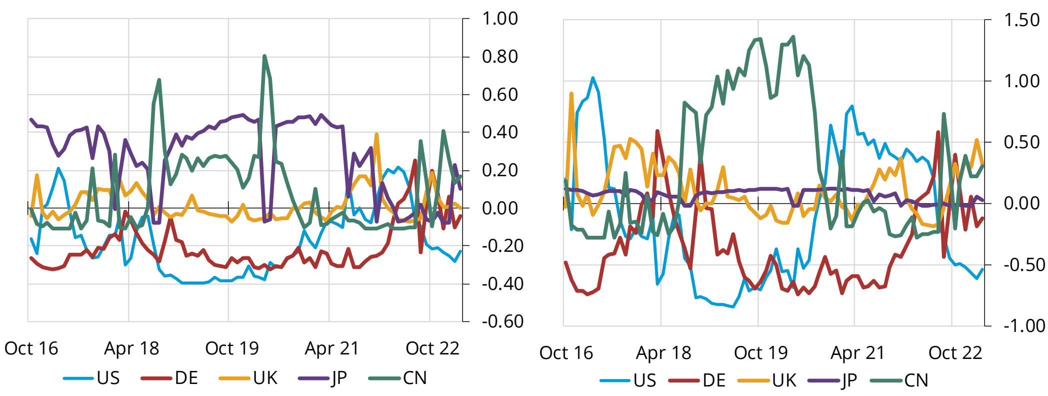
[11] More specifically, time series model errors are drawn at random from the historical distribution of model residuals and applied to the mean projected path.

sum of duration contributions within a given currency. Net positions and duration contributions each sum to zero across currency components, by construction.

Graph 5 shows the resulting portfolio weights for the "carry strategy". A few themes emerge. Over most of the sample's history, long positions in the short-dated JPY security offset a persistently net short EUR position – reflecting in part the well documented attractiveness of the JPY FX basis. A net long CNY duration position tended to offset a net short USD position from roughly mid-2018 to early 2020. There are also at least two instances of sharp reallocations between JPY and CNY assets that are quickly reversed, which upon further examination reflect volatility in the excess hedged FX return for CNY. We do not observe similar volatility in duration contributions, however, reflecting that this "switching" occurs between short-dated assets.

Net positions (lhs) and duration contributions (rhs) of carry strategy — Graph 3

Values (lhs); years(rhs)



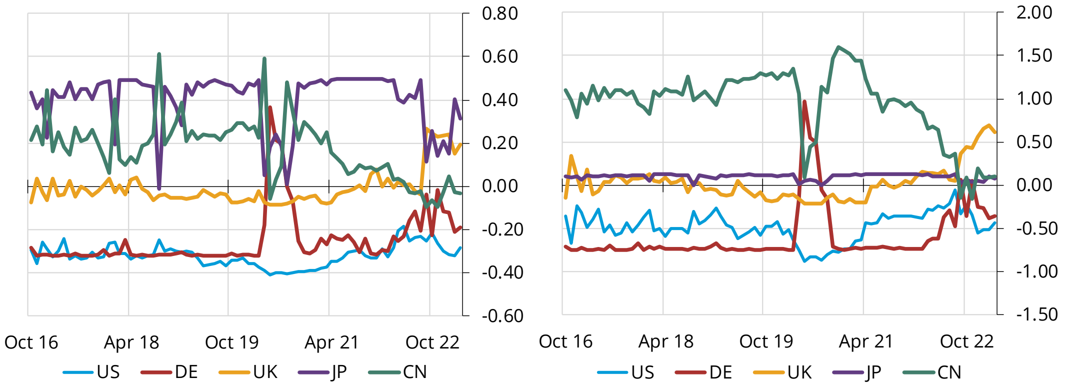
Sources: Bloomberg, author's calculations.

Graph 4 shows the resulting portfolio weights for the "term premium strategy". Themes are not wholly dissimilar from the carry portfolio, though differences are evident. The long positions in CNY and short-dated JPY are again clear themes – though more consistently across the sample – and are funded by persistent short positions in USD and EUR. The contribution from GBP is fairly neutral. We see significant volatility around the onset of the Covid-19 pandemic, when the term premium in EUR rose quickly before falling back to previous levels – a fact that may reflect a lag between changes in the economic outlook as reflected by asset prices and updates to economist forecasts that help drive term premium estimation. We again see some position switching between short-dated CNY and JPY assets, reflecting the same dynamic witnessed in the carry portfolio.

Net positions (lhs) and duration contributions (rhs) of term premium strategy

Graph 4

Values (lhs); years(rhs)



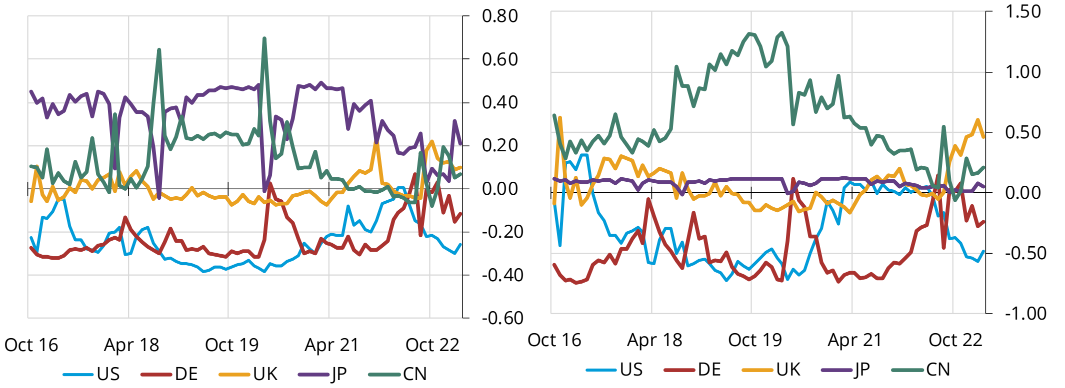
Sources: Bloomberg, author's calculations.

Graph 5 shows the results for the portfolio that takes the simple average of the carry and term premium strategy weights (the "combined strategy"), reflecting a blend of the themes discussed above. Importantly, we are still left with the observation that the weights can be rather volatile.

Net positions (lhs) and duration contributions (rhs) of combined strategy

Values (lhs); years(rhs)

Graph 5



Sources: Bloomberg, author's calculations.

4. Evaluating with a paper portfolio

Our portfolio optimisation exercises make, on an ex ante basis, a number of simplifying assumptions. For example, the use of generic assets implies that each holding is a zero coupon bond with a precise duration. Generic assets are also priced from fitted yield curves, which come with fitting errors. Furthermore, our procedure of monthly re-optimisation implies that portfolio weights will be tweaked on a monthly basis; furthermore, absent any changes in monthly weights, asset and portfolio return projections implicitly assume that the portfolio will be rebalanced back to its assigned weights at a monthly frequency. Finally, the portfolio optimisation problem incorporates no assumed transaction costs.

We seek to build a more realistic setting to perform our backtesting exercise with the use of a paper portfolio represented by actual securities. Given the issues reviewed above, we also need to address the various decisions required to apply model-theoretic weights to a more realistic setting. This includes accounting for transaction costs in some form, and designing a rule for rebalancing a portfolio of actual securities.

To build our paper portfolio, we assign each duration point an actual security holding for the start of our performance period, 1 October 2016, by finding the outstanding government bond with the nearest modified duration to our generic holding (whether the difference is positive or negative). Given that it is unlikely that we will find exact matches for many of our holdings, this will give rise to some error at individual points, though at the overall portfolio level, any drift is unlikely to be significant.

We apply the same approach to each monthly reoptimisation period. However, to economise on transaction costs, we allow positions to roll down one year beyond their generic duration assignment. For example, the US Treasury security corresponding to the four-year US generic asset is permitted to remain in the portfolio until it has a residual modified duration of three or more years. As this rule is applied across securities, it is unlikely to result in systematic duration drifts over time. For three-month assets, we simply allow them to mature. We also assume that JPY holdings are left in a cash account receiving 0% interest.

We apply a static set of transaction cost assumptions unique to each yield curve point. We do this by exercising judgment for each point on each yield curve, based on the observed bid-offer spreads across on- and off-the-run securities over our sample period. As evident in Table 3 below, bid-offer spreads are assumed to be notably higher for Chinese government bonds than for their German, UK and US counterparts. Note that JPY cash is assumed to be held in a cash account bearing a 0% interest rate.

Bid-offer spread assumptions Table 3

In cents

	US	DE	UK	JP	CN
0.25Y	0.003	0.010	0.015	0.000	0.005
2Y	0.006	0.020	0.015	NA	0.040
4Y	0.010	0.020	0.015	NA	0.080
6Y	0.012	0.025	0.020	NA	0.120
8Y	0.016	0.030	0.020	NA	0.160

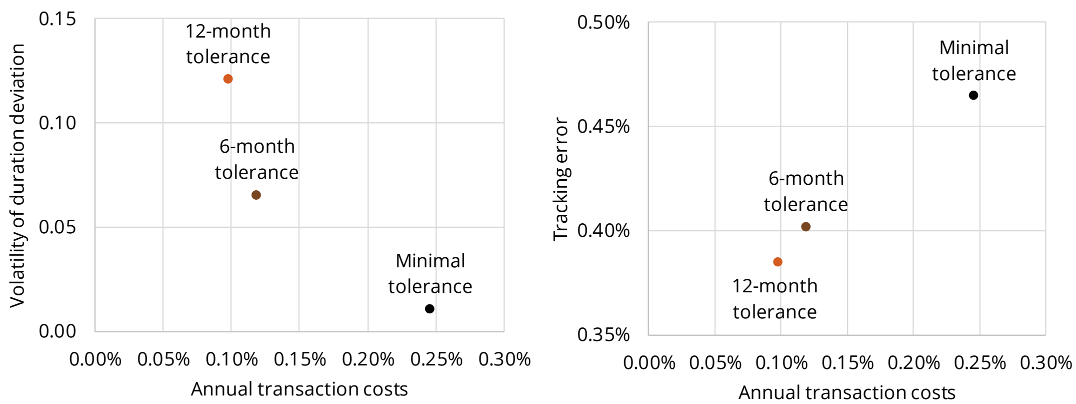
Source: Author.

There is a clear trade-off between our tolerance for duration drift at the portfolio level and the level of transaction costs we will incur. The relationship is intuitive: the less tolerance for duration drift, the more tightly duration positions will need to be managed. Such a reduced tolerance will create more frequent rebalancing and result in greater transaction costs.

Graph 6 displays the implications of various rebalancing rules for transaction costs, the volatility of portfolio duration deviation and tracking error. As seen in the left-hand panel, having minimal tolerance for duration deviation implies heavy transaction costs of around 25 basis points per annum, given that one is required to switch into the nearest duration security at each monthly rebalancing cycle. Allowing securities to roll down the curve six months past their target durations cuts transaction costs in half, at the sacrifice of a slight uptick in the volatility of overall duration deviations. Allowing securities to roll down a full year further cuts transaction costs to below 10 basis points per annum, although at modestly higher deviations in duration.

Annual transaction costs versus volatility of duration deviation (lhs) and tracking error (rhs) for the combined strategy

Years and percentage (lhs); percentage (rhs) Graph 6



Sources: Bloomberg, author's calculations.

One might expect deviations from a net duration position of 0 to result in higher portfolio volatility. The right-hand panel of Graph 6 shows the implications for tracking error with each of the rules. Allowing rolldown actually results in reduced levels of tracking error, despite the fact that duration deviation is more volatile. The desire to minimise transaction costs without taking on more risk motivates the choice of the one-year tolerance rule. Of course, the discussion around the management of duration assumes that that our hypothetical investor has only physical securities at their disposal. The ability to use futures would likely simplify the management of the overall duration position, probably at lower transaction costs.

5. Assessing performance

We examine the performance of each portfolio from multiple vantage points, including the excess return, ex post volatility (or tracking error), information ratio and maximum drawdown. We will do so gross and net of transaction costs, as well as decompose returns into their fixed income, FX and transaction cost components. We report results using an SDR and USD numeraire, recognising that the USD perspective may be more relevant for many investors.

Table 4 displays the performance across the three strategies before transaction costs. Viewed from the vantage point of the SDR, the carry strategy delivered flat returns over the sample before transaction costs, while the term premium strategy has an impressive information ratio of close to 1. The combined strategy has an information ratio of 0.69 and a reduced tracking error, owing to the diversification benefits of combining the signals.

Annual performance statistics gross of transaction costs — Table 4

Various units

	SDR numeraire				USD numeraire			
	Excess return	Tracking error	Information ratio	Maximum drawdown	Excess return	Tracking error	Information ratio	Maximum drawdown
Carry	0.05%	0.49%	0.10	2.25%	0.00%	0.49%	0.00	2.17%
Term premium	0.52%	0.52%	0.99	0.84%	0.57%	0.52%	1.10	0.81%
Combined	0.29%	0.41%	0.69	1.12%	0.28%	0.41%	0.69	1.03%

Sources: Bloomberg, author's calculations. Daily performance data from 1 October 2016 to 31 March 2023.

Table 5 shows the net performance once accounting for transaction costs. These push the carry strategy into negative territory, while the term premium strategy retains an impressive information ratio of 0.77 and the combined a respectable 0.47. The maximum drawdown is notably higher for the carry portfolio – over 2% – while those associated with the term premium and combined portfolios are much reduced, a little under or over 1%.

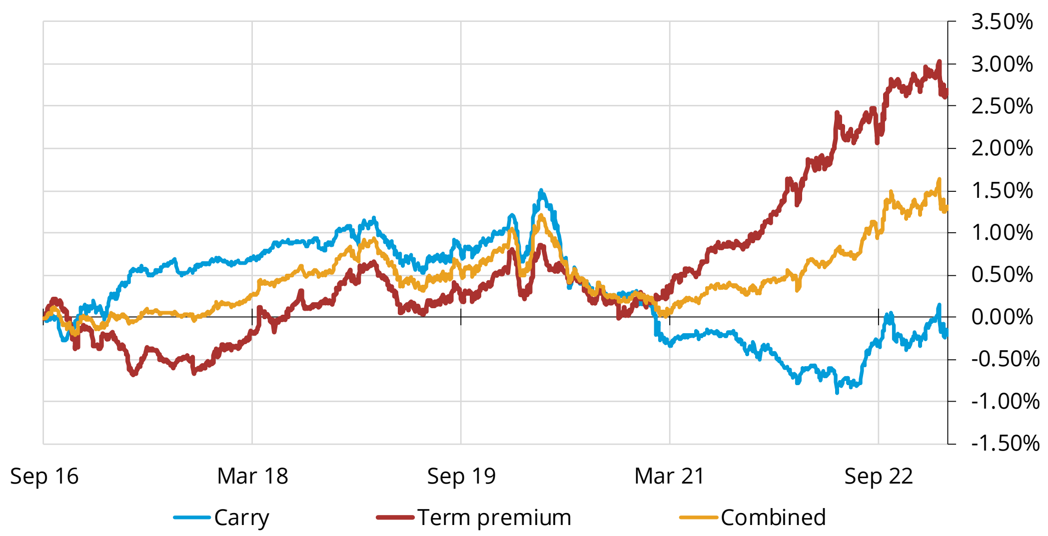
Annual performance statistics net of transaction costs | Table 5
Various units

	SDR numeraire				USD numeraire			
	Excess return	Tracking error	Information ratio	Maximum drawdown	Excess return	Tracking error	Information ratio	Maximum drawdown
Carry	−0.03%	0.49%	−0.07	2.40%	−0.08%	0.49%	−0.17	2.33%
Term premium	0.40%	0.53%	0.77	0.90%	0.45%	0.52%	0.87	0.90%
Combined	0.19%	0.42%	0.47	1.22%	0.19%	0.41%	0.47	1.13%

Sources: Bloomberg, author's calculations. Daily performance data from 1 October 2016 to 31 March 2023.

Graph 7 shows a time series of the cumulative returns of each strategy over the period. Several observations stand out. The early and latter parts of the sample show considerable divergence between the performance of the two portfolios, while during the middle years the performance was much more correlated. Comparing Graphs 5 and 7 in Section 3.2 helps explain why: positive duration contributions from CNY holdings are a dominant feature of both portfolios during the middle portion of the sample, while the weights are much less correlated during the earlier and later parts.

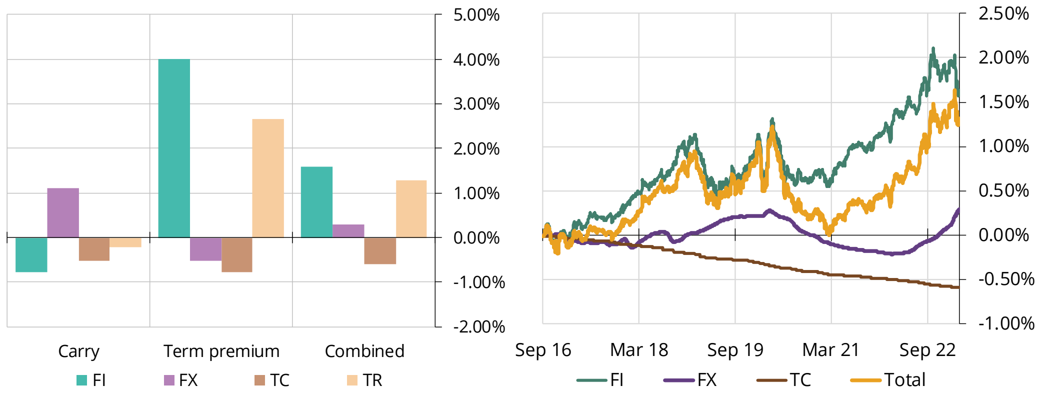
Cumulative returns of the carry, term premium and combined strategies
Percentage
Graph 7



Sources: Bloomberg, author's calculations. Results shown in SDR terms.

Graph 8 shows the decomposition of the combined strategy by return component, specifically the fixed income (FI) return, FX return and transaction costs (TC), which collectively comprise the total return (TR). As evident in the left-hand panel, FI returns were negative on net for the carry strategy, while the FX component actually lifted performance. In contrast, FI returns fully explain the more impressive performance of the term premium strategy, while FX returns were a modest drag. The combined strategy's positive performance is mostly explained by the FI component, while the FX component has been close to neutral, both on net and historically. This decomposition may reflect more favourably on the term premium and combined strategies, given that our signals convey information about carry and value from a fixed income perspective, rather than an FX one.

Decomposition of net returns across strategies (lhs) and cumulative return by component over time of the combined strategy (rhs) Graph 8
Percentage



Source: Bloomberg, author's calculations. Results shown in SDR terms.

6. Considering alternative setups

The constraints put in place reflect our hypothetical reserve manager, who as an investor with a low risk appetite is wary of currency and outright duration risk. In this section, we relax these assumptions by considering two alternative setups, with all other portfolio construction aspects remaining unchanged (unless otherwise noted):

a. Flexible duration
Outright duration positions of +/– 0.5 years are permitted at the portfolio level.

b. Unhedged CNY
The currency risk of CNY positions are not hedged back to the SDR.

We examine the portfolio weights and performance of each.

6.1 Flexible duration

The first alternative setup relaxes the constraint that overall duration must sum to zero. Instead, constraints are put in place to permit portfolio duration positions up to a minimum of –0.5 years and a maximum of 0.5 years.

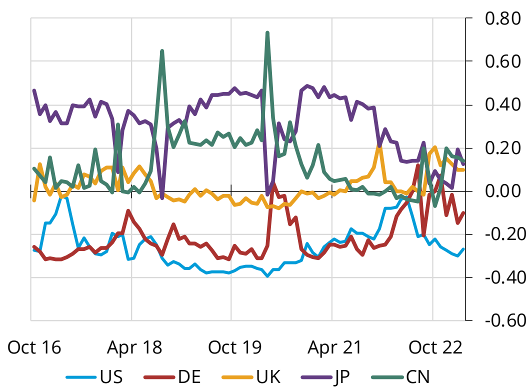
The reasoning behind relaxing this constraint is straightforward: just as our signals are used to exploit compensation for risk premia at the individual asset level, they can also be used to inform the direction of an overall duration position. For example, higher levels of term premium might motivate longer target durations for tactical or strategic asset allocation purposes, while conversely, negative term premia can motivate shorter target durations. With this alternative formulation, we test whether there is a meaningful performance impact from allowing this flexibility.

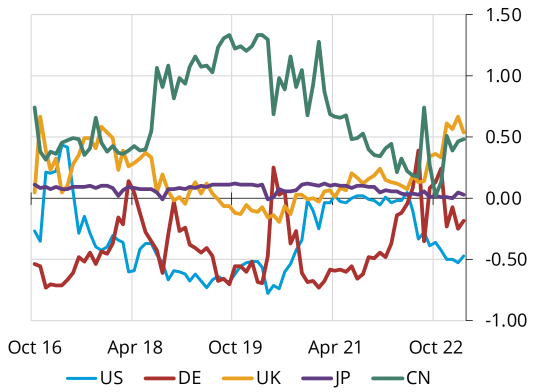
Graph 9 summarises the portfolio weights of the combined strategy under the flexible duration ("FD") setup. Thematically, they are quite similar to the baseline setup, generally featuring net long CNY and JPY positions over the history, funded with short USD and EUR. Duration contributions come largely from the CNY curve. Furthermore, the additional degree of freedom reduces portfolio turnover slightly, from 32% in the baseline strategy to about 30%. This will help with transaction costs, even if only slightly.

Net positions (lhs) and duration contributions (rhs) of combined strategy with flexible duration

Graph 9

Values (lhs); years(rhs)





Sources: Bloomberg, author's calculations.

As shown in Table 6 the flexible duration setup exhibits similar though slightly inferior performance to the fully hedged baseline. It is thus hard to conclude that allowing drift in overall duration adds meaningful value, at least over our sample.

Annual performance statistics net of transaction costs with flexible duration

Table 6

Various units

	SDR numeraire				USD numeraire			
	Excess return	Tracking error	Information ratio	Maximum drawdown	Excess return	Tracking error	Information ratio	Maximum drawdown
Carry	–0.08%	0.56%	–0.13	3.05%	–0.10%	0.56%	–0.17	2.99%
Term premium	0.34%	0.53%	0.64	0.99%	0.40%	0.53%	0.74	1.01%
Combined	0.15%	0.43%	0.34	1.36%	0.16%	0.43%	0.38	1.25%

Sources: Bloomberg, author's calculations. Daily performance data from 1 October 2016 to 31 March 2023.

6.2 Unhedged CNY

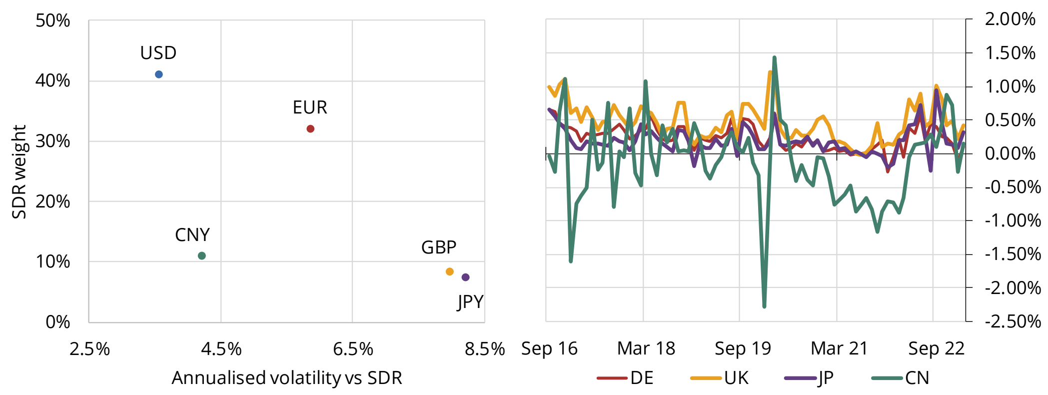
The second alternative setup relaxes the constraints around currency risk, specifically allowing open FX exposure for any CNY positions.[12] Most other characteristics of the baseline setup remain in place, including the requirement that duration be neutral at the portfolio level. The one exception is the target tracking error, which we allow to be 100 basis points given the additional FX risk. Of course, the tracking error in itself may not be very important, given that the size of an actual portfolio can simply be scaled to give the desired risk level in nominal terms.

