



Evolving Practices in Public Investment Management

Proceedings of the Seventh
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 Italy on 22-23 October 2018. 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).

© *Bank for International Settlements 2020. All rights reserved. Brief excerpts may be reproduced or translated provided the source is stated.*

ISSN 2708-0358 (online)
ISBN 978-92-9259-354-4 (online)

Preface

Public investors are facing a host of challenges. One is the macroeconomic environment, which – for many years – has been characterised by low yields and compressed risk premia across a range of reserve assets. At the same time, an evolving regulatory environment, potential shifts in the geopolitical landscape, dramatic changes in financial technology, and an awareness of the challenges stemming from investments in less-conventional reserve assets, such as equities and corporate bonds, have added complexity to the investment process for many institutions. As a result, there is now a confluence of factors raising questions about public investors' governance, decision-making processes, risk preferences, modelling approaches, and investment strategies.

Against this backdrop, the Bank for International