The motivation behind choosing not to hedge some portion of FX risk may be on shakier ground than the previous formulation. Our strategy seeks to exploit information content from signals that reflect interest rate risk, not currency risk. There is certainly empirical evidence that FX carry trades can deliver excess returns (see Fama (1984)), but they reflect a broader set of risks arguably less connected to our signals. For example, Menkhoff et al (2012) show that global FX volatility is an important risk factor in explaining the performance of FX carry trades.

Why do we choose to leave the CNY unhedged as opposed to a different currency or currencies? Several reasons motivate the choice. First, the managed flexible exchange rate exchange rate regime – in place since mid-2015 and paving the way for SDR inclusion – is of a different nature than the free-floating regimes in the euro area, Japan, the United Kingdom and the United States and, potentially helping to mitigate, relatively speaking, unhedged FX risk. The left-hand panel of Graph 10 demonstrates this point, displaying the volatility of each SDR constituent vis-à-vis the SDR's weights. As shown, the CNY-SDR FX rate displays comparable volatility with that of the USD, despite the fact that the CNY's weight in the SDR basket is only about one-quarter in magnitude. In contrast, the volatilities of GBP and JPY – whose weights are of similar magnitude to the CNY – display volatilities around twice that of the CNY.

12 Given the lack of hedging, the FX component for the CNY term premium signal is set to zero in this formulation.

Annualised daily FX volatility versus the SDR (lhs) and excess hedged FX returns vis-à-vis the USD (rhs) Graph 10
Percentage



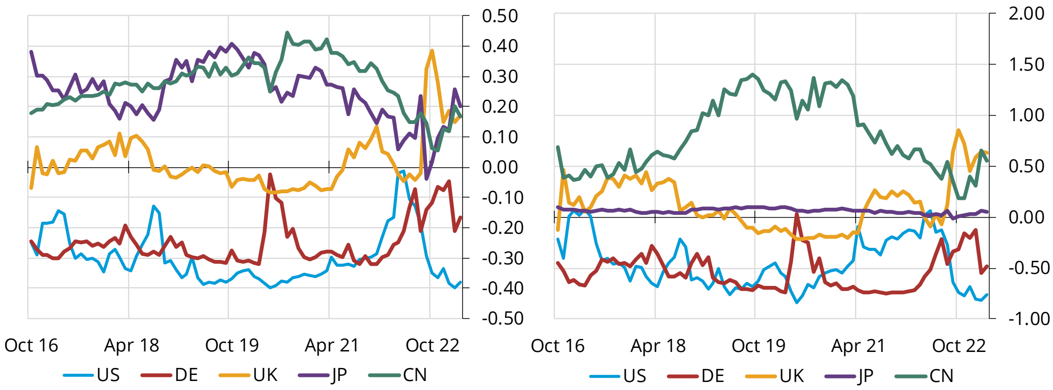
Sources: Bloomberg, author's calculations.

Furthermore, hedging the CNY comes with drawbacks relative to the other currencies. As shown in the right-hand panel of Graph 10, the CNY-USD excess hedged FX return, as measured using three-month government bond yields (the relevant perspective for our strategies), is considerably more volatile than its EUR, GBP and JPY counterparts. As discussed in more detail below, this fact helps explain the volatility in baseline strategy portfolio weights, and by extension the strategy's relatively high turnover.

Graph 11 summarises the portfolio weights of the unhedged CNY formulation. They are thematically similar to the baseline and flexible duration setups, but illustrate more persistent biases towards long CNY duration positions – unsurprising perhaps, given the relatively attractive interest rate levels coupled with the lack of hedging costs. Portfolio weights are also considerably more stable than in the baseline and flexible duration cases, cutting turnover by more than a third to around 20% and helping to reduce transaction costs.

Net positions (lhs) and duration contributions (rhs) of combined strategy with CNY positions unhedged — Graph 11

Values (lhs); years(rhs)



Sources: Bloomberg, author's calculations.

As shown in Table 7, the unhedged CNY formulation dominates the other two in terms of performance, delivering an information ratio – after transaction costs – of over 1 when using an SDR numeraire. The greater allocation to CNY and lack of hedging costs each play important roles in boosting the performance. More recently, the appreciation of the CNY has also contributed heavily, accounting on net for about 1 percentage point of the sample-wide performance. The smaller transaction costs – about 6 basis points per year, down from about 10 in the baseline setup – also help. Importantly, the higher tracking error helps reduce their *relative impact*, with transaction costs as a percentage of total performance dramatically lower when the CNY is left unhedged than in the baseline or flexible duration formulations.

Annual performance statistics net of transaction costs with CNY positions unhedged — Table 7

Various units

	SDR numeraire				USD numeraire			
	Excess return	Tracking error	Information ratio	Maximum drawdown	Excess return	Tracking error	Information ratio	Maximum drawdown
Carry	1.00%	1.31%	0.76	2.51%	0.55%	1.03%	0.53	2.52%
Term premium	1.15%	1.19%	0.97	1.76%	0.78%	1.08%	0.72	2.76%
Combined	1.08%	1.21%	0.89	1.91%	0.67%	1.02%	0.66	2.30%

Sources: Bloomberg, author's calculations. Daily performance data from 1 October 2016 to 31 March 2023.

Performance is modestly less impressive when viewing results from the USD numeraire perspective. Indeed, the choice of numeraire is more consequential when FX risk is left unhedged.

As noted above, tracking error may not be the most relevant risk metric for certain types of investor. For example, reserve managers, who can be quite sensitive to the near-term performance of their tactical positions, may see maximum drawdown as a more relevant consideration. Strategies that have periods of sharp drawdowns may come under pressure from a board or investment committee with limited risk appetite, even if medium-term expectations are consistent with a recovery in performance. As evident in Table 7, the maximum drawdown of the unhedged CNY formulation, in *percentage terms*, is quite elevated relative to the baseline and flexible duration versions, reflecting in large part the FX risk stemming from CNY positions. While reserve managers could scale down the size of their portfolios in nominal terms to account for the higher risk (as noted above with the TE), investment committees and risk departments may still be troubled by the larger drawdown risk per unit invested.

Concluding remarks

This paper has drawn from previous research at the Public Investors Conference to formulate and test systematic investment strategies based on two well known concepts in the fixed income space: carry and term premium. Our main conclusion – that term premium strategies deliver promising excess returns – is consistent with the earlier papers, lending additional confidence to the finding.

What can we say about the lukewarm results delivered by the carry strategy? Coche et al (2018) document that positive excess returns associated with the carry strategy began to take hold persistently only in the 1980s – the beginning of a period marked by a secular decline in interest rates, when high-carry assets would have naturally outperformed. Our sample, however, is taken after this period, with most SDR yield curves having reached their historical lows in the early to mid-2010s. We may thus be suspicious of the value of the carry signal outside such environments.

Why might an investor want to nonetheless consider a combined strategy, when the carry strategy looks less compelling and the information ratio of the combined strategy appears inferior, at least over our sample? One reason may be the management of model risk, as estimating a term premium requires an unobservable element to be estimated – the expected short rate path. While our macro-based approach is consistent with the general understanding of the drivers of central bank policy, it may not always represent the market's physical expectations for the evolution of the policy rate. The combined strategy helps to reduce this model risk, given that carry is a model-free measure. It also helps to diversify the risk – as evident in the lower tracking error, and a maximum drawdown figure that is much closer to that of the term premium strategy than the carry strategy. Finally, it's worth noting that the combined strategy has spent nearly the entire sample above water, save for a few days here and there in slightly negative territory.

There are nonetheless many reasons to be cautious with strategies, despite the promising results reviewed here and in earlier research. Any backtesting exercise is by virtue a reflection of past market dynamics, and provides no guarantee of a future relationship. Systematic strategies also remove discretion by design, and may be slow to recognise regime changes. For example, estimation of term premia may lag important market developments, such as the onset of a pandemic, implying that signals may have less meaning during times of rapid changes in the financial environment.

These cautionary points lead to a few simple final remarks. First, investors in such strategies cannot necessarily expect immediate results, and should be prepared for periods – perhaps sustained periods – of underperformance. This can be a challenge, particularly in the presence of investment boards with a low tolerance for P&L volatility. Second, risk should be appropriately calibrated, such that periods of underperformance and incidents of sharp drawdowns can be sustained without abandoning the strategy. Third, it is useful to keep in mind why we would expect signal-based strategies to work in the first place. As Bjorheim et al (2018) point out, strategies such as these may be powerful tools not due to the superior quality of the models and their predictive power, but because they give rise to reasonable estimates of risk premia. The investor is thus merely being compensated for exposure to risk. A key element to investigate further may be how correlated such risk factors are with other elements of active management, or with the underlying strategic asset allocation. At the end of the day, systematic strategies such as these may prove most useful as a way of diversifying elements of a broader strategic and active management approach.

References

Adrian, T, R Crump and E Moench (2013): "Pricing the term structure with linear regressions", *Staff Reports*, Federal Reserve Bank of New York, no 340.

Ahmerkamp, J and J Grant (2013): "The returns to carry and momentum strategies", Imperial College Business School Working Paper.

Asness, C, T Moskowitz and L Pedersen (2013): "Value and momentum everywhere", *The Journal of Finance*, vol 68, no 3, pp 929–85.

Baz, J, N Granger, C Harvey, N Le Roux and S Rattray (2015): "Dissecting investment strategies in the cross section and time series", working paper.

Bjorheim, J, J Coche, A Joia and V Sahakyan (2018): "A macro-based process for actively managing sovereign bond exposures", *Advances in the Practice of Public Investment Management*, pp 103–29, Palgrave Macmillan.

Coche, J, Knezevic, M and V Sahakyan (2018): "Carry on", *Advances in the Practice of Public Investment Management*, pp 131–59.

Fama, E (1984): "Forward and spot exchange rates", *Journal of Monetary Economics*, no 14, pp 319–38.

Fender, I, M McMorrow, V Sahakyan and O Zulaica (2020): "Reserve management and sustainability: the case for green bonds?", *BIS Working Papers*, no 849.

Fender, I, M McMorrow and O Zulaica (2022): "Sustainable management of central banks' foreign exchange (FX) reserves", *The Inspire Sustainable Central Banking Toolbox*, Policy Briefing Paper, no 6.

International Monetary Fund (2016): "IMF adds Chinese Renmimbi to Special Drawing Rights basket", press release, 30 September.

Kim, D and J Wright (2005): "An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates", *Finance and Economics Discussion Series*, Federal Reserve Board, no 2005-33.

Koijen, R, Moskowitz, T, Pedersen, L and Vrugt, E (2017): "Carry", *Journal of Financial Economics*, pp197–225.

Lombardi, M and F Zhu (2018): "A shadow policy rate to calibrate US monetary policy at the zero-lower bound", *International Journal of Central Banking*, vol 14, no 5, pp 305–46.

Markowitz, H (1952): "Portfolio selection", *The Journal of Finance*, vol 7, no 1, pp 77–91.

Menkhoff, L, L Sarno, M Schmeling and A Schrimpf (2012): "Currency momentum strategies", *Journal of Financial Economics*, vol 106, issue 3, pp 660–84.

Nelson, C and A Siegel (1987): "Parsimonious modelling of yield curves", *The Journal of Business*, vol 60, issue 4, pp 473–89.

Nyholm, K (2015): "A rotated dynamic Nelson-Siegel model with macro-financial applications", *ECB Working Paper Series*, no 1851.

Stock, M and M Watson (2007): "Why has US inflation become harder to forecast?", *Journal of Money, Credit and Banking*, supplement to vol 39, no 1.

The backtesting of value-at-risk and expected shortfall: evidence from the Bank of Italy's foreign reserves portfolio

Marco Fruzzetti,[1] Davide Nasti[2] and Marco Orlandi[3]

Abstract

Value-at-risk (VaR) and expected shortfall (ES) are widely used by financial institutions as tools for measuring market risk. Despite its well known properties as a risk measure, a debate on ES backtestability and its use for risk models validation is still ongoing. We assess the innovative ES Acerbi-Szekely Ridge Backtesting approach by using Bank of Italy foreign reserves portfolio data. We compute the VaR and ES estimates with a well established, multivariate, conditional normal RiskMetrics model. As expected, the model passes traditional VaR backtests, showing accuracy in estimating VaR; however, ES estimates calculated using the RiskMetrics model often fail to pass the Ridge Backtest. These results show that the ES Ridge Backtesting can accurately identify mispredictions of real, thick-tailed, unconditional return distributions.

JEL classification(s): C52, G32.

Keywords: backtesting, risk measures, value-at-risk, expected shortfall.

[1] Bank of Italy, Economic Outlook and Monetary Policy Directorate, marco.fruzzetti@bancaditalia.it.

[2] Bank of Italy, Financial Risk Management Directorate, davide.nasti@bancaditalia.it.

[3] Bank of Italy, Financial Risk Management Directorate, m.orlandi@bancaditalia.it.

1. Introduction

Central banks focus on monetary policy and emergency liquidity assistance, which can involve very high financial risks. Consequently, a high risk aversion is common among central bank investment portfolio operations; large financial losses can lead to capital depletion and undermine a central bank's ability to perform its institutional functions. Although a central bank can technically operate even with negative capital, a positive net value is generally desirable because it bolsters public confidence in the institution and helps to preserve its independence. Furthermore, losses on the investment portfolio expose the bank to reputational risks. Therefore, it is essential for a central bank, when assessing the risks of its investment portfolio, to adequately take into account the interdependencies between the risks of its activities and the current and future adequacy of its assets.

To support decision-making on financial risks, a central bank must have a measurement system that is as accurate as possible. To this end, the Bank of Italy (henceforth BdI) uses specific models and seeks to verify both forecasting accuracy and the adequacy of the assumptions. Various methods can be used to validate these models, including backtesting.

Backtesting has long been a topic of interest for the Basel Committee on Banking Supervision, which has set out a risk management framework for commercial banks. And a number of central banks have also decided to adopt risk practices along the lines of these principles.

For some years, BdI has estimated the market risk of its portfolios using not only the value-at-risk (VaR) metric but also expected shortfall (ES), in line with the market risk regulations set out in the Basel Committee's Fundamental Review of the Trading Book (FRTB).

To monitor the risks of foreign exchange reserves, BdI, like other central banks, uses the RiskMetrics model to calculate a VaR. BdI is therefore interested in backtesting this risk measure. As BdI mainly focuses on the distribution tails, it has recently derived ES measures by applying a transformation to the RiskMetrics VaR and has subsequently applied a backtesting procedure to such measures.

Established backtesting techniques are available for VaR, while the literature on the ES backtesting is more fragmented. Nonetheless Acerbi and Szekely (2014), hereafter "AS", have recently proposed a well received approach that allows a validation procedure for ES models within a first-order approximation.

After a brief review of the underlying theoretical foundations, this article presents the results of a backtesting strategy for BdI's foreign reserves risk measures, with a specific focus on the AS procedure and its innovative characteristics.

2. VaR backtesting techniques

Backtesting techniques designed to assess the accuracy of VaR estimates typically focus on the **unconditional coverage** and **conditional coverage** of VaR.

The simplest test of a VaR model involves **counting the number** of days in which the actual portfolio loss is greater than the VaR forecast, which results in a VaR exception or violation.

Unconditional coverage tests aim to determine whether this number is statistically different from the number of breaches expected from a VaR estimate. For example, a VaR model with a confidence level of 95% would expect five breaches out of 100 observations. If the distance between the number of expected and realised exceptions is not (statistically) negligible, the estimates need to be improved.

In contrast, conditional coverage tests consider the exact point at which the violations occurred. Precise VaR models can react to changes in volatility and correlations to maintain **independence** of breaches. VaR modellers are interested in having a risk measure that can manage volatility clustering, as significant losses that occur in rapid succession increase the probability of default of the financial institutions compared with widely distributed individual losses. Conditional coverage tests address this problem by assessing whether the distribution of breaches over time is significantly different from the expected random one.

It is important to note that both types of test must be satisfied by an accurate VaR model. Tests that jointly examine the unconditional and conditional coverage provide an option to detect VaR measures shortfall.

Following the related literature and best practices, BdI decided to adopt the "Binomial" and the "Proportion of failures (POF)" tests for unconditional coverage tests and the "Christoffersen" and the "Time Between Failures Independence (TBFI)" for conditional coverage tests.

The sequence of successes and failures resulting from the comparison between portfolio effective results and VaR forecasts is known as a *Bernoulli trial* and the number of VaR exceptions follows a binomial distribution.

The binomial test (Haas (2001)) compares the number of VaR exceptions with the expected number, considering exceptions variability. Under the null hypothesis of having a VaR model with good predictive power, the null is rejected when the number of VaR exceptions, given a confidence level of the test (typically 95%), lies beyond the threshold value. Also, the POF test uses the binomial distribution (Kupiec (1995); Haas (2001)). In addition to the Binomial test, the POF verifies through a likelihood ratio (LR) test that the probability of exceptions is consistent with the probability p implied in the confidence level of VaR. If the data suggest that the probability of exceptions is different than p, the VaR model is rejected. According to the Neyman-Pearson Lemma, the LR test is the most powerful of its class.

The Christoffersen test (Christoffersen (1998)) examines the concept of independence between VaR violations by examining whether the probability of a violation on any given day depends on the previous day result. The relevant test statistic for verifying this kind of independence is an LR. The Christoffersen test considers only the dependence between observations on two subsequent days. However, it is possible that today's VaR violation does not depend on the violation that occurred yesterday but on the violation that occurred, for example, a week ago. The TBFI test (Haas (2001)) measures the time between exceptions. For the first VaR exception, the TBFI test is implemented as a standard time-until-first-failure (TUFF) test (Kupiec (1995); Haas (2001)); the TUFF is based on a LR and similar assumptions with respect to the POF test, measuring the number of days until the first VaR violation occurs (here the target variable is the distance, in days, between the first VaR forecast and the first VaR violation). After calculating the LR ratio statistic for each exception, a test is obtained with a null hypothesis that assumes that each exception is independent from the others.

In order to conduct a more comprehensive analysis, we also performed two joint tests: the first is obtained combining the Christoffersen's statistic with the TBFI ("Mixed Kupiec" joint test; Haas (2001)), the second results from the sum of the Christoffersen's statistic and the POF (we named it the "Mixed Christoffersen" joint test; Nieppola (2009)).

3. ES backtesting techniques: the Acerbi-Szekely test

The literature on the backtesting of ES forecasts is quite fragmented. The ES belongs to the class of "coherent" risk measures (Acerbi and Tasche (2002)), whose concept was introduced (Artzner et al (1997)) and formalised (Artzner et al (1999)) in the late 1990s. However, only in 2011 emerged a stream of research that identified the properties of backtestable statistics and ascertained if ES belonged to this class of measures. After Gneiting's breakthrough result (Gneiting (2011)) that showed that ES is not elicitable,[4] it gradually became clear (Acerbi and Szekely (2014); Acerbi and Szekely (2017)) that ES is not backtestable either and, therefore, that an exact validation procedure cannot exist for ES models.

Nonetheless, it has been proven that VaR and ES are jointly elicitable (Fissler and Ziegel (2016).) This allows **selection procedures** based on the ranking of different risk models.[5]

Building on these concepts, Acerbi and Szekely have proposed an innovative approach, which they named *ridge backtesting*. This approach allows, within a first order approximation, a **validation procedure of an absolute type** for ES models (Acerbi and Szekely (2017); Acerbi and Szekely (2019)).

The new approach is based on the following *ridge backtesting* function:

$$Z_{ES_\alpha} = e - v - \frac{1}{\alpha}(x + v)_- \qquad (1)$$

where e, v and x are respectively the ES forecast, the VaR forecast and the observed return (X is the random variable for the portfolio returns), α is the confidence level of the ES and VaR forecasts and $(x + v)_-$ stands for $min(x + v, 0)$ [6]. The authors showed that a backtest based on Z_{ES_α} suffers from a *bias*, $B(v)$, that : a) has a prudential effect, b) is small around $v = \boldsymbol{VaR_\alpha}$[7] (Acerbi and Szekely (2017)); consequently, it is possible to use Z_{ES_α} in order to validate ES models, as long as VaR forecasts are sufficiently accurate.

4 A statistic is said to be *elicitable* if it can be obtained as the minimiser of its expected scoring function.

5 Comparative backtests can be carried out using Fissler and Ziegel's strictly consistent loss functions (containing forecasts of both risk measures, ie VaR and ES) and the Diebold-Mariano test (Diebold and Mariano (1995)).

6 The function in (1) is based on the following way of writing ES: $ES_\alpha = \min_v E\left[v + \frac{1}{\alpha}(X + v)_-\right]$ (Acerbi and Szekely (2019)).

7 Under weak regularity conditions, the sensitivity of a ridge-backtest to the auxiliary variable (in the ES case such variable is the VaR) forecast is zero at first order (Acerbi and Szekely (2017)).

Acerbi and Szekely also showed that an ES *ridge* backtest satisfies, at first order, a highly desirable property that they named *sharpness*: the expected value of Z_{ES_α} provides a measure of the average forecast error (Acerbi and and Szekely (2019)) given by the difference between: a) the average ES forecast and b) an estimator of the average true value of ES,[8] the latter quantity leads to a notion of *realised* ES which can be calculated as follows (Acerbi and Szekely (2019)):

$$\widehat{ES}_\alpha = \frac{1}{T}\sum_{t=1}^{T}[v_t + \frac{1}{\alpha}(x_t + v_t)_-] \quad (2)$$

This means that the ES *ridge* backtest depends on the amplitude of the forecast errors, in contrast to the VaR backtest (the binomial test).

Practically, the new backtesting procedure consists of a standard, one-sided hypothesis test, based on the mean backtest function:

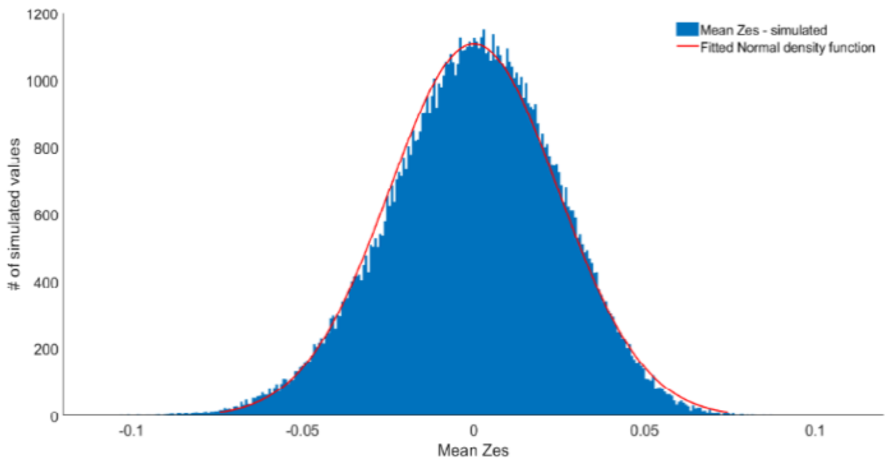
$$\bar{Z}_{ES_\alpha} = \frac{1}{T}\sum_{t=1}^{T} Z_{ES_\alpha}(e_t, v_t, X_t) \quad (3)$$

where $t = 1, \dots, T$ is a discrete sequence of time, e_t and v_t are series of ES and VaR model forecasts respectively, and $X_t \sim P_t$, being P_t the predictive distribution for X given by the model at time t. The one-sided null hypothesis ($\bar{Z}_{ES_\alpha} \geq 0$) is verified comparing the predictive distribution $P_{\bar{Z}_{ES}}$ of the test statistic and its realised value $\bar{z}_{ES}$:

$$\bar{z}_{ES} = \frac{1}{T}\sum_{t=1}^{T} Z_{ES}(e_t, v_t, x_t) \quad (4)$$

where x_t is a series of realisations of X [9]. $P_{\bar{Z}_{ES}}$ is obtained through a Monte Carlo simulation (Acerbi and Szekely (2017)); an example is shown in Graph 1.

Simulated distribution for the $\overline{\boldsymbol{z_{ES_\alpha}}}$ test statistic — Graph 1



This chart shows the $\overline{Z_{ES_\alpha}}$ test statistic distribution, obtained through a Montecarlo simulation (100,000 scenarios) on the BdI's USD portfolio (see below).

Source: Bank of Italy.

[8] Such estimate is biased for the *ridge backtesting* whereas it is correct for sharp strict backtests.

[9] So, $\bar{z}_{ES}$ is the average realisation of the backtest function over the period $t = 1, \dots, T$; considering different values of x_t, drawn from the distribution P_t, it is possible to provide the distribution of $\bar{z}_{ES}$ (see Chart 1).

One can also consider a relative and dimensionless version of the test by normalising the *ridge backtesting* function Z_{ES_α} with ES forecasts. This yields a new backtesting function denoted as $Z^{rel}_{ES_\alpha}$, which is defined as $Z^{rel}_{ES_\alpha} = \frac{Z_{ES_\alpha}}{e}$. The resulting test statistic is:

$$\bar{z}^{rel}_{ES_\alpha} = 1 - \widehat{\varphi_{ES}} \quad (5)$$

where

$$\widehat{\varphi_{ES}} = \frac{1}{T}\sum_{t=1}^{T} \frac{v_t + \frac{1}{\alpha}(x_t + v_t)_-}{e_t} \quad (6)$$

is a positively biased estimator of the average prediction ratio $\frac{ES_\alpha}{e}$ between the true and the predicted ES (Acerbi and Szekely (2019)). This estimator is conservative and provides a prudential estimate. The scaling factor that needs to be applied to ES forecasts to bring $\bar{z}^{rel}_{ES_\alpha}$ back to zero is simply the estimated ratio $\widehat{\varphi_{ES}}$.

4. Data description

To construct the reference data set for this study, we used data from 1 January 2012 to 31 March 2021. Specifically, we followed these steps:

1. We obtained a daily time series of 95% VaR measures[10] for each BdI foreign reserves portfolio;
2. We then associated each time series with actual returns in local currency.

Due to space limitations, we only present evidence relating to two of the Bank's foreign exchange reserve portfolios: those in United States dollars and in Japanese yen.

Firstly, we examined the statistical properties of the daily return series. The only notable feature that warrants consideration is the high value for the *asymmetry and kurtosis* parameters of the Japanese yen portfolio (Table 1), which may have a significant impact on our analysis.

Descriptive statistics for foreign reserves portfolios — Table 1

Foreign reserves portfolio	Mean	Standard Deviation	Asymmetry	Kurtosis
$ US	0.005%	0.01%	-0.1%	5.1
¥ JP	0.001%	0.03%	-1.1%	14.7

Source: Bank of Italy.

Secondly, we had to make a decision about the sample period to use for the backtesting analysis. Regulation often requires using the latest available year of

[10] Calculated with the RiskMetrics™ methodology.

data,[11] but statistical techniques show that longer time periods can significantly increase the power of the tests.

For both risk measures, VaR and ES, we gradually increased the *backtesting period*[12] with one-day steps for VaR and 10-day steps for ES[13]. This allowed us to evaluate the outcome of the tests on the entire data set and understand its dynamics over time. To have a short-term result comparable to the "Basel" logic, we also conducted the same exercise, with a fixed backtesting period over one year of rolling data, updated on a daily basis for the VaR and every 10 days for the ES.

The RiskMetrics™ model is used in this paper to calculate risk measures. It provides a daily conditional forecast of return volatility for various risk factors. Assuming the returns' conditional distributions to be normal, the estimates of VaR and ES are at the 95% confidence level.

The daily volatility forecast is based on the assumption that only recent returns affect the volatility estimates (the EWMA decay parameter λ is 0.94). For data availability reasons, we used monthly volatility forecasts (in that case the optimum λ is 0.97), appropriately rescaled to daily frequency using the "square root of time" rule. The bias imposed by this transformation is negligible (Hendricks (1996)).

As previously mentioned, the chosen approach assumes that the risk factors returns are distributed according to a multivariate normal distribution with zero mean. This conceptual framework allows, for each confidence level, to pass from VaR to ES forecast using the following formula:

$$ES_\alpha = \frac{VaR_\alpha}{z_\alpha \cdot (1-\alpha)} \cdot \phi[\Phi^{-1}(\alpha)] \qquad (7)$$

where VaR_α e ES_α are, respectively, the VaR and ES of the portfolio at the selected confidence level α (which is 95% in the data set being examined), while z_α, ϕ e Φ are, respectively, the quantile of order α, the probability density function and the distribution function of the standard normal distribution. The ES forecasts for each foreign reserve portfolio were obtained using equation (7).[14]

5. VaR backtesting results

Table 2 shows the results of joints tests applied to the US dollar portfolio, which appear to be **contradictory**. Using the increasing time window, the Mixed Christoffersen test fails (p-value of 2%) while the Mixed Kupiec test is not rejected. **The frequency tests** suggest that the VaR model is not entirely reliable since they were **rejected, with p-values** of **3.1%** (Table 2). Further inspection reveals that the reference model tends to **overestimate** the VaR, leading to a lower number of violations than expected (Graph 2, left-hand panel). The Christoffersen independence

11 Market risk standard set out by the Basel Committee on Banking Supervision.

12 To be distinguished from the *lookback period*, which refers to the issue of estimating risk measures (in other words, one wonders how much history must be considered in order to produce the most accurate VaR forecast possible).

13 The different size of the time steps is due to computational reasons.

14 In this regard, it is interesting to put in evidence that: using RiskMetrics, for $\alpha = 0.95$ it is possible to say that $ES = 1.254 * VaR = 2.1 * standard\ deviation$. So, (conditional) gaussian models present a fundamental limit: going from basic to more sophisticated risk measures does not bring any truly new information about risk.

test and the TBFI were **not rejected** (Table 2) with regard to the **independence** of the violations.

Test results for USD portfolio — Table 2

Whole period

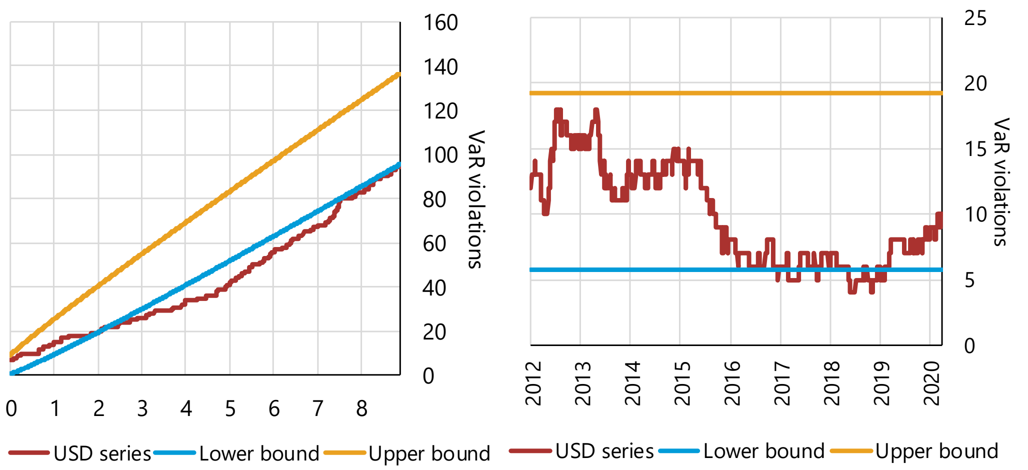
Class of test	Test	p-value / #breaches	Rejection
Joint	Mixed Kupiec	26.8%	5%
	Mixed Christoffersen	2.00%	5%
Frequency	Binomial	94	< 95.3 , > 136.5
	POF	3.1%	5%
Independence	Christoffersen	7.7%	5%
	TBFI	26.8%	5%

Source: Bank of Italy.

Lastly, the binomial test was applied to a rolling time window with a fixed length of 250 trading days (Graph 2, right-hand panel). This was to ensure compliance with the Basel framework, as the binomial test fundamental to the traffic light approach is used to validate banks' risk models.

USD portfolio binomial test results — Graph 2

Increasing and rolling backtesting period



Results of the binomial test. The amount of data used for the tests (in years, left-hand panel; equal to 250 days, right-hand panel), starting from the most recent observations, is shown on the x-axes; the number of VaR violations on the y-axes. The rejection region is represented by the area that is not included between the orange line, which indicates an underestimate of the VaR, and the blue one, which indicates an overestimate.

Source: Bank of Italy.

Regarding the Japanese yen portfolio, the results of the **joint tests**, over a large part of the increasing size windows, fall in the **rejection region** at the 5% confidence level (Table 3, mixed Christoffersen test). However, **the frequency tests** indicate that the VaR model appears to be **well calibrated**. Both tests are passed in the presence of p-values significantly higher than the threshold (Table 3 and Graph 3, left-hand panel).

Test results for JPY portfolio — Table 3

Whole period

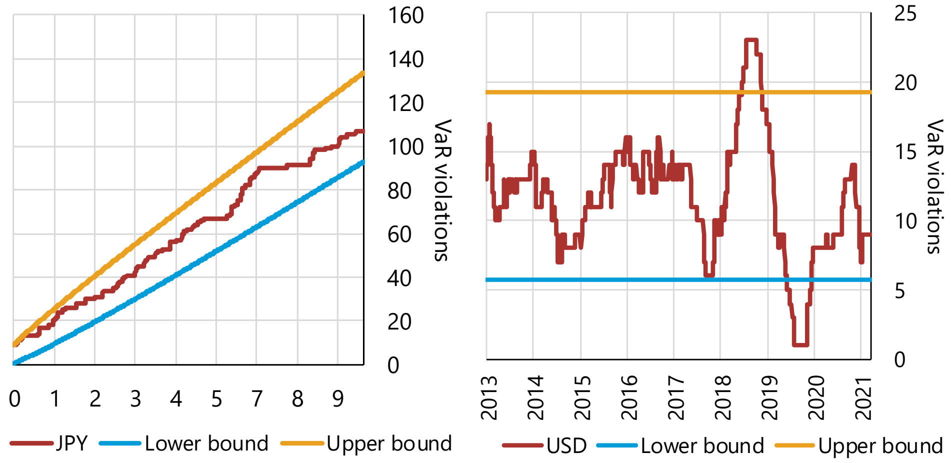
Class of test	Test	p-value / #breaches	Rejection
Joint	Mixed Kupiec	0.00%	5%
	Mixed Christoffersen	0.00%	5%
Frequency	Binomial	107	< 92.7 , > 133.4
	POF	55.30%	5%
Independence	Christoffersen	0.00%	5%
	TBFI	0.00%	5%

Source: Bank of Italy.

The results obtained with a reference time window of fixed length confirm the good performance of the model, although the series shows alternating periods of over and underestimation of risks (Graph 3, right-hand panel).

JPY portfolio binomial test results — Graph 3

Increasing and rolling backtesting period



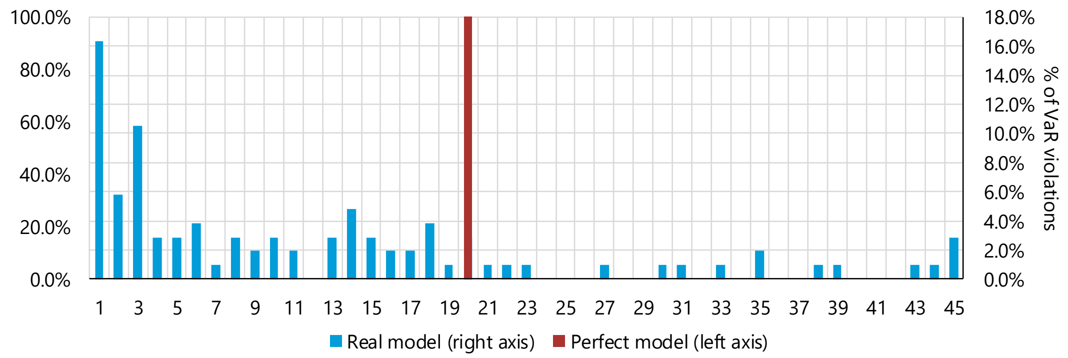
Results of the binomial test. The amount of data used for the tests (in years, left-hand panel; equal to 250 days, right-hand panel), starting from the most recent observations, is shown on the x-axes; the number of VaR violations on the y-axes. The rejection region is represented by the area that is not included between the orange line, which indicates an underestimate of the VaR, and the blue one, which indicates an overestimate.

Source: BdI.

The clustering of the violations might be the reason behind the rejection of the joint tests, as both the Christoffersen's **independence test** and the TBFI were

rejected. The Christoffersen independence test shows good results for half of the sample, where there is a 16% chance of having another violation the next day and almost a 40% chance within five days (Graph 4). This evidence highlights the limitations of a model designed to be responsive to market conditions.

TBFI test – focus on the breach distance Graph 4



Distribution of distances (in days) between VaR violations. The orange histogram represents the theoretical model of the TBFI test (left axis, all violations are spaced 1/α days apart), the blue histograms indicate the real observations (right axis).

Source: Bank of Italy.

6. ES backtesting results

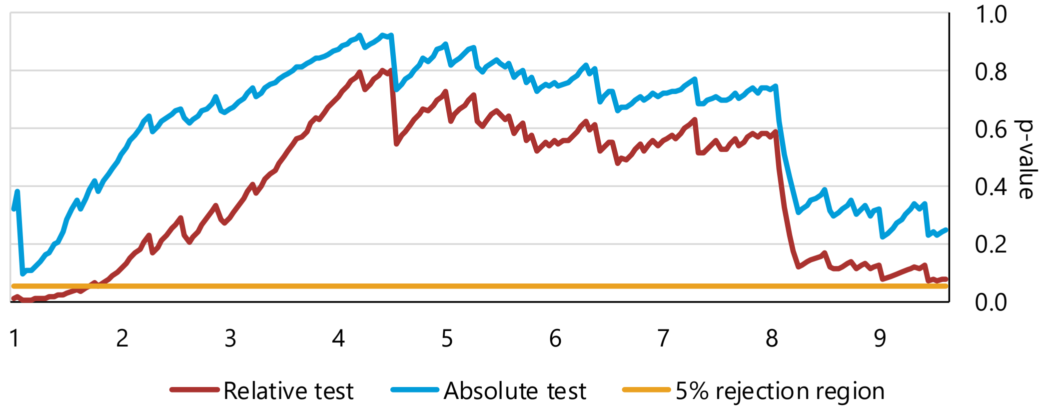
For the USD portfolio, **ES forecasts calculated using the** RiskMetrics™ model **show an 'average performance' that is satisfactory** when considering AS tests with increasing backtesting periods (Graph 5). The p-values for the entire data set (spanning over nine years, 01/04/2012–31/03/2021) are 0.25 and 0.08 for the absolute and relative test respectively.

Despite large fluctuations in the p-values, **absolute tests are always passed**. However, **relative tests are rejected** for backtesting periods shorter than about two years. For longer periods, the relative p-values are always greater than 0.05, but lower than the absolute ones.

USD portfolio AS test results

Increasing backtesting period

Graph 5



The figure shows the results of relative and absolute AS tests; the backtesting periods, indicated on the horizontal axis, are of increasing size (in years), defined starting from the most recent observation in the data set.

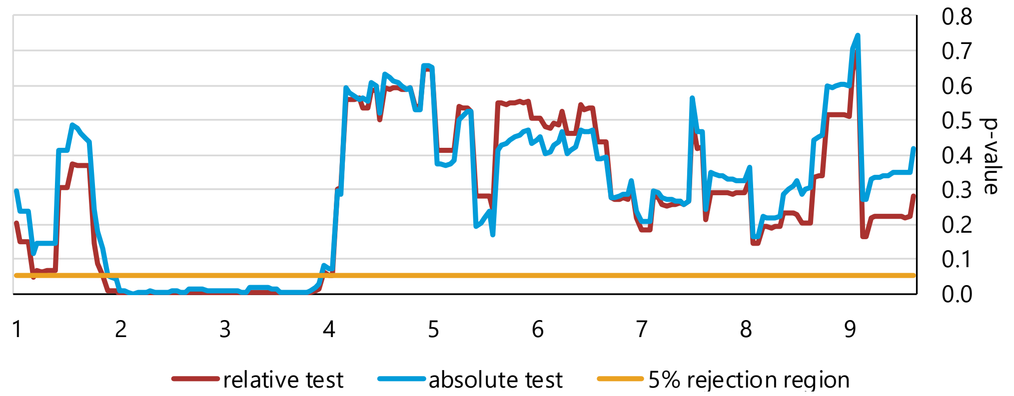
Source: Bank of Italy.

The AS test results over rolling windows (spanning 250 trading days) are presented in Graph 6, which sheds light on the significant drop in p-values observed in the right-hand side of Graph 5. Specifically, the rolling windows p-values approach zero in the period from the beginning of May 2013 to that of the following July.[15]

USD portfolio AS test results

Rolling backtesting period

Graph 6



The figure shows the results of relative and absolute AS tests; the backtesting periods consist of rolling observation windows of constant width (250 days), starting at different dates indicated on the horizontal axis.

Source: Bank of Italy.

15 During this sample period, data and news began to circulate about the possible shift of the Federal Reserve towards a restrictive monetary policy stance; this led to sudden increases in Treasury yields on certain dates and consequently to significant violations of the VaR, concentrated in a short period of time.

As previously explained, the AS test is "sharp" at first order, which enables the assessment of the impact of forecast errors. Such errors can be evaluated, on average, using the concept of realised ES (2) and/or the scaling factor $\widehat{\varphi_{ES}}$ (6). However, to analyse the effect of daily forecast errors on p-values, we found that $Z^{rel}_{ES_\alpha}$ could be expressed as follows:

$$Z^{rel}_{ES_\alpha} = \frac{\alpha \cdot (e-v) - (x+v)_-}{\alpha \cdot e} \qquad (7)$$

Given that $\alpha = 0.05$, conditional on VaR violations, the term $\alpha \cdot (e - v)$ is small compared with $(x + v)_-$. Therefore, $Z^{rel}_{ES_\alpha}$ can be approximated as:

$$Z^{rel}_{ES_\alpha} \cong -\frac{(x+v)_-}{\alpha \cdot e} \quad (8)$$

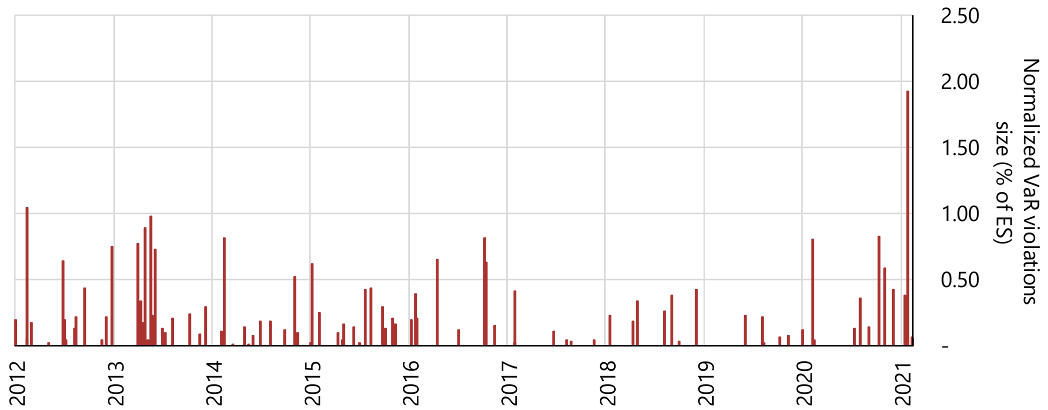
which implies:

$$-\alpha \cdot Z^{rel}_{ES_\alpha} \cong \frac{(x+v)_-}{e} \qquad (8')$$

Thus, we observed that $Z^{rel}_{ES_\alpha}$ (resp. $-\alpha \cdot Z^{rel}_{ES_\alpha}$) tends to be negative (resp. positive) conditional on VaR violations, which contributes to rejecting the null hypothesis. The normalised size of VaR forecasts violations, $\frac{(x+v)_-}{e}$ (ie the size of the VaR forecast violation as a percentage of the ES forecast), can be considered a proxy of daily relative errors on ES forecasts and used to analyse the effect of such errors on p-values. We calculated these sizes for each day in the data set.

Graph 7 illustrates the normalised violations of VaR forecasts for the USD portfolio. We observe that the low p-values reported for the 2021 period in Graph 6 are due primarily to the large violation occurred in February 2021,[16] which was seven times greater than the average.

USD portfolio normalised VaR violations — Graph 7



The figure shows the size of daily VaR violations (difference between the observed return and the VaR forecast) expressed as a percentage of the ES forecasts for the same date.

Source: Bank of Italy.

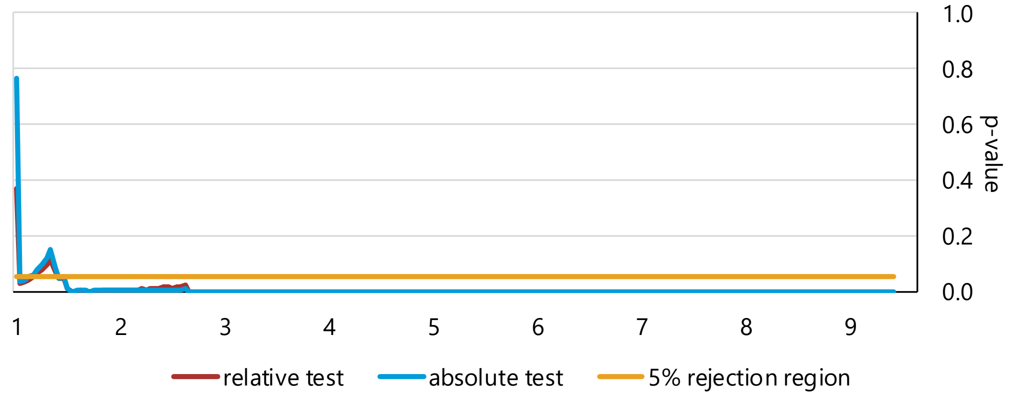
[16] This violation could be related to the 24 February FOMC meeting, whose monetary policy decision surprised markets that were expecting an interest rates increase.

Furthermore, the critical period of 2013 accounts for four out of the eight violations in the sample period that exceed a size of 0.7. The observed frequency of VaR forecasts violations in the entire data set is 0.0405, which is below the expected value of 0.05. On average, these violations have a size of 0.28.

In contrast, when considering the **JPY portfolio** and using the RiskMetrics™ model to calculate ES forecasts, the results show **unsatisfactory "average performance"** as the size of the backtesting period increases (Graph 8). For almost every backtesting period longer than one year, p-values are in fact close to zero.

JPY portfolio AS test results
Increasing backtesting period

Graph 8



The figure shows the results of the relative and absolute AS tests; the backtesting periods, indicated on the horizontal axis, are of increasing size, defined starting from the most recent observation in the data set (29/3/2021).

Source: Bank of Italy.

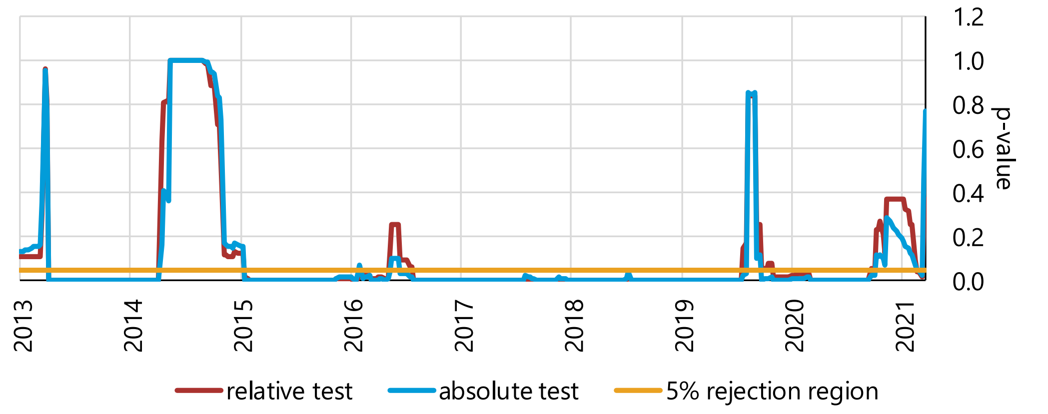
Graph 3, which presents the VaR binomial test on increasing time windows, indicates that the JPY portfolio's number of VaR violations aligns with its expected value. On average, the model correctly estimates portfolio volatility and VaR. However, the AS tests on rolling time windows are not rejected when the number of VaR violations is very close to or smaller than the lower bound of the binomial test (as shown in Graphs 3 and 9). During these periods, VaR violations are much lower than their expected value (which is 12.5 violations for 250 observations). This implies that the model overestimates the VaR as well as the ES (which is a consequence of overestimating the VaR).

In contrast, when the number of VaR violations exceeds the upper limit of the binomial test (backtesting periods starting between April and September 2015) the p-value of the AS test is very close to zero, as shown in Graph 9; even when the number of violations is close to its expected value (as depicted in Graph 3) the AS test still yields a low p-value.

JPY portfolio AS test results

Rolling backtesting period

Graph 9



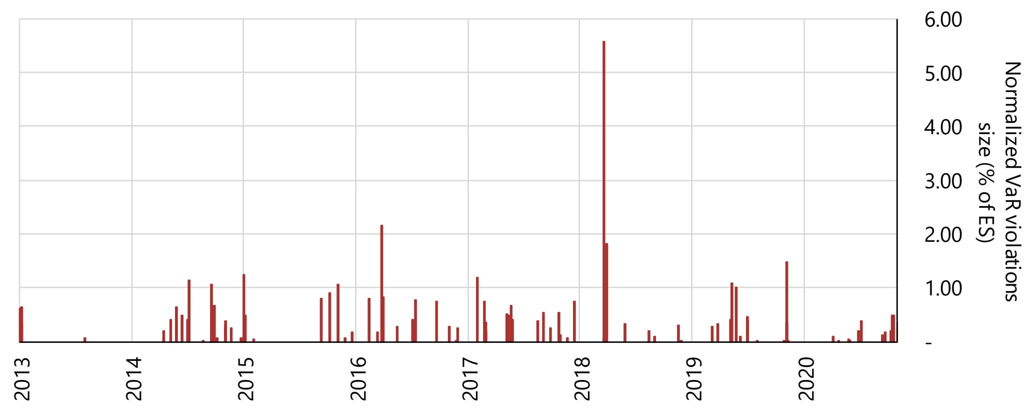
The figure shows the results of relative and absolute AS tests; the backtesting periods consist of rolling observation windows of constant width (250 days), starting at different dates indicated on the horizontal axis.

Source: Bank of Italy.

These observations, combined with the high kurtosis of the observed portfolio returns, suggest that **the AS test tends to reject the null hypothesis for the JPY portfolio**, despite the VaR forecasts being generally accurate. This is because **the model's conditional distribution fails to predict extreme returns**, resulting in a negative difference between the average forecast of the ES and the "realised" ES, which leads to the rejection of the ES model.

JPY portfolio normalised VaR violations

Graph 10



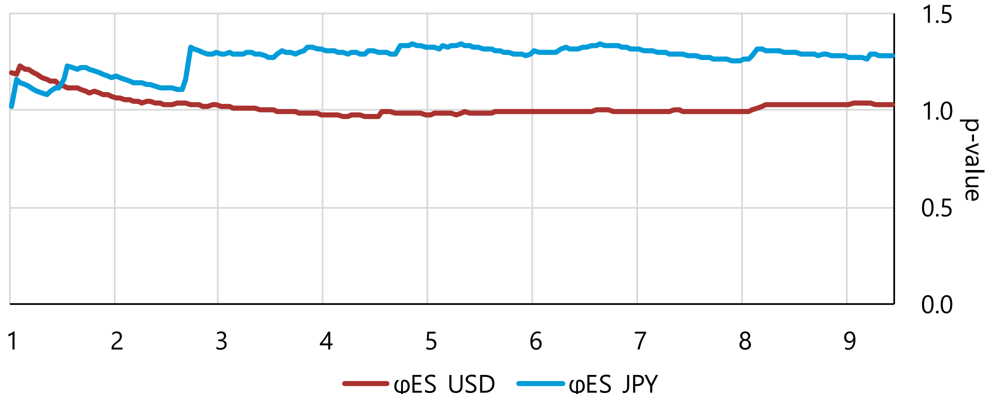
The figure shows the size of daily VaR violations (difference between the observed return and the VaR forecast) expressed as a % of the ES forecasts for the same date.

Source: Bank of Italy.

Graph 10 presents data which confirm that the frequency of VaR violations over the entire historical series is only slightly lower than expected (0.0473 versus 0.05). Additionally, for the JPY portfolio, normalised VaR violations are on average larger than for USD and show very high spikes, reaching values greater than 5.

Finally, according to Acerbi and Szekely (2019), bias effects become significant when $\widehat{\varphi_{ES}}$ deviates from 1 by more than approximately ± 60%. Graph 11 displays $\widehat{\varphi_{ES}}$ values for USD and JPY portfolios, which indicate that it stays within the threshold recommended by the authors. Moreover, the $\widehat{\varphi_{ES}}$ for USD data is very close to 1 on average, while for the JPY portfolio, it is consistently greater than 1 (with an average value of 1.27); this implies that the average ES forecasts are lower than the realised ES for this currency.

$\widehat{\varphi_{ES}}$ for USD and JPY portfolios Graph 11



The figure shows $\widehat{\varphi_{ES}}$ values for both USD and JPY portfolios, calculated for backtesting periods of increasing size (in years).

Source: Bank of Italy.

Conclusions

This paper presents a strategy for implementing effective backtesting of the risk measures most commonly adopted for managing financial portfolios in central banks.

We used well established techniques to validate VaR forecasts while, for ES, we focused on the innovative method proposed by Acerbi and Szekely. The latter provides an absolute type validation procedure for ES models, despite the fact that this measure is not strictly backtestable. All techniques were applied to the daily time series of risk measures for two foreign reserves portfolios of the BdI.

Results show that VaR forecasts are well specified, although doubts remain about the independence of the related violations, which is a desirable feature of a sound financial risk models. Regarding ES forecasts, the presence of inadequately captured fat tails leads to the rejection of the model, which often tends to underestimate risk. Furthermore, implementing the Acerbi-Szekely test on these real data has shown its peculiar characteristics: the Acerbi-Szekely test is a conservative framework that provides robustness to the validation process and is particularly suitable for risk-averse institutions quantifying discrepancies between prediction and actual values.

The empirical exercise has shown that the ridge backtesting could represent a powerful analysis tool for a central bank too. It is useful for a deeper assessment of the reliability of actual models, model selection when considering alternative models or, at least, for a more aware management of model risk. As regards avenues for further research, efforts should be aimed at providing solutions to overcome the problems encountered, by extending, for instance, the backtesting activity towards non-parametric models. In theoretical terms, the challenge will be defining a coherent power analysis framework for the Acerbi-Szekely test.

References

Acerbi, C and B Szekely (2014): "Backtesting expected shortfall", MSCI Inc, December.

——— (2017): "General properties of backtestable statistics", MSCI Inc, January.

——— (2019): "The minimally biased backtest for ES", *Risk*, September.

Acerbi, C and D Tasche (2002): "Expected Shortfall: A Natural Coherent Alternative to Value at Risk", *Economic Notes*, vol 31, issue 2, pp 379–88.

Artzner, P, D Heath, F Delbaen and J Eber (1997): "Thinking coherently" *Risk*, vol 10, pp 68–71.

——— (1999): "Coherent measures of risk", *Mathematical Finance*, vol 9, pp 203–28.

Christoffersen, P (1998): "Evaluating interval forecasts", *International Economic Review*, vol 39, no 4, pp 841–62.

Diebold, F and R Mariano (1995): "Comparing predictive accuracy", *Journal of Business & Economic Statistics*, vol 13, no 3, pp 253–63.

Fissler, T and J Ziegel (2016): "Higher order elicitability and Osband's principle", *The Annals of Statistics*, August, vol 44, no 4, August, pp 1680–707.

Gneiting, T "(2011): "Making and evaluating point forecasts", *Journal of the American Statistical Association*, vol 106, pp 746–62.

Haas, M (2001): "New methods in backtesting", Financial Engineering, Research Center Caesar, Bonn.

Hendricks, D (1996): "Evaluation of value-at-risk models using historical data", *Economic Policy Review*, pp 39–69.

Kupiec, P (1995): "Techniques for verifying the accuracy of risk management models", *Journal of Derivatives*, vol 3, pp 73–84.

Nieppola, O (2009): "Backtesting value-at-risk models", master's thesis, Helsinki School of Economics.

Machine learning applied to active fixed income portfolio management: a Lasso logit approach[1]

Mercedes de Luis[2], Emilio Rodriguez[3] and Diego Torres[4]

Abstract

The use of quantitative methods constitutes a standard component of the institutional investor's portfolio management toolkit. In the last decade, several empirical studies have used probabilistic or classification models to predict stock market excess returns, to model bond ratings and default probabilities, or for yield curve forecasting. To the authors' knowledge, there is little previous research on their application to active fixed income management. This paper contributes to filling this gap by comparing a machine learning algorithm, the Lasso logit regression, with a passive investment strategy (buy and hold) in the construction of a duration management model for high-grade bond portfolios, specifically focusing on US Treasuries. Additionally, we propose a two-step procedure, together with a simple ensemble averaging aimed at minimising the potential overfitting of traditional machine learning algorithms. We also introduce a method to select thresholds that translate probabilities into signals based on conditional probability distributions. A large set of financial and economic variables serves as an input to obtain a signal for active duration management relative to a passive benchmark portfolio. As a first result, the model selects variables related to financial flows and economic fundamentals, but the parameters appear to be unstable over time, suggesting that the variable relevance may be dependent on the timing. Backtesting of the model, conducted on a sovereign bond portfolio denominated in US dollars, yields a small but statistically significant outperformance over the benchmark index in the out-of-sample data set while controlling for overfitting. These results support the case for incorporating quantitative tools in the active portfolio management process for institutional investors while paying special attention to potential overfitting and unstable parameters. Quantitative tools should be considered as a complementary input to the qualitative and fundamental analysis, in conjunction with the portfolio manager's expertise, to facilitate better informed investment decisions.

JEL classifications(s): C45, C51, C53, E37, G11.

1 We thank all participants at the internal seminar of the Bank of Spain for their comments, and we also thank Mario Bajo from the Bank of Spain and Carmen Herrero from the World Bank for their comments and suggestions. The views expressed in this paper are our own and do not necessarily reflect the views of the Bank of Spain or the European System of Central Banks (ESCB).

2 mercedes.luis@bde.es

3 e.rodriguez@bde.es

4 diego.torres@bde.es

1. Introduction and literature review

In the last decade, a vast empirical econometric literature has been devoted to the prediction of financial market variables using classification-based qualitative models combined with machine learning (ML) techniques. Most of the studies have been applied to predict the direction of excess returns in the stock market and, to a lesser extent, in the FX markets. In the equity space, for example, Nyberg (2011) uses a dynamic error correction probit model, incorporating a binary recession indicator, for the prediction of S&P excess returns, finding better sign predictions and higher investment returns than in previous probit and ARMAX (autoregressive moving average with exogenous inputs) models.

Kara et al (2011) apply two models based on ML techniques (artificial neural networks – ANNs – and support vector machines – SVMs) to the prediction of daily directional movements in the Istanbul Stock Exchange National 100 Index, showing superior experimental performance of the first class of models. This result contrasts with those generated by Kumar and Thenmozhi (2006), who try to predict the direction of the S&P CNX NIFTY market index using several ML tools, resulting in a superior performance of SVMs compared to random forest, ANNs and other traditional discriminant analysis and logit models. Rapach et al (2013) apply the least absolute shrinkage and selection operator (Lasso) method to predict global equity market returns using lagged returns in different countries. Nasekin (2013) uses adaptive Lasso quantile regression in an empirical application designed as a "Lasso quantile trading (hedging) strategy" in comparison to other strategies related to the S&P 500 index.

Other authors use hybrid approaches to combine the strengths of parametric (logistic regressions) and non-parametric models or tree-based models (such as classification and regression trees – CART). This is the case in Zhu et al (2011), which applies these models to North American stock selection of defensive companies. In the same vein, Zaidi and Amirat (2016) combine logistic regression and artificial neural networks to predict stock market trends in Saudi Arabia. Additionally, some researchers apply ML to the design of trading strategies in the stock market, such as Beaudan and He (2019), who use a logistic regression algorithm to build a time series dual momentum trading strategy on the S&P 500 index with successful risk-adjusted overperformance. Another application can be found in Roy et al (2015), where a Lasso method based on a linear regression model is proposed as a method to predict stock market behaviour. Finally, Gu et al (2020) perform a comparative analysis of ML methods for measuring asset risk premia, identifying neural networks and regression trees as the best-performing tools for predicting stock returns.

Regarding the FX markets, literature is less abundant compared to the stock market, but a good example can be found in Sermpinis et al (2012), where the authors investigate the use of different ML methods, mainly neural networks, for forecasting and trading the EUR/USD exchange rate, finding significant outperformance evidence.

With regard to the application of ML techniques to fixed income markets, a limited amount of research has been conducted, most of which has primarily been focused on the modelling and prediction of yield curves. Some examples can be found in Castellani and Santos (2006), who do not find significant outperformance of data-driven artificial intelligence approaches in building reliable predictions for US 10-year Treasury bonds, and Dunis and Morrison (2007), who find mixed evidence for the advanced time series models compared to more traditional ones. Nunes et al (2018) apply several ANN models for forecasting the main benchmarks of the European yield curve, concluding that, in general, neural network models tend to improve results and comparing favourably to Dunis and Morrison's results, in spite of the different data set used.

Another example of yield curve forecasting with neural network models can be found in Rosadi et al (2011), where no outperformance is observed for neural network models in the prediction accuracy of the yield curve compared to more traditional methods, such as Nelson-Siegel or vector auto-regression (VAR), at least for long-term bonds. This result contrasts with those in Sambasivan and Das (2017), who, applying a Gaussian process to model the yield curve, find superior performance in forecasting the yields in the medium- and long-term segments of the yield curve.

With regard to empirical analysis using classification-based qualitative models, it has predominantly been devoted to modelling bond ratings or predicting bond defaults. Some examples can be found in Westgaard and Wijst (2001) for default rates estimation of a retail bank portfolio, or Bandyopadhyay (2006), where traditional Z-score discriminant analysis is complemented with logistic regression analysis to achieve a more accurate default prediction.

Nevertheless, to the best of the authors' knowledge, few research examples can be found on the application of classification models to the active management of bond portfolios. In Larsen and Wozniak (1995), regression models are applied for market timing in active portfolio management of different combinations of stocks, bonds and cash, finding superior performance over passive fixed-weight strategies. Berardi et al (2004) estimate a logistic econometric model for forecasting default probabilities of US dollar-denominated emerging market bonds. They construct a naïve trading strategy based on the signals of the out-of-sample forecasts of the logit model, which obtain risk-adjusted returns outperforming those derived from a buy and hold indexed strategy. As an example of the application of ML techniques, Pollege and Posch (2013) use a Lasso algorithm to find the optimal set of explanatory variables in the design of an arbitrage strategy to benefit from the non-zero basis between European sovereign credit default swaps (CDS) and cash bonds.

Despite the significant attention given to ML techniques by academia, their adoption in the asset management industry has not been as widespread as in other sectors. The performance of active exchange-traded funds (ETFs) using ML in their investment strategies tends to be mixed, as shown by Bartram et al (2021). López De Prado (2018) concludes that these mixed results are mainly due to the fact that financial data sets violate standard assumptions of ML applications, including stationarity, independence, data labelling and sampling. In this study, we have carefully considered these factors and accounted for the unique statistical properties of the series when selecting the variable groups, as detailed in Section 2.2.

2. Methodology

The present study adopts a simple approach for modelling the future performance of a fixed income portfolio, assuming that its expected market value can be explained by a set of potential variables. Following Nyberg (2011), the goal is to predict the future direction, not the level, of the fixed income portfolio market value[5] (let $y_t^* = 1$ if we observe a positive total return, ie if $I_t - I_{t-1} > 0$ where I_t is the index value at time t; $y_t^* = 0$, otherwise). The logistic Lasso approach is applied to handle the high number of predictors.[6] This approach predicts y^* conditioned on a set of k explanatory variables $[\boldsymbol{x}]$ as reflected in equation (1):

$$E(y^*/_x) \quad = g(\boldsymbol{x\beta}, \varepsilon) \tag{1}$$

assuming $g(\boldsymbol{x\beta}, \varepsilon)$ to have the same structure as a traditional logit regression but with a penalised version of the log-likelihood function. A simple logistic Lasso is selected because, given the literature in other areas different from fixed income asset management, it is not clear that complex models, such as XGBoost or neural networks, are more accurate than simpler ones. For example, Palomares-Salas et al (2009) found that autoregressive integrated moving average (ARIMA) models outperformed neural networks for short-term wind speed forecasting in terms of lower root mean square errors (RMSE), while Rahman et al (2022) found that an ARIMA model performed better than an XGBoost model for predicting Covid-19 in Bangladesh. However, different results were achieved in the Fang et al (2022) study for the United States.

2.1. Hyperparameter, lambda or regularisation parameter

The penalising component included in the definition of equation (1) is the sum of the absolute value of the k parameters incorporated in the model scaled by a hyperparameter[7] λ such that the final log-likelihood is given by equation (2):

$$L(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left[y_i x_i \boldsymbol{\beta} - \boldsymbol{log}\left(\boldsymbol{1} + \boldsymbol{e}^{x_i\boldsymbol{\beta}}\right)\right] + \lambda \sum_{j=1}^{k}\left|\beta_J\right| \tag{2}$$

The penalty used in Lasso logit regression works as a variable selection and shrinkage procedure: when λ is sufficiently large it forces some of the coefficient estimates to be exactly equal to zero. From a Bayesian perspective, Park and Casella (2008) conclude that λ can be interpreted as the prior uncertainty of the model parameters. For example, when λ is small it could be interpreted as the true model a priori (ie the one that includes most of the variables). Usually when only a few predictors have large coefficients, one can expect Lasso to have a good performance but when all the coefficients are roughly of equal size, or when the number of predictors is much larger than the number of observations (*n*),[8] Pereira et al (2016) suggest that other regularisation techniques are more appropriate (eg ridge regressions, elastic net, etc).

[5] Portfolio is used interchangeably to refer to the benchmark index used in this paper: the Bloomberg-Barclays fixed income index for US bonds, which contains US domestic government debt with maturities higher than one year (Section 3).

[6] We opted for Lasso over ridge and stepwise regression primarily due to our data set's high multicollinearity. Lasso handles this more effectively, selecting and regularising variables simultaneously. In contrast, stepwise regression's sequential approach may lead to suboptimal results depending on the order of incorporation.

[7] In the literature, the hyperparameter is also known as the regularisation parameter or just lambda. This paper will use these terms interchangeably.

[8] This is not the case in this data set; there are 201 variables and 275 observations.

To deal with potential overfitting, which can be more severe in more complex ML algorithms such as XGBoost due to the subjectivity associated with the selection of hyperparameters, the Lasso hyperparameter λ[9] is selected using the cross-validation algorithm. Cross-validation is a resampling method that uses different portions of the data to train and test a model on different iterations. The same data that were used to fit the model are divided into K (K=10 in this study) approximately equally sized and mutually exclusive subsamples called folds. For each fold k, the model is refit on the data using 100 different λ in the other K–1 folds. Finally, λ is selected so as to minimise the cross-validation deviance[10] defined in the algorithm as minus twice the log-likelihood on the left-out data.

The estimation exercise is performed by dividing the database into two parts: a training set (in-sample) and a testing set (out-of-sample). The training set starts in January 2004 and ends recursively at the end of 2011 throughout 2020; leaving the testing sample, also recursively, from 2012 throughout 2021. For example, the first loop has a training set from 2004 to 2011[11] and leaves the 2012 for testing (out-of-sample).

2.2. The two-step procedure: an error correction approach and a simple ensemble averaging

The method proposed to overcome potential worse out-of-sample performance in the Lasso logit algorithm consists of a two-step procedure. In the first step, a Lasso logit model is estimated with a long time span (ie 10 years, in order to include around two economic cycles) that is interpreted as the long-term relationship between financial and economic variables and the portfolio performance. In the second step, another model is estimated; a simple logit regression with the error committed in t–1 and the prediction made by the Lasso logit estimated with the long time span as explanatory variables. This step is estimated for the last four years (around a standard economic cycle). If the error committed in t–1 is statistically significant, the probability eventually used will be the one obtained in the second step; otherwise, it will be the one obtained in the first step.

[9] Only one hyperparameter in the simple Lasso approach compares to more complex models that have more than one hyperparameter to choose from.

[10] Cross-validation deviance is a statistical technique commonly used in model evaluation to assess the predictive performance of a statistical or ML model. It is particularly useful when working with complex models that may have a high risk of overfitting or poor generalisation to new data. Cross-validation deviance involves dividing the available data set into multiple subsets or "folds". The model is then trained on a combination of folds and tested on the remaining fold. This process is repeated several times, with each fold serving as a testing set exactly once. The deviance, which quantifies the model's fit to the data, is calculated for each fold. By averaging the deviances across all folds, a robust estimate of the model's performance can be obtained. Cross-validation deviance provides researchers with a reliable measure of a model's ability to generalise to unseen data, enabling them to make informed decisions regarding model selection and refinement.

[11] The two-step procedure, as explained in Section 2.2, involves the division of the trained set into two subsamples. The first subsample contains no more than 10 years of data, while the second encompasses a period of four years. The length of the test sample is always fixed at one year. The rationale behind this approach is to approximate the widely accepted five-year business cycle duration reported in existing literature. By combining the subsample comprising four years of data with the one-year test data, this approximation is achieved.

Additionally, another model proposed to overcome possible overfitting or model misspecification is a simple ensemble averaging. Ensemble averaging is the process of creating multiple models and combining them to produce a desired output, as opposed to creating just one model. The ensemble of models frequently performs better than any individual model, because the various errors of the different models tend to "average out". An advanced methodology that can address misspecification and model uncertainty is the Bayesian model averaging approach, as discussed by Fragoso et al (2018). However, in the present study, a simpler approach is adopted, where three distinct models are estimated so as to tackle non-stationarity of the variables. Each model pertains to a specific type of statistical property observed in the data, namely level, first difference, and monthly growth stationary variables. The final estimation is obtained by taking a simple average of the three estimates.

2.3. Thresholds based on conditional probability distributions to translate probabilities into portfolio duration signals

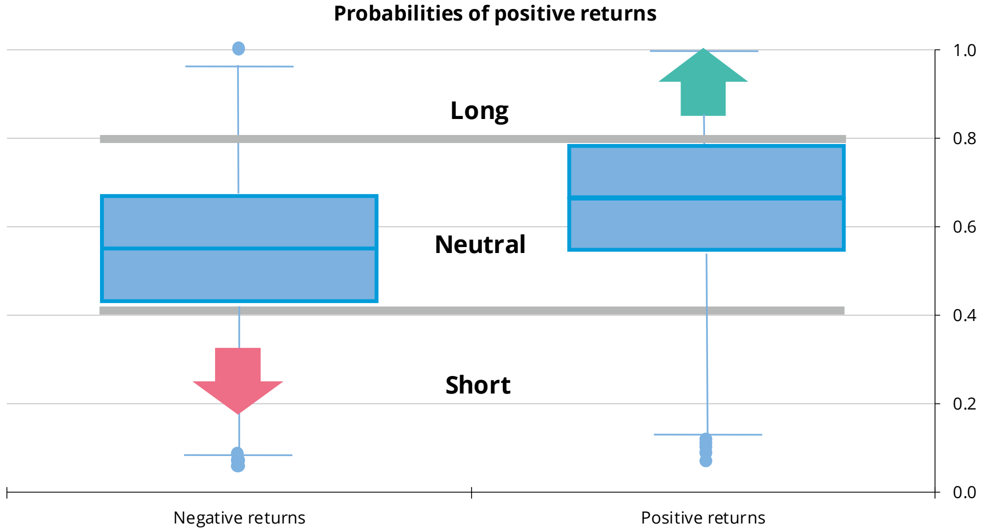
We propose an algorithm to translate probabilities into three portfolio duration signals (short, neutral and long). The algorithm checks the distribution of probabilities given by the model in the in-sample period conditioned on the observed direction of the portfolio market value (ie if it had positive or negative returns). We derive short signals from probabilities lower than the 25th percentile of the distribution when the index was down and long signals when probabilities are higher than the 75th percentile, but in this case, conditioned on cases where the index was up (Graph 1). We assign neutral signals to probabilities between those two previously defined thresholds (ie higher than the 25th percentile when the index presented negative returns and lower than the 75th percentile when positive returns were observed). The proposed mapping is also compared to a naïve threshold or "rule of thumb" (probability up to 33% short position, 33%–66% neutral and more than 66% long).

It is assumed that a short position means reducing the investment by 10%, neutral implies maintaining the previous invested amount and a long position derives in increasing the investment by 10%.[12] For instance, assuming that in $t = 0$ we invest I_0 and the return of the portfolio is r_1 in $t = 1$, then we would have $I_1 = (1 + r_1)I_0$. Assuming that in $t = 0$ the model gives a "long signal", then we invest an additional 10%, $I_{long,0} = (1 + 0.1)I_0$ and in $t = 1$ we would have $I_{long,1} = (1 + r_1)I_{long,0}$. But this equation is equivalent to $I_{long,1} = (1 + r_1)(1 + 0.1)I_0$, so that every month, we would scale the returns of the portfolio to obtain the returns of the model depending on the signal:

Long → 1.1 * $(1 + r_t)$
Neutral → $(1 + r_t)$
Short → 0.9 * $(1 + r_t)$

[12] The percentage of investment withdrawal or increase is selected to be 10% to match standard constraints in active portfolio management, but is proposed to be a topic for further research.

Thresholds based on conditional probability distributions | Graph 1



Source: Authors' elaboration.

2.4. Model-based strategies compared to passive investment

We compare five models to a passive investment:

1. **LassoDefault**: A Lasso logit that selects the hyperparameter applying cross-validation as defined in Section 2.1.
2. **LassoDefaultTwoStep**: A Lasso logit that uses the two-step procedure defined in Section 2.2.
3. **LassoSimpleEnsemble**: A Lasso logit for every group of variables as defined in Section 2.2 (we estimate three models according to different data features: a first one in levels, a second using first differences, and a third using monthly growth stationary variables).
4. **LassoSimpleEnsembleTwoStep**: The same as in the LassoSimpleEnsemble model (3) but applying the two-step procedure defined in Section 2.2.
5. **Always long strategy**: A strategy that always invests 10% more than the passive investment.

These models are compared to a passive investment algorithm. The passive investment is the trading strategy that invests in the benchmark or tries to replicate its total return while minimising the tracking error. In this study passive investment will be referred to as being neutral or investing in the same constituents and same weights as the benchmark.

3. Data description

A large set of financial and economic indicators are used as input information to obtain a signal for active duration management relative to a passive benchmark portfolio. The series used in the estimation of this model are listed in the Annex (Table A1). The selection of these indicators was driven by a focus on maximising the available variables without adhering to a specific rule or hypothesis testing. Our aim was to include as many variables as possible to capture a broad range of market dynamics and information. We start with 250 indicators, almost half of which are macroeconomic series, 18% financial data and 9% fixed income market variables (Table A2). In addition, these indicators include mixed frequencies, 107 of which are updated on a monthly basis, 85 daily and 24 quarterly (Table A3).

The economic and financial variables should start from 2004, on a monthly basis, in order to encompass a 10-year out-of-sample period so as to include the Great Financial Crisis (GFC). Only 201 out of 250 analysed variables fulfil the requirement imposed by this study.[13]

As regards the missing values, an imputation method is applied, in which the last non-missing observed value is used to assign a particular missing value. This approach allows us to maintain the temporal structure of the data and minimise any potential biases introduced by imputation. For daily and weekly data, we take the average value of the corresponding month to ensure consistency in the frequency of the variables. It is worth noting that we have explored alternative approaches, including excluding missing values entirely, and found that the results remain unchanged. Additionally, the variables are standardised in order to ease the comparison of scores measured on different scales. It is important to note that only the in-sample values are standardised, not the out-of-sample data.

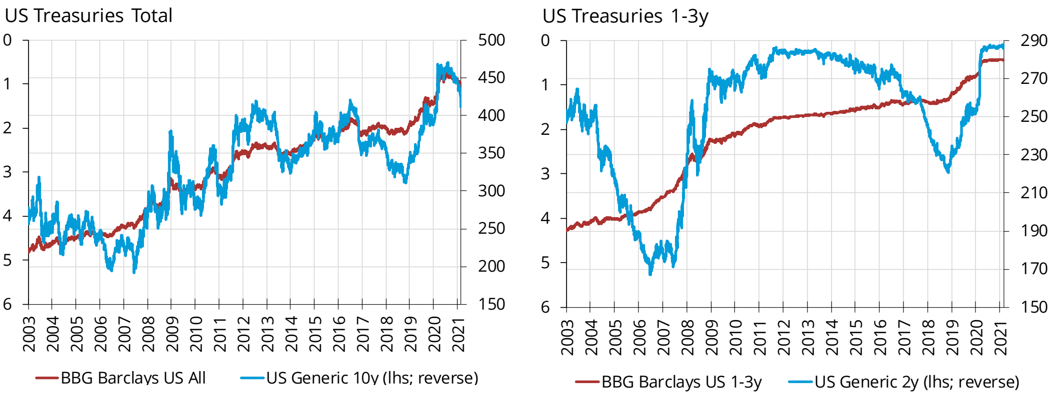
Before using the data as input for the model an augmented Dickey-Fuller test is run for every variable to check if they are stationary. If they are not stationary, a first difference or a percentage change transformation is applied. Finally, variables are allocated to three different groups: 1) stationary variables in levels without transformation; 2) first difference of stationary variables; and 3) percentage change of stationary variables. These three groups are going to be used for simple ensemble averaging.

With respect to the dependent variable, the Bloomberg Barclays USD Treasuries Total Return Index is used, which includes all domestic government debt with maturities higher than one year. This total return index has been selected for two reasons. First, it represents the whole yield curve and has a higher correlation with the 10-year US Generic Government Bond than with the two-year (see Graph left- and right-hand panels). Second, it is transparent in its constituents and their weights.

13 From a theoretical standpoint, we anticipate that the state of the macroeconomy can impact interest rates and inflation pressures. When the macroeconomy performs well, inflation pressures may arise, potentially leading to an increase in interest rates. Additionally, financial variables such as interest rates in other countries, like Japan or the euro zone, may influence the flow of capital and impact the US interest rate. If interest rates increase in Japan, for instance, capital may flow to Japan, exerting upward pressure on US interest rates to prevent outflows. These theoretical expectations underpin our inclusion of certain indicators and provide a framework for interpreting the results in the context of market dynamics and interdependencies.

Benchmark indices and generic Treasury bonds

Graph 2



Source: Authors' elaboration based on Bloomberg data.

Table 1 shows time-lagged correlations between monthly growth rates (month-over-month, MoM) and the benchmark index. The financial variables are the most contemporaneously correlated, specifically fixed income series. However, economic variables increase in importance when the forecasting horizon is higher than one month. This result supports the empirical evidence found by Cerniglia and Fabozzi (2020) that the variance-covariance matrix depends on the forecast horizon analysed. In the short-term, financial variables have the greatest impact on fixed income markets, while macroeconomic variables seem to impact more in the long-run behaviour of fixed income portfolios.

Correlation of explanatory variables (month-over-month (MoM) vis-à-vis the benchmark)

Table 1.a

	MoM					
			Lags			
	0		1		2	
BBG US Ser-E Gov > 1Y Bond Index	100%	S&P 500	26%	MOVE	20%	US JPM Ts Investor Sentiment
iBoxx US Ts 7-10Y TRI	98%	DAX	24%	London Metal Exchange Index	18%	Generic Spain 30y Government Bond
iBoxx US Ts 5-7Y TRI	96%	IBEX 35	20%	BBG USDJPY 3M Hedging Cost	18%	Conference Board Consumer Conf
iBoxx US Ts 3-5Y TRI	91%	CBOE Volatility Index	18%	Langer US National Economy Exp	16%	US Capacity Utilization % of Total
iBoxx US Ts 1-3 TRI	77%	Federal Reserve Balance Sheet	17%	iBoxx Euro Spain Sovereign TRI	13%	Langer US Nat. Eco. Expect. Diffus. Index
iBoxx Euro Germany Sovereign TRI	75%	USD INFL SWAP ZC 10Y	16%	BBG Commodity	12%	ECB Survey of Professional Forecasters
iBoxx Euro Germany Covered TRI	66%	Private Housing Authorized by Bldg Permits	16%	Federal Reserve Balance Sheet	12%	Phil. Fed Survey of Professional Forecasters
iBoxx Euro Spain Sovereign TRI	41%	Generic 1st 'CO' Future	14%	iBoxx US Ts 7-10Y TRI	11%	U. of Michigan Current Eco. Conditions Inde
iBoxx EUR Spain Covered	35%	Nat. Assoc. of Home Builders Market Index	13%	EUR-JPY X-RATE	10%	Market News International Chic
Gold Spot $/Oz	26%	iBoxx US Trs 1-3 TRI	13%	iBoxx EUR Spain Covered	10%	Private Housing Authorized by Bldg Permits

Correlation of explanatory variables by categories (month-over-month (MoM) vis-à-vis the benchmark)

Table 1.b

	MoM					
			Lags			
	0		1		2	
Fixed Income	100%	Equity	26%	Uncertainty	20%	Survey
Fixed Income	98%	Equity	24%	Commodity	18%	Financials
Fixed Income	96%	Equity	20%	Financials	18%	Economics
Fixed Income	91%	Uncertainty	18%	Economics	16%	Economics
Fixed Income	77%	Monetary	17%	Fixed Income	13%	Economics
Fixed Income	75%	Financials	16%	Commodity	12%	Survey
Fixed Income	66%	Economics	16%	Monetary	12%	Survey
Fixed Income	41%	Commodity	14%	Fixed Income	11%	Economics
Fixed Income	35%	Economics	13%	Currency	10%	Economics
Commodity	26%	Fixed Income	13%	Fixed Income	10%	Economics

Source: Authors' elaboration.

4. Results

First, we compare the naive threshold or "rule of thumb" method with the proposal based on conditional probabilities. Table 2 shows the confusion matrix using the "naive" threshold, rows show the signals predicted by the model, while columns show the actual benchmark's movements (negative or positive returns). This "naive" threshold gives only one correctly identified long signal in the out-of-sample period (2012–21). The signals given by this approach are too small (one out of 120 months) compared to the proposed thresholds based on conditional probabilities (22 out of 120 months). Based on this low active ratio, this study will only use the threshold obtained by the proposed approach, discarding the "naive" threshold so that we obtain a more "active" model.

Confusion matrix

Table 2

Lasso with rule of thumb thresholds

	Loss	Gain	Total_Signals
Short	0%	0%	0%
Neutral	49%	51%	99%
Long	0%	100%	1%
Total_observed	49%	51%	100%

HIT Ratio	100%
Active	1%

Source: Authors' elaboration.

Second, we carry out performance analysis for the LassoDefault model. This model has a hit ratio[14] of 58% in the in-sample period (2004–11), but only because it correctly predicts the direction of returns when they are positive (74%). It is not able to identify the negative returns, and correctly predicts a bearish move only 43% of the time. It is very active in the in-sample period, where it gives a long/short signal 42% of the time (Table 3; first panel). In the out-of-sample period, the LassoDefault's hit ratio drops 8pp to 50%, not better than a random model, and gives signals only 27% of the time. The performance loss comes from a decrease in accuracy when the model predicts positive returns (long signals), down from 74% to only 52%. There is an increase in the accuracy of negative returns, but not enough to achieve a hit ratio higher than a random model (up to 45% from 43%; Table 3; second panel). The two-step procedure is not able to increase the performance of the model in the out-of-sample period (Table 3; third panel).

14 Hit ratio is defined as the sum of correct long and short signals given by the model divided by the number of periods in which the model delivers a signal.

Model performance

Table 3

Lasso default (CrossValidation - insample)

	Loss	Gain	Total_Signals
Short	43%	57%	22%
Neutral	40%	60%	58%
Long	26%	74%	20%
Total_observed	38%	62%	100%

HIT Ratio	58%
Active	42%

Lasso default (CrossValidation - outsample)

	Loss	Gain	Total_Signals
Short	45%	55%	9%
Neutral	50%	50%	73%
Long	48%	52%	18%
Total_observed	49%	51%	100%

HIT Ratio	50%
Active	27%

Lasso default Two Step (CrossValidation - outsample)

	Loss	Gain	Total_Signals
Short	44%	56%	8%
Neutral	50%	50%	73%
Long	48%	52%	19%
Total_observed	49%	51%	100%

HIT Ratio	50%
Active	27%

Source: Authors' elaboration.

The "poor" out-of-sample performance of the LassoDefault model could be explained by possible overfitting when choosing the lambda's value in the cross-validation exercise. To try to overcome this possible overfitting, the simple ensemble model is compared. This model has a really good performance in the in-sample period, with a hit ratio of 88%, but it gives few signals (only 27% of the time). The accuracy between negative and positive returns looks more balanced (a hit ratio of 82% when the index has losses and 93% when the returns are positive; Table 4, first panel). The simple ensemble model also behaves well in the out-of-sample period. The hit ratio decreases, but it maintains a level that is above the 50% threshold (59%). In this model, the accuracy is concentrated in the short signals, where 67% correctly identified a loss trend and 56% the positive returns. The model is slightly less active than in the in-sample period, down to 18% from 27% (Table 4, second panel). In this case, the two-step procedure is able to increase the performance slightly to achieve a hit ratio of 62%, with correct short signals 71% of the time, and correct long signals 57% of the time (Table 4, third panel).

Model performance

Table 4

Lasso Simple Ensemble (CrossValidation - insample)

	Loss	Gain	Total_Signals
Short	82%	18%	12%
Neutral	38%	62%	73%
Long	7%	93%	16%
Total_observed	38%	62%	100%

HIT Ratio	88%
Active	27%

Lasso Simple Ensemble (CrossValidation - outsample)

	Loss	Gain	Total_Signals
Short	67%	33%	5%
Neutral	49%	51%	82%
Long	44%	56%	13%
Total_observed	49%	51%	100%

HIT Ratio	59%
Active	18%

Lasso Two Step Simple Ensemble (CrossValidation - outsample)

	Loss	Gain	Total_Signals
Short	71%	29%	6%
Neutral	48%	52%	83%
Long	43%	57%	12%
Total_observed	49%	51%	100%

HIT Ratio	62%
Active	18%

Source: Authors' elaboration.

Previously it has been shown that the Lasso regression with default options is worse than the model with the simple ensemble. However, it is possible that a model with a low hit ratio may outperform one with a high hit ratio if the former produces signals whenever there are significant movements in returns. This could happen if the LassoDefault model identifies a period of extreme return movements compared to the simple ensemble (many mistakes with small losses but hitting very big and extreme returns). In Table 5, both models (with the two-step procedure applied to each) are compared to the passive investment strategy. As illustrated there, the LassoDefault model is not able to beat the passive investment strategy in the out-of-sample period either in absolute returns or in Sharpe ratio terms. The passive investment strategy also has better risk ratios than the LassoDefault model. The LassoDefault model suffers the worst drawdown across the models analysed. The two-step procedure (LassoDefaultTwo-Step) does not improve the model. The Sharpe ratio and the conditional value-at-risk at the 95% confidence level (CVaR95) remain virtually the same. The accumulated return is the only indicator that improves slightly when the two-step procedure is applied (+4 bp).

The simple ensemble model outperforms passive investment strategy on almost all indicators except for maximum drawdown, where it slightly underperforms (–7.75% versus –7.72%). However, the simple ensemble has the lowest CVaR95 and the highest excess return. When a two-step procedure is applied to the simple ensemble (LassoSimpleEnsembleTwoStep), the model's risk ratios improve, and it achieves the highest Sharpe ratio and information ratio and the lowest drawdown. Although the outperformance of the simple ensemble begins from 2011, the two-step procedure only shows significantly better performance since 2019, as indicated by Graph 4. The always long strategy has almost the same information ratio as the Lasso Simple Ensemble (Two-Step), but it has the worst Sharpe ratio and risk ratios.

Performance and risk ratios

Table 5

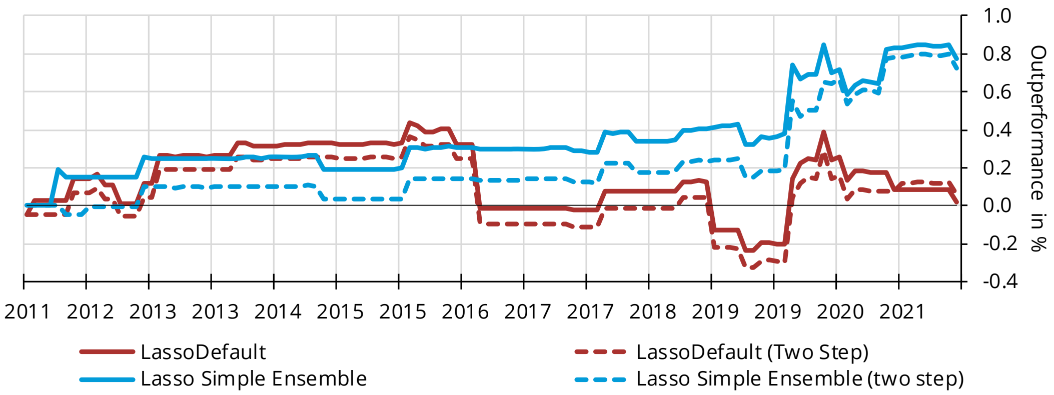
	Performance Ratios					Risk Ratios				
	Annual Returns	Sharpe Ratio	Accum. Return	Exc. Return[1]	Info. Ratio	SE	Tracking Error	MaxDD	VaR 5%	CVaR
Passive Investment	2.18	0.58	23.47	n/a	n/a	3.78	n/a	-7.72	-16.48	-23.28
Lasso Default	2.18	0.57	23.49	0.02	0.02	3.83	0.19	-7.94	-15.83	-23.67
Lasso Default (Two step)	2.19	0.57	23.53	0.06	0.04	3.83	0.19	-7.93	-15.83	-23.67
Lasso Simple Ensemble*	2.25	0.59	24.25	0.78	0.42	3.82	0.15	-7.75	-16.44	-22.97
Lasso Simple Ensemble (Two Step)*	2.24	0.59	24.20	0.73	0.44	3.81	0.14	-7.60	-16.44	-22.97
Always Long	2.60	0.56	26.00	2.53	0.44	4.67	0.96	-8.69	-19.53	-28.89

SE = statistical error, MaxDD = maximum drawdown, VaR = value-at-risk and CVaR = conditional value-at-risk. Note: * p < 0.1. Period 2012-2021. [1] Excess returns compared to Passive Investment's cumul. returns.

Source: Authors' elaboration.

Outperformance

Graph 4



Source: Authors' elaboration.

The alpha generated by the best model (Lasso Simple Ensemble Two-Step) is statistically significant at 10%, but the most interesting result is that most of the alpha generated comes from correctly identifying "extreme" movements (return movements higher than 1.3 standard deviations; Table 6).

Performance and risk, Lasso Simple Ensemble Two-Step: alpha

Table 6

	alpha (%)[1]	p-value	n
Whole sample	0.06	0.09	120
returns > 1.28σ	**0.42**	**0.05**	**18**
returns > 1.64σ	0.31	0.17	11
returns > 1.96σ	0.86	0.20	4

alpha (%)[1]	coefficient	Std. Err.	p-value
percentile 80 (> 1.28σ)	0.42	0.37	0.002
constant	0.00	0.01	0.979

[1] Annualised figures.

Source: Authors' elaboration.

Table 7 shows the set of variables that the ML algorithm is selecting every year, jointly with their betas. Only two variables appear recurrently in the whole out-of-sample exercise, one related to economics (US import prices) and another to financial flows (Japanese two-year government bond). US capacity utilisation appears nine out of 10 years, jointly with the seven-year Japanese bond and the euro swap overnight index rate (OIS) for one week. Remarkably, the betas are not stable year-over-year. For instance, it seems that at the end of the out-of-sample period, US capacity utilisation is losing forecasting power compared to US 1y1y inflation forward rate.

Set of variables selected by year

Table 7

Variable	Type	Year incl.	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Mean beta
US Capacity Utilization	ECON	9		0,38	0,42	0,31	0,42	0,38	0,37	0,27	0,19	0,07	0,31
Unemployment	ECON	1									-0,27		-0,27
Japan (7 year Issue)	FINANCIAL	9		-0,23	-0,33	-0,25	-0,25	-0,29	-0,25	-0,28	-0,26	-0,17	-0,26
Dax Index	FINANCIAL	7				-0,02	-0,18	-0,28	-0,27	-0,23	-0,30	-0,26	-0,22
Japan (2 year Issue)	FINANCIAL	10	-0,05	-0,24	-0,22	-0,24	-0,26	-0,34	-0,23	-0,28	-0,24	-0,10	-0,22
Home Builders Market Index	ECON	8		-0,26	-0,32	-0,14	-0,14	-0,10	-0,10	-0,12	-0,17		-0,17
JPMorgan Investor Sentiment Survey Active (long)	SURVEY	4		-0,05	-0,14	-0,17	-0,24						-0,15
VIX	FINANCIAL	1										0,13	0,13
Euro Swap 1 week	FINANCIAL	9	0,02	0,13	0,15	0,17	0,18	0,20	0,10	0,14	0,11		0,13
Adjusted Retail Sales	ECON	6		-0,12	-0,18	-0,20	-0,13	-0,12	-0,04				-0,13
US Import Prices	ECON	10	-0,19	-0,13	-0,06	-0,10	-0,21	-0,09	-0,12	-0,15	-0,07	-0,08	-0,12
5-year US Open Interest	FINANCIAL	8		-0,22	-0,16	-0,16	-0,17	-0,08	-0,09	0,00	-0,02		-0,11
Japan (5 year Issue)	FINANCIAL	3	-0,32	-0,01							-0,01		-0,11
US Retail Sales	ECON	2				-0,01	-0,20						-0,11
Economic Condition Michigan	SURVEY	2		-0,10	-0,04								-0,07
OIL Open Interest	COMMODITIES	2				0,00	0,13						0,07
US Export Prices	ECON	5			-0,06	-0,06	-0,08	-0,10		-0,02			-0,06
US Inflation Forward Rate 1y1y	FINANCIAL	8		0,01	0,00	0,02	0,09	0,10	0,02	0,06	0,09		0,05
PMI Services	SURVEY	4					-0,01		-0,03	-0,11	-0,04		-0,04
US Manufacturers New Orders	ECON	4			-0,04	-0,11			-0,01	-0,02			-0,04
JPMorgan Investor Sentiment Survey All (long)	SURVEY	4		-0,03	-0,02	-0,04	-0,08						-0,04
Eurostoxx Implied Volatility	FINANCIAL	1										0,03	0,03
US 3 month	FINANCIAL	1				-0,03							-0,03
Global Implied Volatility	FINANCIAL	3							0,02	0,05	0,01		0,03
OIL	COMMODITIES	1									-0,02		-0,02
US Industrial Production	ECON	1					0,02						0,02
Man. Activity (Kansas)	ECON	3			-0,01	-0,01	-0,01						-0,01

Source: Authors' elaboration.

This result is in line with the increase in inflation uncertainty observed in 2021 when Covid-19 measures started to be loosened and demand began to lift out.

Conclusions and further research

This paper tries to fill the gap related to the application of machine learning in the context of active fixed income management. It compares the performance of a machine learning (ML) algorithm, "the Lasso logit regression", with a passive investment strategy and proposes a simple ensemble alternative and a two-step model to reduce overfitting problems. It also presents an algorithm to select thresholds that map probabilities into signals based on conditional probability distributions.

The algorithm proposed to translate probabilities into signals is more active than the "rule of thumb" alternative and performs better. That is, choosing as the higher threshold (long signals) the 75th percentile of the distribution of probabilities given that the benchmark has positive returns, and as the lower threshold (short signals) the 25th percentile of the distribution of probabilities given that the index has negative returns, using only the in-sample data set. The probabilities between the higher and lower thresholds are assigned as "neutral".

The ML algorithm that only applies the Lasso logit regression with default options is not able to beat the passive investment strategy. Even applying the two-step procedure, the performance is not increased. The algorithm that seems to work well is the simple ensemble alternative, which achieves the best risk and return ratios. This algorithm splits the data set into three different sets of variables, based on their statistical properties (being stationary or not), and then a Lasso logit regression is applied to every set. The two-step procedure applied to the simple ensemble improves the risk ratios of the model, achieving the highest Sharpe ratio and information ratio and the lowest maximum drawdown.

The variables selected by the ML algorithm behave as expected a priori. For the evolution of our monthly fixed income portfolio, economic variables and financial flows are the most relevant. For most of the years the following variables are selected: US capacity utilisation, Japanese bonds, import prices and euro swap OIS. But a signal of caution is observed because the relevance of the variables is not stable and changes over time. Nevertheless, this makes sense. As an example, inflation expectations increased their forecasting power in 2021 compared to US capacity utilisation, something that one should expect because of the increasing inflation uncertainty post Covid-19 in 2021.

The alpha generated by the Lasso Simple Ensemble model after applying the two-step procedure is positive and statistically significant at 10%, but the most interesting result is that most of the alpha comes from correctly identifying "extreme" movements (returns movements higher than 1.3 standard deviations).

These results provide evidence to support the advantages of incorporating quantitative tools in the active portfolio management process for institutional investors but taking into account that some overfitting could occur. All in all, ML algorithms should be applied as a complementary input to the qualitative or fundamental analysis together with the portfolio manager's expertise, in order to make better informed investment decisions.

There are some limitations that could be explored in further research. First of all, the amount of money invested (divested) when there is a long (short) signal (+10%/–10%) could be tied to the probabilities, maybe applying the Kelly criterion, in order to find if there are some improvements compared to the fixed 10% approach. The lack of stability in the parameters of the model could be an additional line of research, including some feature selection algorithms like Bayesian model averaging. Another extension could be to include as inputs for the model some technical indicators that are widely used in investment decisions, like Bollinger Bands, relative strength index (RSI), moving average oscillator, or Ichimoku, among others. It could also be interesting to test other types of ML algorithms like XGBoost, among others, to investigate if they are less prone to overfitting issues.

References

Bajo, M and E Rodríguez (2011): "Gestión activa de una cartera de bonos: un modelo cuantitativo de duración", *Análisis Financiero*, no 115, pp 72–89.

Bandyopadhyay, A (2006): "Predicting probability of default of Indian corporate bonds: logistic and Z-score model approaches", *Journal of Risk Finance*, vol 7, no 3, pp 255–72, https://doi.org/10.1108/15265940610664942.

Bartram, S, J Branke, G De Rossi and M Motahari (2021): "Machine learning for active portfolio management", *The Journal of Financial Data Science*, vol 3, no 3, pp 9–30, https://doi.org/10.3905/jfds.2021.1.071.

Basak, S, S Kar, S Saha, L Khaidem and S Dey (2019): "Predicting the direction of stock market prices using tree-based classifiers", *The North American Journal of Economics and Finance*, vol 47, pp 552–67.

Beaudan, P and S He (2019): *Applying machine learning to trading strategies: using logistic regression to build momentum-based trading strategies*, available at SSRN 3325656.

Berardi, A, S Ciraolo and M Trova (2004): "Predicting default probabilities and implementing trading strategies for emerging markets bond portfolios", *Emerging Markets Review*, vol 5, no 4, pp 447–69, https://doi.org/10.1016/j.ememar.2004.05.004.

Castellani, M and E Santos (2006): "Forecasting long-term government bond yields: an application of statistical and AI models", *Working Papers Department of Economics* 2006/04, ISEG - Lisbon School of Economics and Management, Department of Economics, Universidade de Lisboa.

Cerniglia, J and F Fabozzi (2020): "Selecting computational models for asset management: financial econometrics versus machine learning – is there a conflict?", *The Journal of Portfolio Management*, vol 47, no 1, pp 107–18, https://doi.org/10.3905/jpm.2020.1.184.

Clewell, D, C Faulkner-Macdonagh, D Giroux, S Page and C Shriver (2017): "Macroeconomic dashboards for tactical asset allocation", *The Journal of Portfolio Management*, vol 44, no 2, pp 50–61, https://doi.org/10.3905/jpm.2018.44.2.050.

Colianni, S, S Rosales and M Signorotti (2015): *Algorithmic trading of cryptocurrency based on Twitter sentiment analysis*, CS229 Project, 1-5.

Dunis, C and V Morrison (2007): "The economic value of advanced time series methods for modelling and trading 10-year government bonds", *The European Journal of Finance*, vol 13, no 4, pp 333–52.

Fang, Z, S Yang, C Lv, S An and W Wu (2022): "Application of a data-driven XGBoost model for the prediction of COVID-19 in the USA: a time-series study", *BMJ Open*, vol 12, no 7, e056685.

Fragoso, T, W Bertoli and F Louzada (2018): "Bayesian model averaging: a systematic review and conceptual classification", *International Statistical Review*, vol 86, no 1, pp 1–28.

Gentry, J, D Whitford and P Newbold (1988): "Predicting industrial bond ratings with a probit model and funds flow components", *The Financial Review*, vol 23, no 3, pp 269–86, https://doi.org/10.1111/j.1540-6288.1988.tb01267.x.

Gu, S, B Kelly and D Xiu (2020): "Empirical asset pricing via machine learning", *The Review of Financial Studies*, vol 33, no 5, pp 2223–73, https://doi.org/10.1093/rfs/hhaa009.

Kara, Y, M Boyacioglu and Ö Baykan (2011): "Predicting direction of stock price index movement using artificial neural networks and support vector machines: the sample of the Istanbul Stock Exchange", *Expert Systems with Applications*, vol 38, no 5, pp 5311–19, https://doi.org/10.1016/j.eswa.2010.10.027.

Kauppi, H and P Saikkonen (2008): "Predicting U.S. recessions with dynamic binary response models", *The Review of Economics and Statistics*, vol 90, no 4, pp 777–91, https://doi.org/10.1162/rest.90.4.777.

Kumar, M and M Thenmozhi (2006): *Forecasting stock index movement: a comparison of support vector machines and random forest*, Indian Institute of Capital Markets 9th Capital Markets Conference Paper.

Larsen, G and G Wozniak (1995): "Market timing for active asset allocation: a discrete regression model approach", *Journal of Applied Business Research*, vol 11, no 1, pp 125–35, https://doi.org/10.19030/jabr.v11i1.5899.

Li, Y, D Turkington and A Yazdani (2020): "Beyond the black box: an intuitive approach to investment prediction with machine learning", *The Journal of Financial Data Science*, vol 2, no 1, pp 61–75, https://doi.org/10.3905/jfds.2019.1.023.

Lipton, A and M López De Prado (2020): "A closed-form solution for optimal Ornstein-Uhlenbeck driven trading strategies", *International Journal of Theoretical and Applied Finance*, vol 23, no 8, https://doi.org/10.1142/S0219024920500569.

López De Prado, M (2018): "The 10 reasons most machine learning funds fail", *The Journal of Portfolio Management*, vol 44, no 6, pp 120–33, https://doi.org/10.3905/jpm.2018.44.6.120.

——— (2019): *Tactical investment algorithms*, available at SSRN 3459866.

Nasekin, S (2013): *High-dimensional Lasso quantile regression applied to hedge funds' portfolio*, Master's thesis, Humboldt-Universität zu Berlin.

Nunes, M, E Gerding, F McGroarty and M Niranjan (2018): *Artificial neural networks in fixed income markets for yield curve forecasting*, available at SSRN 3144622.

Nyberg, H (2011): "Forecasting the direction of the US stock market with dynamic binary probit models", *International Journal of Forecasting*, vol 27, no 2, pp 561–78, https://doi.org/10.1016/j.ijforecast.2010.02.008.

Palomares-Salas, J, J De La Rosa, J Ramiro, J Melgar, A Aguera and A Moreno (2009): "ARIMA vs neural networks for wind speed forecasting", 2009 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, pp 129–33, IEEE.

Park, T and G Casella (2008): "The Bayesian lasso", *Journal of the American Statistical Association*, vol 103, no 482, pp 681–6.

Pereira, J, M Basto and A Ferreira da Silva (2016): "The logistic lasso and ridge regression in predicting corporate failure", *Procedia Economics and Finance*, vol 39, pp 634–41, https://doi.org/10.1016/S2212-5671(16)30310-0.

Pollege, S and P Posch (2013): "Managing and trading sovereign risk using credit derivatives and government markets", *Journal of Risk Finance*, vol 14, no 5, pp 453–67, https://doi.org/10.1108/JRF-03-2013-0019.

Rahman, M, A Chowdhury and M Amrin (2022): "Accuracy comparison of ARIMA and XGBoost forecasting models in predicting the incidence of COVID-19 in Bangladesh", *PLOS Global Public Health*, vol 2, no 5, e0000495.

Rapach, D and Z Guofu (2013): "Forecasting stock returns", Chapter 6 in G Elliott and A Timmerman (eds), *Handbook of Economic Forecasting*, vol 2, pp 328–83, Elsevier, https://doi.org/10.1016/B978-0-444-53683-9.00006-2.

Rapach, D, J Strauss and G Zhou (2013): "International stock return predictability: what is the role of the United States?", *The Journal of Finance*, vol 68, no 4, pp 1633–62, https://doi.org/10.1111/jofi.12041.

Rosadi, D, A Nugraha and K Dewi (2011): *Forecasting the Indonesian government securities yield curve using neural networks and vector autoregressive model*, Bank for International Settlements, September.

Roy, S, D Mittal, A Basu and A Abraham (2015): "Stock market forecasting using LASSO linear regression model", Afro-European Conference for Industrial Advancement, pp 371–81, https://doi.org/10.1007/978-3-319-13572-4_31.

Sambasivan, R and S Das (2017): "A statistical machine learning approach to yield curve forecasting", 2017 International Conference on Computational Intelligence in Data Science (ICCIDS).

Sermpinis, G, C Dunis, J Laws and S Charalampos (2012): "Forecasting and trading the EUR/USD exchange rate with stochastic neural network combination and time-varying leverage", *Decision Support Systems*, vol 54, no 1, December, pp 316–29, https:/doi.org/10.1016/j.dss.2012.05.039.

Shynkevich, A (2016): "Predictability in bond returns using technical trading rules", *Journal of Banking & Finance*, vol 70, September, pp 55–69, https://doi.org/10.1016/j.jbankfin.2016.06.010.

Westgaard, S and N van der Wijst (2001): "Default probabilities in a corporate bank portfolio: a logistic model approach", *European Journal of Operational Research*, vol 135, no 2, December, pp 338–49, https://doi.org/10.1016/S0377-2217(01)00045-5.

Yang, J and J Parwada (2012): "Predicting stock price movements: an ordered probit analysis on the Australian Securities Exchange", *Quantitative Finance*, vol 12, no 5, pp 791–804.

Zaidi, M and A Amirat (2016): "Forecasting stock market trends by logistic regression and neural networks: evidence from KSA stock market", *International Journal of Economics, Commerce and Management*, vol 4, no 6, June, pp 220–34.

Zeugner, S (2011): "Bayesian model averaging with BMS", tutorial to the R-package BMS 1e30.

Zhu, M, D Philpotts, R Sparks and M Stevenson (2011): "A hybrid approach to combining CART and logistic regression for stock ranking", *The Journal of Portfolio Management*, vol 38, no 1, pp 100–09, https://doi.org/10.3905/jpm.2011.38.1.100.

ANNEX

Data description (final indicators **in bold**) Table A1

Variable	Classification	Start year	Periodicity	Max	Min	Mean	Indicator
US Govt 2yr-3yr-5yr Butterfly	Butterfly	2003	Intraday	10,7	-80,1	-22,9	Rate
US Govt 2yr-5yr-10yr Butterfly	Butterfly	1999	Intraday	80,9	-75,7	-0,3	Rate
US Govt 5yr-10yr-30yr Butterfly	Butterfly	1999	Intraday	87,9	-58,9	-0,7	Rate
BBG Commodity	Commodity	1999	Daily	238,0	59,5	121,0	Price
Crude Oil Open Interest Combined	Commodity	1999	Daily	3.785.409,0	439.268,0	2.068.316,0	Net balance
Generic 1st 'CO' Future	Commodity	1999	Daily	146,1	14,2	63,4	Price
Gold Spot $/Oz	Commodity	1999	Daily	2.063,5	252,6	956,5	Price
LMEX LONDON METALS INDEX	Commodity	2000	Daily	4.556,6	958,3	2.662,2	Price
EUR-JPY X-RATE	Currency	1999	Daily	169,5	89,5	127,1	Currency
EURO 1 MO	Currency	1999	Daily				Rate
EURO 10 YR	Currency	1999	Daily				Rate
EURO 12 MO	Currency	1998	Daily				Rate
EURO 2 MO	Currency	1999	Daily				Rate
EURO 3 YR	Currency	1999	Daily				Rate
EURO 5 YR	Currency	1999	Daily				Rate
EURO 7 YR	Currency	1999	Daily				Rate
Euro Spot	Currency	1999	Daily	1,6	0,8	1,2	Currency
EUR-USD OPT VOL 1M	Currency	1999	Daily	28,9	3,8	9,6	Index (state)
CFTC CME Euro Fx Total Open Interest/Combined	Currency	1999	Weekly on Tuesday	795.353,0	33.215,0	305.334,0	Net balance
Citi Economic Surprise - United States	Economics	2003	Intraday	270,8	-144,6	4,3	Index
Citi Economic Surprise Index - Eurozone	Economics	2003	Intraday	212,4	-304,6	4,4	Index
Adjusted Retail & Food Services Sales	Economics	1999	Monthly	18,2	-14,7	0,4	Rate
Adjusted Retail Sales Less Autos and Gas Stations	Economics	1999	Monthly	12,1	-14,2	0,4	Rate
Adjusted Retail Sales Less Autos	Economics	1999	Monthly	12,2	-15,1	0,4	Rate
ADP National Employment Report	Economics	2002	Monthly	4.485,5	-19.408,9	47,4	Net balance
Capital Goods New Orders Nondefense Ex Aircraft & Parts	Economics	1999	Monthly	9,2	-10,8	0,1	Rate
Capital Goods Shipments Ex Air	Economics	1999	Monthly	5,2	-8,1	0,1	Rate
Census Bureau US Construction	Economics	1999	Monthly	2,8	-3,7	0,3	Rate
Challenger US Job Cut Announce	Economics	2000	Monthly	1.576,9	-77,4	31,6	Rate
Chicago Fed National Activity	Economics	1999	Monthly	6,0	-17,7	-0,1	Confidence/survey
Conference Board Consumer Confidence	Economics	1999	Monthly	144,7	25,3	95,0	Confidence/survey
Conference Board Consumer Confidence Expectations	Economics	1999	Monthly	119,2	27,3	88,0	Confidence/survey
Conference Board Consumer Confidence Present Situation	Economics	1999	Monthly	186,8	20,2	105,4	Confidence/survey
Conference Board US Leading Index	Economics	1999	Monthly	3,1	-7,6	0,1	Confidence/survey
Dallas Fed Manufacturing Outlook Level of General Business Activity	Economics	2004	Monthly	48,0	-72,2	2,2	Rate
Federal Reserve Consumer Credit	Economics	1999	Monthly	29,3	-64,0	10,3	Net balance
FHFA US House Price Index Purchase Only	Economics	1991	Monthly	-	-	-	Rate
ISM Manufacturing PMI SA	Economics	1999	Monthly	61,4	34,5	52,8	Confidence/survey
ISM Manufacturing Report on Business Employment	Economics	1999	Monthly	62,3	28,0	50,8	Confidence/survey
ISM Manufacturing Report on Business New Orders	Economics	1999	Monthly	71,3	25,9	55,7	Confidence/survey
ISM Manufacturing Report on Business Prices	Economics	1999	Monthly	92,1	17,1	60,0	Confidence/survey
ISM Services PMI	Economics	1999	Monthly	61,3	37,8	54,6	Confidence/survey
Kansas City Federal Reserve SA	Economics	2001	Monthly	25,0	-30,0	4,7	Confidence/survey
Langer US National Economy Expectations Diffusion Index	Economics	1999	Monthly	63,0	8,5	42,0	Confidence/survey
Market News International Chicago Business Barometer	Economics	1999	Monthly	68,6	32,5	54,8	Confidence/survey
Markit US Composite PMI SA	Economics	2018	Monthly				Confidence/survey
Markit US Manufacturing PMI SA	Economics	2018	Monthly				Confidence/survey
Markit US Services PMI Busines	Economics	2018	Monthly				Confidence/survey
Merchant Wholesalers Inventories Total	Economics	1999	Monthly	2,1	-2,0	0,3	Rate
Merchant Wholesalers Sales Total	Economics	1999	Monthly	9,0	-16,4	0,4	Rate
National Association of Home Builders Market	Economics	1999	Monthly	90,0	8,0	50,5	Difference, employees
NFIB Small Business Optimism Index	Economics	1999	Monthly	108,8	81,6	97,6	Index
Philadelphia Fed Business Outlook Survey	Economics	1999	Monthly	37,0	-46,8	7,5	Confidence/survey
Private Housing Authorized by Building Permits	Economics	1999	Monthly	2.263,0	513,0	1.335,0	Level
Private Housing Units Started	Economics	1999	Monthly	24,0	-26,4	0,3	Rate
Private Total Housing Authorized by Building Permits	Economics	1999	Monthly	18,6	-21,9	0,2	Rate
Retail Inventories Seasonally	Economics	1999	Monthly	1,6	-6,2	0,2	Rate
Retail Sales Less Food Service	Economics	1999	Monthly	10,4	-12,4	0,3	Rate
S&P CoreLogic Case-Shiller 20-City Composite Home Price Index	Economics	2000	Monthly	-	-	-	Index
S&P CoreLogic Case-Shiller 20-City Composite Home Price MoM	Economics	2000	Monthly	-	-	-	Rate
S&P CoreLogic Case-Shiller 20-City Composite Home Price YoY	Economics	2001	Monthly	-	-	-	Rate
S&P CoreLogic Case-Shiller U.S	Economics	1987	Monthly	-	-	-	Index
S&P CoreLogic Case-Shiller U.S YoY	Economics	1988	Monthly	-	-	-	Rate
U-3 US Unemployment Rate Total	Economics	1999	Monthly	14,8	3,5	5,9	Rate
UMich Expected Change in Prices During the next 5-10y	Economics	1999	Monthly	3,4	2,2	2,8	Confidence/survey
UMich Expected Change in Prices During the next year	Economics	1999	Monthly	5,2	0,4	3,0	Confidence/survey
University of Michigan Consumer Expectations Index	Economics	1999	Monthly	108,6	47,6	78,6	Confidence/survey
University of Michigan Consumer Sentiment Index	Economics	1999	Monthly	112,0	55,3	86,2	Confidence/survey
University of Michigan Current	Economics	1999	Monthly	121,2	57,5	98,2	Confidence/survey
US Auto Sales Total Annualized	Economics	1999	Monthly	21,8	8,6	15,8	Rate
US Average Hourly Earnings All Employees Total Private MoM	Economics	2006	Monthly				Rate
US Average Hourly Earnings All Employees Total Private YoY	Economics	2007	Monthly				Rate
US Average Weekly Hours All Employees	Economics	2006	Monthly				Index Level
US Capacity Utilization % of Total Capacity	Economics	1999	Monthly	82,3	64,2	77,0	Rate
US CPI Urban Consumers Less Food & Energy YoY	Economics	1999	Monthly	4,5	0,6	2,0	Rate
US CPI Urban Consumers Less Food & Energy Index	Economics	1999	Monthly	279,1	175,7	222,4	Index
US CPI Urban Consumers Less Food & Energy MoM	Economics	1999	Monthly	0,9	-0,4	0,2	Rate
US CPI Urban Consumers MoM SA	Economics	1999	Monthly	1,4	-1,8	0,2	Rate
US CPI Urban Consumers NSA	Economics	1999	Monthly	261,6	165,0	215,7	Index
US CPI Urban Consumers YoY NSA	Economics	1999	Monthly	5,6	-2,1	2,1	Rate

US Durable Goods New Orders Industries	Economics	1999	Monthly	23,0	-18,8	0,2	Rate
US Durable Goods New Orders Total ex Transportation	Economics	1999	Monthly	6,3	-10,3	0,1	Rate
US Empire State Manufacturing	Economics	2001	Monthly	39,0	-78,2	7,9	Confidence/survey
US Employees on Nonfarm Payrolls Total Private MoM	Economics	1999	Monthly	4.807,0	-19.731,0	62,5	Difference
US Employees on Nonfarm Payrolls Total MoM	Economics	1999	Monthly	4.846,0	-20.679,0	69,6	Difference
US Employees on Nonfarm Payrolls Manufacturing Industry	Economics	1999	Monthly	342,0	-1.304,0	-18,7	Difference, employees
US Existing Homes Sales MoM SA	Economics	1999	Monthly	23,7	-22,5	0,2	Rate
US Existing Homes Sales SAAR	Economics	1999	Monthly	7,3	3,5	5,3	Difference
US Export Price By End Use All Commodities MoM	Economics	1999	Monthly	2,7	-3,5	0,1	Rate
US Export Price By End Use All Commodities YoY	Economics	1999	Monthly	17,6	-8,3	1,3	Rate
US Foreign Net Transactions	Economics	1999	Monthly	157,8	-134,9	42,6	Net balance
US Import Price Index by End Use All MoM	Economics	1999	Monthly	3,2	-7,4	0,2	Rate
US Import Price Index By End Use Ex-Petroleum MoM	Economics	1999	Monthly	1,3	-1,7	0,1	Rate
US Import Price Index by End Use All YoY	Economics	1999	Monthly	-	-	-	Rate
US Industrial Production Industry Groups Manufacturing	Economics	1999	Monthly	7,7	-15,8	0,1	Rate
US Industrial Production MoM	Economics	1999	Monthly	6,2	-12,7	0,1	Rate
US Job Openings By Industry Total	Economics	2000	Monthly	-	-	-	Level
US Labor Force Participation	Economics	1999	Monthly	67,3	60,2	64,7	Rate
US Manufacturers New Orders Excluding Transportation	Economics	1999	Monthly	4,8	-8,9	0,2	Relative change
US Manufacturers New Orders Total	Economics	1999	Monthly	10,3	-13,5	0,2	Net balance
US Manufacturing & Trade Inventories Total	Economics	1999	Monthly	1,3	-2,3	0,2	Rate
US New One Family Houses Sold Annual Total MoM	Economics	1999	Monthly	21,0	-33,6	0,2	Rate
US New One Family Houses Sold Annual Total Units/Persons	Economics	1999	Monthly	1.389,0	270,0	708,4	Level
US New Privately Owned Housing	Economics	1999	Monthly	2.273,0	478,0	1.287,0	Level
US Pending Home Sales Index YoY	Economics	2002	Monthly	29,3	-34,6	1,1	Rate
US Personal Consumption Expenditures Chained 2012 $ MoM	Economics	1999	Monthly	8,5	-12,2	0,2	Rate
US Personal Consumption Expenditure Core Price Index MoM	Economics	1999	Monthly	0,7	-0,6	0,2	Rate
US Personal Consumption Expenditures Nominal $ MoM	Economics	1999	Monthly	8,6	-12,6	0,4	Rate
US Personal Consumption Expenditure Core Price Index YoY	Economics	1999	Monthly	3,6	0,6	1,7	Rate
US Personal Consumption Expenditures Chain Type Price Index MoM	Economics	1999	Monthly	1,0	-1,2	0,2	Rate
US Personal Consumption Expenditures Chain Type Price Index YoY	Economics	1999	Monthly	4,2	-1,5	1,8	Rate
US Personal Income MoM SA	Economics	1999	Monthly	12,4	-4,7	0,4	Rate
US PPI Final Demand Less Foods and Energy MoM	Economics	2010	Monthly				Rate
US PPI Final Demand Less Foods Energy and Trade Services MoM	Economics	2013	Monthly				Rate
US PPI Final Demand Less Foods Energy and Trade Services YoY	Economics	2014	Monthly				Rate
US PPI Final Demand Less Foods and Energy YoY	Economics	2010	Monthly				Rate
US PPI Final Demand MoM SA	Economics	2009	Monthly				Rate
US PPI Final Demand YoY NSA	Economics	2010	Monthly				Rate
US Real Average Hourly Earning	Economics	2007	Monthly				Rate
US Real Average Weekly Earning	Economics	2007	Monthly				Rate
US Trade Balance of Goods and Services	Economics	1999	Monthly	-17,7	-69,0	-44,3	Net balance
US Trade in Goods Balance Total	Economics	1999	Monthly	-23,8	-86,1	-57,0	Net balance
US Treasury Federal Budget Debt Summary	Economics	1999	Monthly	214,3	-864,1	-57,8	Difference
US Treasury International Capital	Economics	1999	Monthly	317,0	-194,6	36,6	Net balance
US U-6 Unemployed & Part Time	Economics	1999	Monthly	22,9	6,8	10,7	Rate
Bureau of Labor Statistics Employment Cost	Economics	1999	Quarterly	1,2	0,2	0,7	Rate
Delinquencies As % Of Total Loans	Economics	1979	Quarterly	-	-	-	Rate
FHFA US Purchase-Only	Economics	1991	Quarterly	-	-	-	Rate
FOF Federal Reserve US Households	Economics	1946	Quarterly	-	-	-	Difference
Foreclosures As % Of Total Loans	Economics	1979	Quarterly	-	-	-	Rate
GDP US Chained 2012 Dollars	Economics	1999	Quarterly	33,4	-31,4	2,1	Rate
GDP US Personal Consumption	Economics	1999	Quarterly	41,0	-33,2	2,4	Rate
US GDP Personal Consumption	Economics	1999	Quarterly	3,4	-0,8	1,7	Rate
US GDP Price Index QoQ SAAR	Economics	1999	Quarterly	4,2	-1,8	1,9	Rate
US Labor Productivity Output	Economics	1999	Quarterly	10,6	-4,8	2,0	Rate
US Nominal Account Balance	Economics	1960	Quarterly	-	-	-	Net balance
US Nominal Output Gap as a Percentage of GDP	Economics	1999	Quarterly	2,1	-10,1	-1,4	Rate
US Unit Labor Costs Nonfarm Business Sector	Economics	1999	Quarterly	15,6	-13,4	1,5	Rate
MBA US US Mortgage Market	Economics	1999	Weekly on Friday	112,1	-38,8	0,5	Rate
US Continuing Jobless Claims	Economics	1999	Weekly on Friday	24.912,0	1.649,0	3.282,0	Level (state)
US Initial Jobless Claims	Economics	1999	Weekly on Friday	6.867,0	201,0	397,7	Level (state)
Langer US Weekly Consumer Comf	Economics	1999	Weekly on Sunday	69,0	23,0	43,3	Confidence/survey
DAX INDEX	Equity	1999	Daily	14.109,0	2.203,0	7.628,0	Price
IBEX 35 INDEX	Equity	1999	Daily	15.946,0	5.364,0	9.684,0	Price
S&P 500 INDEX	Equity	1999	Daily	3.934,8	676,5	1.657,3	Price
Bloomberg USDEUR 3 Month Hedging Cost	Financials	1999	Daily	3,6	-2,0	0,7	Rate
Bloomberg USDJPY 3 Month Hedging Cost	Financials	1999	Daily	6,9	0,1	2,1	Rate
BONOS Y OBLIG DEL ESTADO	Financials	1999	Daily	7,5	0,8	4,2	Rate
EUR Eonia Forward 1Y1Y	Financials	2000	Daily	7,5	-0,8	1,7	Rate
EUR SWAP (EONIA) 1WK	Financials	1999	Daily	5,0	-0,5	0,7	Rate
USD INFL FORWARD RATE 1Y1Y	Financials	1999	Daily	5,6	-2,1	0,7	Rate
USD INFL SWAP ZC 10Y	Financials	2004	Daily	3,1	0,8	2,4	Rate
USD INFL SWAP ZC 5Y	Financials	2004	Daily	3,3	-0,6	2,1	Rate
USD SWAP OIS 18M	Financials	2001	Daily	5,6	0,0	1,6	Rate
USD SWAP OIS 1M	Financials	2001	Daily	5,4	0,0	1,4	Rate
USD SWAP OIS 1W	Financials	2001	Daily	5,3	0,0	1,4	Rate
USD SWAP OIS 1Y	Financials	2001	Daily	5,7	0,0	1,5	Rate
USD SWAP OIS 2M	Financials	2001	Daily	5,4	0,0	1,4	Rate
USD SWAP OIS 2W	Financials	2001	Daily	5,4	0,0	1,4	Rate
USD SWAP OIS 2Y	Financials	2001	Daily	5,6	0,0	1,7	Rate
USD SWAP OIS 3M	Financials	2001	Daily	5,4	0,0	1,4	Rate
USD SWAP OIS 3Y	Financials	2002	Daily	4,6	0,0	1,2	Rate
USD SWAP OIS 4M	Financials	2001	Daily	5,5	0,0	1,4	Rate
USD SWAP OIS 4Y	Financials	2002	Daily	5,0	0,0	1,4	Rate
USD SWAP OIS 5M	Financials	2001	Daily	5,5	0,0	1,4	Rate
USD SWAP OIS 5Y	Financials	2002	Daily	5,7	0,0	2,4	Rate
USD SWAP OIS 6M	Financials	2001	Daily	5,5	0,0	1,4	Rate
USD SWAP OIS 9M	Financials	2001	Daily	5,6	0,0	1,5	Rate
USD SWAP SEMI 30/360 10Y	Financials	1999	Daily	7,9	0,5	3,7	Rate
USD SWAP SEMI 30/360 7YR	Financials	1999	Daily	7,8	0,4	3,4	Rate
US Breakeven 10 Year	Financials	1999	Intraday	2,8	0,0	2,0	Rate
Bloomberg CFTC CBT 10-Yr US Tr	Financials	1999	Weekly on Tuesday	608.492,0	-756.316,0	-20.779,0	Net balance
Bloomberg CFTC CME Euro Fx Net	Financials	1999	Weekly on Tuesday	211.752,0	-226.560,0	-4.222,0	Net balance
CFTC CBT 10-Year US Treasury N	Financials	1999	Weekly on Tuesday	5.736.552,0	541.198,0	2.556.514,0	Net balance
CFTC CBT 2-Year US Treasury No	Financials	1999	Weekly on Tuesday	4.423.693,0	32.328,0	998.523,0	Net balance
CFTC CBT 5-Year US Treasury No	Financials	1999	Weekly on Tuesday	5.580.720,0	285.337,0	1.939.930,0	Net balance

AUSTRALIAN GOVERNMENT	Fixed Income	1999	Daily	7,3	0,6	4,3	Yield
BELGIUM KINGDOM	Fixed Income	1998	Daily	-	-	-	Rate
BUNDESREPUB. DEUTSCHLAND	Fixed Income	1999	Daily	5,6	-0,9	2,6	Rate
IBOXX (EUR) DESOV OA TR	Fixed Income	1999	Daily	248,2	96,4	170,5	Price
IBOXX (EUR) ES SOV TR	Fixed Income	1999	Daily	289,0	95,5	176,0	Price
IBOXX (EUR) JUMBO OA TR	Fixed Income	1999	Daily	209,0	96,8	159,8	Price
IBOXX (EUR) SPAIN COVRD TR	Fixed Income	2003	Daily	255,1	122,2	186,2	Price
iBoxx US Trs 1-3 TR	Fixed Income	1999	Daily	185,4	100,6	150,4	Price
JAPAN (7 YEAR ISSUE)	Fixed Income	1999	Daily	1,9	-0,4	0,6	Rate
JAPAN (10 YEAR ISSUE)	Fixed Income	1999	Daily	2,0	-0,3	0,9	Rate
JAPAN (1 YEAR ISSUE)	Fixed Income	1999	Daily	0,8	-0,4	0,1	Rate
JAPAN (2 YEAR ISSUE)	Fixed Income	1999	Daily	1,1	-0,4	0,1	Rate
JAPAN (5 YEAR ISSUE)	Fixed Income	1999	Daily	1,6	-0,4	0,4	Rate
US Generic Govt 12 Mth	Fixed Income	1999	Daily	6,4	0,0	1,4	Rate
USD Trsyies 3-5Y Tot	Fixed Income	1999	Daily	243,5	98,0	174,1	Price
USD Trsyies 5-7Y Tot	Fixed Income	1999	Daily	283,2	96,1	186,0	Price
USD Trsyies 7-10Y Tot	Fixed Income	1999	Daily	313,5	93,5	191,7	Price
US Generic Govt 10 Yr	Fixed Income	1999	Intraday	6,8	0,5	3,4	Rate
US Generic Govt 2 Yr	Fixed Income	1999	Intraday	6,9	0,1	2,1	Rate
US Generic Govt 3 Yr	Fixed Income	1999	Intraday	6,9	0,1	2,3	Rate
US Generic Govt 5 Yr	Fixed Income	1999	Intraday	6,8	0,2	2,8	Rate
US Generic Govt 7 Yr	Fixed Income	2009	Intraday				Rate
TREASURY BILL	Monetary	2001	Daily	5,3	-0,1	1,2	Rate
US Treasury 3 Month Bill Money	Monetary	1999	Daily	6,3	0,0	1,7	Rate
Federal Funds Target Rate Mid	Monetary	1999	Intraday	6,5	0,1	1,8	Price
US Federal Funds Effective Rat	Monetary	1999	Intraday	7,0	0,0	1,8	Rate
US Generic Govt 3 Mth	Monetary	1999	Intraday	6,4	-0,1	1,7	Rate
Federal Reserve Balance Sheet	Monetary	1999	Monthly	35,2	5,5	14,5	Balance
BofA Securities GF	Other	2000	Daily	3,0	-0,7	0,1	Index (state)
GeoQuant Italy Extent of Political Risk	Other	2016	Daily				Index
GeoQuant Italy Political Risk Score	Other	2016	Daily				Index
GeoQuant Italy Political Risk Score Forecast	Other	2018	Daily				Index
GeoQuant United States Politic	Other	2016	Daily				Index
Bloomberg Country Risk Politic	Other	2009	Quarterly				Index
ECB Survey of Professional Forecasters HICP 1y Ahead	Survey	1999	Quarterly	2,4	0,8	1,6	Rate
ECB Survey of Professional Forecasters HICP 5y Ahead	Survey	1999	Quarterly	2,0	1,6	1,9	Rate
Survey of Prof Forecasters Moody's BAA Corporate Bond	Survey	2010	Quarterly	6,4	3,4	4,9	Rate
Survey of Prof Forecasters Moody's AAA Corporate Bond	Survey	1999	Quarterly	7,8	2,5	5,3	Rate
Survey of Professional Forecasters 5y CPI Inflation Rate	Survey	2005	Quarterly	2,8	1,9	2,2	Rate
Survey of Professional Forecasters 10y Treasury Bill Current Q	Survey	1999	Quarterly	6,6	0,6	3,4	Rate
Survey of Professional Forecasters Anxious Index Current Q +4	Survey	1968	Quarterly	-	-	-	Rate
Survey of Professional Forecasters Anxious Index Current Q +1	Survey	1999	Quarterly	74,8	4,3	16,3	Rate
Survey of Professional Forecasters 10y Treasury Bill Rate Prior Q	Survey	1999	Quarterly	6,5	0,6	3,4	Rate
Survey of Professional Forecasters 10y Treasury Bill Current Q+4	Survey	1999	Quarterly	6,5	0,9	3,9	Rate
U.S. JP Morgan Treasury Investor Sentiment All Client Long	Survey	2003	Weekly on Monday	50,0	0,0	17,1	Rate
U.S. JP Morgan Treasury Investor Sentiment Active Client Long	Survey	2003	Weekly on Monday	60,0	0,0	12,1	Confidence/survey
U.S. JP Morgan Treasury Investor Sentiment All Client Neutral	Survey	2003	Weekly on Monday	85,0	26,0	58,1	Confidence/survey
U.S. JP Morgan Treasury Investor Sentiment Active Client Short	Survey	2003	Weekly on Monday	70,0	0,0	16,4	Confidence/survey
U.S. JP Morgan Treasury Investor Sentiment Active Client Neutral	Survey	2003	Weekly on Monday	100,0	3,0	37,1	Confidence/survey
U.S. JP Morgan Treasury Investor Sentiment All Client Short	Survey	2003	Weekly on Monday	66,0	0,0	24,8	Confidence/survey
Cboe Volatility Index	Uncertainty	1999	Daily	82,7	9,1	20,1	Index
Índice de Volatilidad Global	Uncertainty	2000	Daily	7,0	-1,3	0,0	Rate
MOVE	Uncertainty	1999	Daily	264,6	36,6	87,5	Index
Geopolitical Risk Index	Uncertainty	1985	Monthly				Index
US Economic Policy Uncertainty	Uncertainty	1999	Monthly	350,5	57,2	119,9	Index
US Treasury Yield Curve Rate T	Yield Curve	2001	Daily	5,3	0,0	1,2	Rate
Market Matrix US Sell 10 Year	Yield Curve	1999	Intraday	159,8	-42,9	62,6	Rate
Market Matrix US Sell 2 Year &	Yield Curve	1999	Intraday	291,0	-56,0	123,9	Rate
Market Matrix US Sell 5 Year &	Yield Curve	1999	Intraday	149,2	-42,2	62,0	Rate

The data are available from Bloomberg, except the Geopolitical Risk Index, which is available from Matteo Iacoviello's website (www.matteoiacoviello.com/gpr.htm#data).

Source: Authors' elaboration.

Data classification

Table A2

Classification	Number of indicators
Butterfly	3
Commodity	5
Currency	11
Economics	123
Equity	3
Financials	45
Fixed Income	22
Monetary	6
Other	6
Survey	17
Uncertainty	5
Yield Curve	4

Source: Authors' elaboration.

Data periodicity

Table A3

Periodicity	Number of indicators
Daily	85
Intraday	17
Monthly	107
Quarterly	24
Weekly	17

Source: Authors' elaboration.

Middle out: only extreme deciles matter[1]

Ashwin Alankar,[2] Philip Maymin[3] and Myron Scholes[4]

Abstract

For many return distributions, both empirical and simulated, replacing the middle with zeroes and keeping only returns in the top decile or bottom decile still yields a 90% correlation with the original and exhibits 90% of its volatility. This heuristic holds broadly across various asset classes, including 20 years of daily returns on broad equity and fixed income indexes, and regardless of whether the breakpoints are calculated on an overall or a rolling historical basis. We show that for the thin-tailed Gaussian, middling out results in a lower correlation of 80%, while for more realistic distributions such as the Student T distribution with three degrees of freedom, the correlation is again 90%, and these results also hold in the presence of changing distributions over time. Further, we document two interesting theoretical properties of the 80% middle-out cutoff for the Gaussian distribution by showing it is the only cutoff that equals the correlation and also the only cutoff whose marginal correlation exactly equals one in magnitude. Middling out is stronger when kurtosis is higher or return horizons are shorter. In addition to these quantitative correlation and volatility results, we argue that tail returns should also have an even greater importance for investment managers and asset allocators because of their outsized effects on compound returns, the difficulty in hedging or reacting to tail conditions when they do occur, and the observation that tail returns are the ones that reveal actual asset information, while the middle returns are merely middling noise. Investment professionals should apply middle-out thinking and focus their risk management and portfolio decision policies on tail events.

Keywords: middles, tails, deciles, risk, compound returns.

[1] The opinions and views expressed are those of the authors and are subject to change without notice. They do not necessarily reflect the views of others in Janus Henderson's organisation and no forecasts can be guaranteed.

[2] Head of Global Asset Allocation and Portfolio Manager, Janus Henderson Investors, ashwin.alankar@janushenderson.com.

[3] Director of Asset Allocation Strategies, Janus Henderson Investors, philip.maymin@janushenderson.com, corresponding author.

[4] Chief Investment Strategist, Janus Henderson Investors, myron.scholes@janushenderson.com.

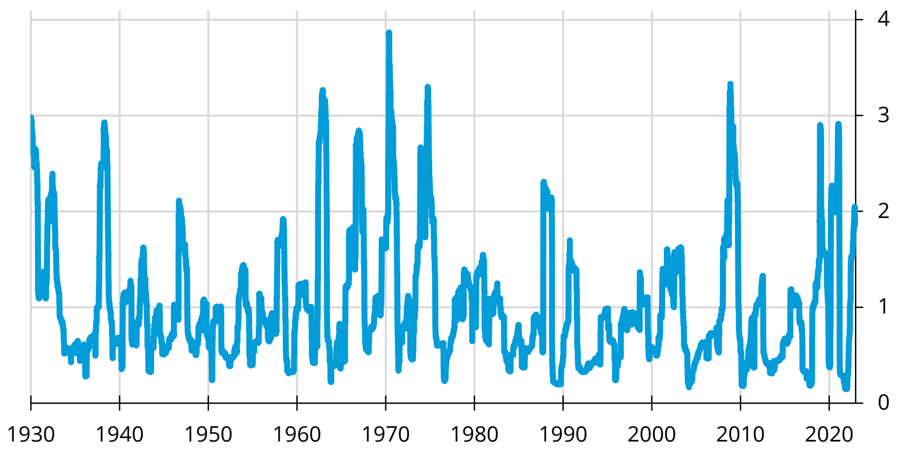
1. Introduction

Among the American Finance Association (2022)'s top 25 cited Journal of Finance articles of all time, 14 use information from financial market prices. Of those, 13 focus on the average returns (Carhart (1997), Fama and French (1992), Jegadeesh and Titman (1993), Daniel et al (1998), Glosten et al (1993), De Bondt and Thaler (1985), Fama and French (1996), Markowitz (1952), Baker and Wurgler (2006), Ang et al. (2006), Sharpe (1964), Bansal and Yaron (2004), and Hong and Stein (1999)); the only exception is Forbes and Rigobon (2002), who distinguish market co-movement from correlation by adjusting for a correlation bias due to market volatility.

This focus on average returns is rampant in the industry as well. Portfolio managers report and are ranked by average or risk-adjusted returns (eg Morningstar (2022)), even though there is little evidence that such ratings have any ability to forecast future performance (Blake and Morey (2009)). Risk managers are effectively required to use metrics that explicitly ignore tail risks. Even the latest Basel (2017) standards for bank capital requirements, which ultimately flow through to govern virtually all risk management with its centralised definition of the "standardised approach," explicitly ignores tail risk by, for example, routinely requiring risk to be measured to within some confidence interval. One implicit problem with confidence intervals is that they highlight a single point such as the 0.05 level. A common misinterpretation is to observe volatility of, say, 15% and erroneously presume this implies a 30% loss on their fund in extreme conditions, or a ratio of 2:1. The actual losses could be as great as 45%, or a ratio of 3:1. Graph 1 illustrates by plotting on a daily rolling basis the ratio of the forward-looking one-year maximum drawdown to the backward-looking one-year realised volatility, on the S&P 500 from 1928 to 2022. Indeed, ratios in excess of 3:1 occurred 10 times more frequently than would be expected by a Gaussian distribution.

Forward-looking one-year maximum drawdown divided by the backward-looking one-year realised volatility on the S&P 500 — Graph 1

Daily rolling ratios



Source: Bloomberg, 1928–2022. As of Nov 2022.

We aim to argue here that this approach, while ubiquitous and conventional, is entirely upside-down. Rather than viewing tails and extreme events as unexplainable, unforeseen, unavoidable outliers that need to be ignored while focusing on the more common situations, we argue that the tails are where all the information and value is, and that the middles of the distribution are the actual noise. Even scenario analysis can't forecast the next tail event; it can only extrapolate from the prior observed tail events.

This paper proceeds as follows. Section 2 describes the simple method of middling out and shows its result on theoretical distributions. Section 3 evaluates middling out on empirical return distributions. Section 4 expands the discussion to implications for portfolio managers, asset allocators, risk managers and other investment professionals. Section 5 concludes.

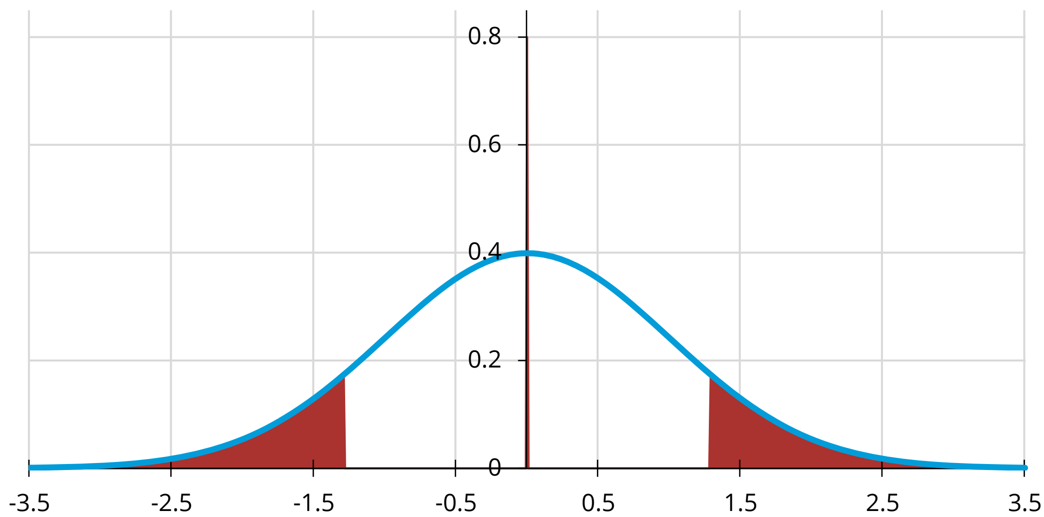
2. Middle out

Middling out a distribution means transforming the distribution to keep only the values in the tails and replacing the middle values with a replacement value. Graph 2 illustrates an 80% middled-out normal distribution, meaning values in the top 10% and bottom 10% of the distribution remain unchanged while all the values in the middle 80% are replaced with zero; the probability density function of the standard normal distribution is superimposed with the middled-out distribution.

While in complete generality a middled-out distribution could use a replacement value other than zero (for example, the mean or the median), for the remainder of this paper we focus only on middle-outs with zero as the replacement value. From the perspective of calculating correlations, volatilities and similar measures, this is a conservative choice.

A standard normal and an 80% middled-out standard normal — Graph 2

Probability density functions

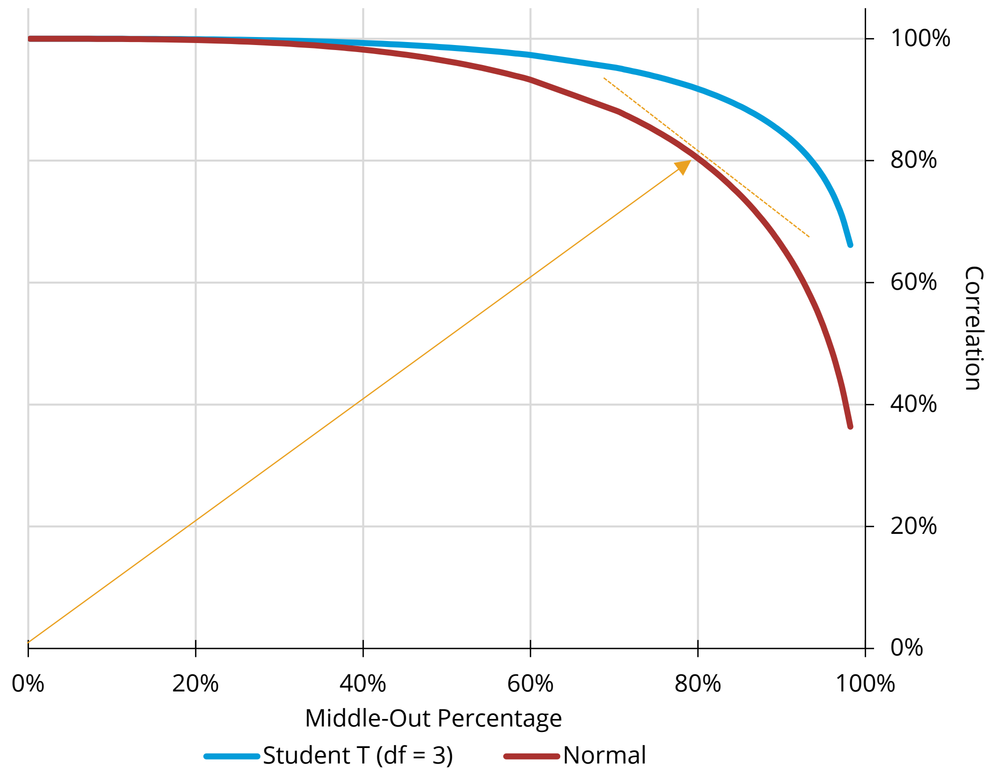


Source: Simulations.

Graph 3 computes the correlation between 10,000 simulated values from a standard normal distribution and a Student T distribution with three degrees of freedom with the middled-out distributions for a variety of middle-out percentages. For example, the 80% middled-out normal distribution has a correlation of 80.74% with its original simulated values and the 80% middled-out Student T distribution with three degrees of freedom has a correlation of 90.96%.

Correlation between middled-out and raw simulation values — Graph 3

For Student T distribution with 3 degrees of freedom, and Normal



Source: Simulations.

We can derive a closed-form formula for the correlation ρ between a standard normal and a middled-out version that replaces the middle m percent of the distribution with zeros:

$$\rho = \sigma = \sqrt{1 - m + \frac{2e^{-\mathrm{erfc}^{-1}(1-m)^2}\mathrm{erfc}^{-1}(1-m)}{\sqrt{\pi}}}$$

where erfc^{-1} is the inverse complementary error function. Note that the standard deviation σ of the middled-out distribution on a standard normal will also equal the correlation ρ. Hence all conclusions about the correlation in this case also apply to the middled-out volatility.

The correlation between the original distribution and the "middled-in" distribution, where only values in the middle are kept while those in the extreme deciles are replaced with zeroes, can be computed similarly. The middled-in correlation begins to exceed the middled-out correlation when ρ = 0.50, which occurs when m = 0.88. In other words, for a standard normal distribution, the middled-out and middled-in correlations are the same when the top and bottom 6% of the distribution are considered, thus providing an upper bound of m = 0.88 for the middle-out percentage. However, the (rounded) value of m = 0.80 displays two striking theoretical facts, graphically shown with the orange lines in Graph 3:

(1) It is the only value for which $\rho(m) = m$. (Exact solution: m = 80.3%.)

(2) It is also the only value for which $\rho'(m) = -1$. (Exact solution: m = 79.7%.)

The above discussion shows that the top and bottom deciles are the primary drivers of risk in terms of volatility and correlation with the original distribution. In other words, if the choice is between the traditional approach of ignoring the extremes or the suggested approach here of ignoring the middles instead, the middle-out approach is superior, as the tails contain the most important risk information. However, this begs the question: if all the returns are available, why not use all of the returns? Granted that ignoring the tails is far worse than ignoring the middles, but why ignore anything at all?

One way to address this question theoretically is to consider compound returns. If we simulate a year's worth of daily returns on a normal distribution with illustrative mean μ = 5% / 252 and standard deviation σ = 15% / √252, we can split the year into the top decile, bottom decile and middle, and compute the compound return of each group. If we do 100,000 simulations, we can report both the average and the standard deviation of the compound returns.

The first column of Table 1 reports the results. First, it is important to note that the compound return of the middle of the distribution, as well as the total, falls short of the expected average annualised return of 5%. This drag is due to the variance of the returns: a risk-free sequence would have had a compound return of exactly 5%. Second, the compounding effect appears to be stronger for the upside than for the downside, but of course the cumulative contributions of both tails combined must be near zero because the distribution had no skew: (1 – 34%) * (1 + 51%) – 1 ≈ 0%. Third, the total compound return exceeds the middle-only compound return, even for the skew-less normal distribution, because the tails incorporate some of the expected returns.

The second column of Table 1 reports similar results for a skew-normal distribution whose expectation and variance match the normal distribution above but with a modest +0.75 positive skewness. Now, the middle of the distribution has a very negative compound return, because it is by construction missing out on the positive skewness. By construction, the average total compound return and the standard deviation of the total compound return are the same across the normal and skew-normal distributions, but the locations of the volatility differ: the downside volatility is much higher absent skew. Incorporating a skew substantially heightens the importance of the tails. Now, if one were able to ignore the middle, the return wouldn't be break-even around zero, but substantially positive: (1 – 28%) * (1 + 61%) – 1 ≈ 16%. In addition, rather than missing out on the 4.4% positive compound of the normal middle, now we are avoiding a 9.3% loss in the skew-normal middle.

Compound returns of 252 normal returns and skew-normal — Table 1

Simulated 100,000 times

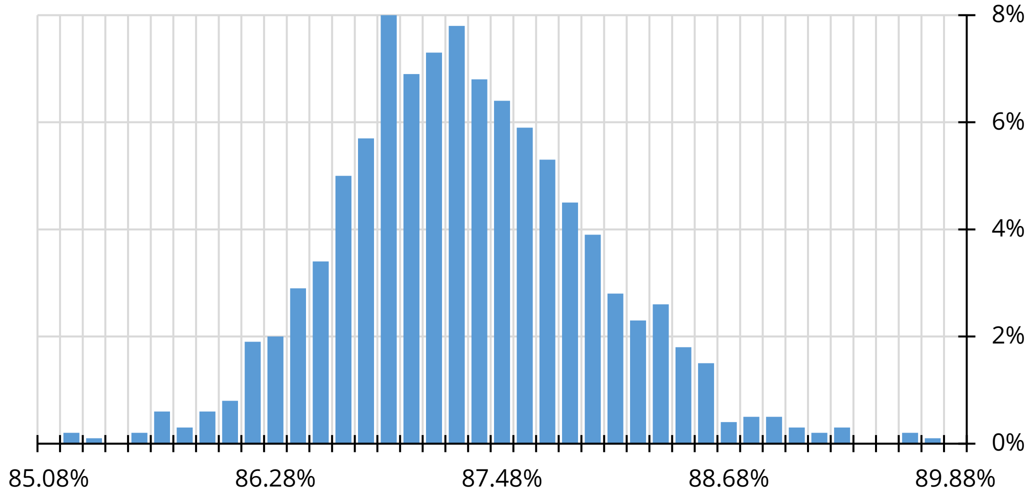
	Normal		Skew-Normal	
	Average	Standard deviation	Average	Standard deviation
Left (bottom decile)	–34%	2%	–28%	1%
Middle	4%	13%	–9%	11%
Right (top decile)	51%	4%	61%	6%
Total	5%	16%	5%	16%

Source: Simulations.

The real world does not represent repeated draws from a single static distribution, of course. The next section explores empirical results, but we can first explore what happens if the distribution changes over time. We simulate 100 years of returns where each year we choose an annualised mean uniformly randomly between –30% and +30% and a volatility uniformly randomly between 5% and 45% to draw 252 daily returns from a normal distribution with that year's mean and volatility. We then compute the correlation between the resulting 25,200 returns and the 80% middled-out returns. Graph 4 plots the histogram of results across 10,000 simulations. The minimum correlation was about 85% and the maximum 90%. In other words, the middle-out perspective continues to hold even for dynamic distributions that vary wildly over time, even for thin-tailed distributions such as the normal. The results would be even more extreme for fat-tailed distributions with varying parameters.

Histogram of correlations — Graph 4

Between middled-out and dynamic simulated random returns



Source: Simulations.

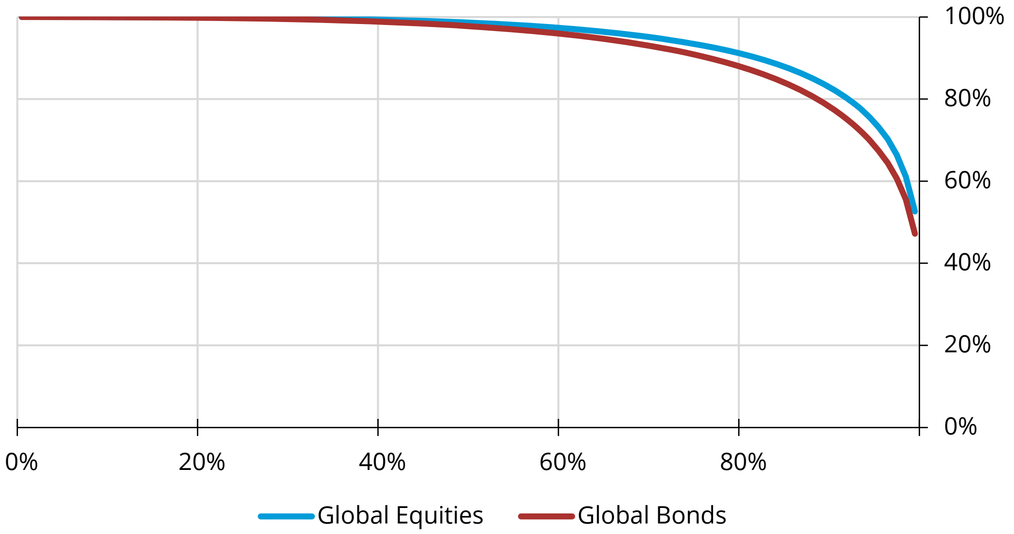
3. Empirical results

To evaluate the empirical results of middle-out, we focus on 20 years of history of global equities and global bonds. For global equities, we use the MSCI ACWI Net Total Return Local Index (Bloomberg ticker: NDLEACWF Index). For global bonds, we use the Bloomberg Global Aggregate Index, described as a flagship measure of global investment grade debt from 24 local currency markets (Bloomberg ticker: LEGATRUH Index). The daily returns were calculated from Bloomberg from 26 February 1999 through 14 April 2023. Similar results were obtained for equities and debt comprising only US assets, using S&P 500 total returns and the returns of the Bloomberg US Government/Credit Bond Index (LUGCTRUU Index).

Graph 5 shows the correlation curve as a function of the middle percentage for both of these global indexes. The correlation between 80% middled-out equities and the original equities is 91%. The correlation between 80% middled-out bonds and the original bonds is 87%. Not shown but virtually identical are the ratio of volatilities between the middled-out and the original distributions: these numbers were 91% and 87%, respectively, as well.

Correlation between middled-out and historical index returns — Graph 5

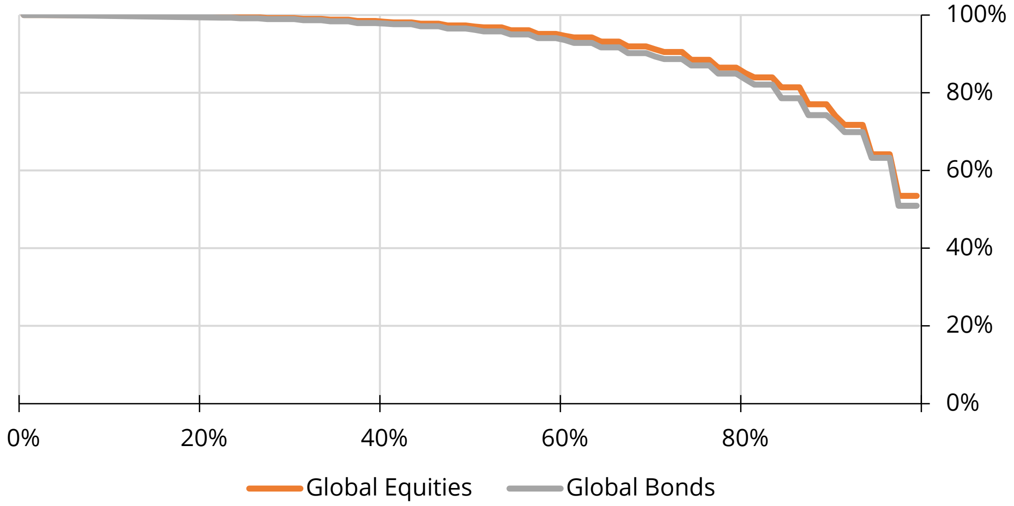
Feb 1999–Apr 2023 for NDLEACWF and LEGATRUH indexes



Source: Simulations.

One potential concern is that the middling-out thresholds are chosen using the entire history, which could potentially hide a form of lookahead bias. To address that concern, we run a rolling middle-out where the thresholds are chosen only from the returns of the past 60 trading days. Graph 6 shows those results, which are broadly the same. The correlation between 80% rolling-middled-out equities and the original equities is 85%. The correlation between 80% middled-out bonds and the original bonds is 83%. Again, the ratios of volatilities were also almost identical at all points.

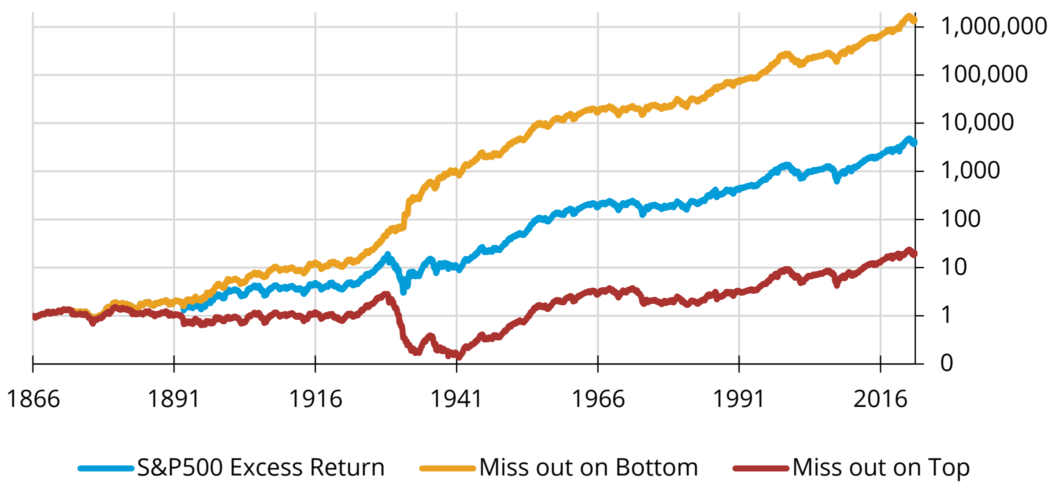
Correlation between rolling middled-out and historical index returns — Graph 6
60-day rolling, Feb 1999–Apr 2023 for NDLEACWF and LEGATRUH indexes



Source: Simulations.

As with the theoretical results, one may note that the above has convinced us, first, that the tails are more important than the middles, and, second, that the optimal threshold for middling out appears to be around 80%. In other words, rather than ignoring the extremes, it is much better to ignore the middles instead. Yet we are faced with the same question as before: why ignore anything at all? The earlier simulation results showed that, for skewed distributions, ignoring the middle can lead to substantially better performance. We now explore the same question in the context of empirical results.

Growth of $1 in excess of cash invested in US large-cap equities — Graph 7
Hypothetical, Ibboton 1857–1925 and S&P 500 1926–2016



Source: Ibbotson and Bloomberg.

Using Ibbotson, for the period 1/1/1857–12/31/1925, individual security returns were gathered from US financial periodicals on a monthly basis, beginning with the official list of the New York Stock Exchange during that time period; from the period 1/1/1926–12/31/2022 returns were represented by the S&P 500 Index.

Graph 7 plots the hypothetical growth of a single $1 investment in US large-cap equities at the beginning of the period for three scenarios: always invested, missing the extreme tail losses, and missing the extreme tail gains. Here, an extreme tail gain or loss is any monthly period whose performance is two standard deviations above or below the average monthly return for the entire period.

Another way of exploring the importance of tails might be to calculate the marginal improvement due to a 50% improvement in either the left tail, the right tail, or the middle. To do that, we use gross daily returns of the S&P 500 since January 1928 through April 2023 from Bloomberg (SPX Index).

Table 2 shows the results.

The left and right tails are almost 40x larger in magnitude than the middle, about 200 bp each versus about 5.5 bp for the middle. The overall compound annualised return is 9.5%. That increases to 15.3% if the middles improve by 50%, ie if we multiply the middle return by 1.5. But it increases to more than 40% if we improve either the left or the right tails by 50% (reduce the downside by 50% or increase the upside by 50%), and to nearly 85% if both tails are improved and the middle remains untouched. It is hard to imagine a more dramatic difference.

Compound annualised returns of the S&P 500
1928–2022

Table 2

	Average
Average of bottom 10% of daily returns	–209 bp
Average of middle 80% of daily returns	5.5 bp
Average of top 10% of daily returns	209 bp

	Compound Annualised Returns
Overall	9.5%
Multiply middle returns by 1.5	15.3%
Multiply downside tail returns by 0.5	43.2%
Multiply upside tail returns by 1.5	41.0%
Multiple upside by 1.5 and downside by 0.5	84.5%

Source: Bloomberg.

4. Discussion and extensions

The results shown here are generalisable to other distributions and other measures of co-movement. For example, the fatter-tailed a distribution is, the higher the correlation would be: a Cauchy distribution is nearly 100% correlated with its middled-out version for essentially any middle-out percentage. On the flip side, using rank correlation generates distribution-free results that are similar to the Normal case.

In addition to these correlation and volatility results, tail returns should also have an even greater importance for investment managers and asset allocators because of their outsize effects on compound returns. The compound return of an investment, or its geometric average, underperforms the simple average return, and this amount of underperformance increases with larger volatility. And as we have seen, virtually all of the volatility of a financial asset comes from its tails.

On a practical level, when markets are in the middle region of a distribution, hedging and reacting to events can appear deceptively smooth and easy. When tail conditions do occur, hedging or reacting can become inordinately difficult, and may come as a surprise to those accustomed to middling returns.

On a philosophical level, tail returns are likely the ones that reveal actual asset information, while the middle returns are merely middling noise. The classic temptation is to treat tail returns as outliers, noise, or acts of God and effectively ignore them in standard portfolio construction methodologies such as mean-variance optimisation and standard risk management metrics such as value-at-risk, but this is exactly backwards. Earnings announcements, news and one-off events drive tail events and change asset prices substantially; without these drivers, prices essentially randomly fluctuate without any signal. Investment professionals should instead apply middle-out thinking and focus their risk management and portfolio decision policies on tail events.

Finally, if we take the view that the returns on a portfolio have a systematic component and a residual component, with tail events, the correlations among assets increase, thus reducing the diversification effect of individual securities. The volatility of the portfolio thus increases as a result. Even more likely, the total volatility is also related to tail events.

Conclusion

Broadly speaking, the two extreme deciles of financial returns represent about 90% of the volatility. Yet typically, extreme events are the ones that are ignored by conventional risk models that focus on estimating risk primarily from the middle of the distribution; tails are incorrectly viewed as “noise.”

This conventional approach is exactly backwards: indeed, it is the extremes where the information is, and the middle of the distribution is the noise. Investors, asset allocators, risk managers and other finance professionals should stop focusing on the middling contribution of the middle and focus instead on the tales of the tails.

References

American Finance Association (2022): "Top 50 cited articles of all time", 21 April, https://afajof.org/most-cited-articles/.

Ang, A, R Hodrick, Y Xing and Z Zhang (2006): "The cross-section of volatility and expected returns", *Journal of Finance*, vol 61, no 1.

Baker, M and J Wurgler (2006): "Investor sentiment and the cross-section of stock returns", *Journal of Finance*, vol 61, no 4.

Bansal, R and A Yaron (2004): "Risks for the long run: a potential resolution of asset pricing puzzles", *Journal of Finance*, vol 59, no 4.

Basel Committee on Banking Supervision (2017): *Basel III: Finalising post-crisis reforms*, 21 April.

Blake, C and M Morey (2009): "Morningstar ratings and mutual fund performance", *Journal of Financial and Quantitative Analysis*, vol 35, no 3, pp 451–83.

Carhart, M (1997): "On persistence in mutual fund performance", *Journal of Finance*, vol 52, no 1.

Daniel, K, D Hirshleifer and A Subrahmanyam (1998): "Investor psychology and security market under- and overreactions", *Journal of Finance*, vol 53, no 6.

De Bondt, W and R Thaler (1985): "Does the stock market overreact?", *Journal of Finance*, vol 40, no 3.

Fama, E and K French (1992): "The cross-section of expected returns", *Journal of Finance*, vol 47, no 2.

——— (1996): "Multifactor explanations of asset pricing anomalies", *Journal of Finance*, vol 51, no 1.

Forbes, K and R Rigobon (2002): "No contagion, only interdependence: measuring stock market comovements", *Journal of Finance*, vol 57, no 5.

Glosten, L, R Jagannathan and D Runkle (1993): "On the relation between the expected value and the volatility of the nominal excess return on stocks", *Journal of Finance*, vol 48, no 5.

Hong, H and J Stein (1999): "A unified theory of underreaction, momentum trading, and overreaction in asset markets", *Journal of Finance*, vol 54, no 6.

Jegadeesh, N and S Titman (1993): "Returns to buying winners and selling losers: implications for stock market efficiency", *Journal of Finance*, vol 48, no 1.

Markowitz, H (1952): "Portfolio selection", *Journal of Finance*, vol 7, no 1.

Sharpe, W (1964): "Capital asset prices: a theory of market equilibrium under conditions of risk", *Journal of Finance*, vol 19, no 3.

Since 2008, the Bank for International Settlements and the World Bank have organised – jointly with cosponsoring central banks – the Public Investors Conference to discuss policy issues, quantitative methods and current challenges for central banks, sovereign wealth funds and public pension plans.

This publication covers many of the advances in the practice of public investment management presented at the eighth edition of the Conference. The volume is the second of a biennial series that is edited by staff members of the cosponsoring institutions and is published by the BIS.

Public Investors Conference
www.publicinvestorsconference.org
publicinvestorsconference@worldbank.org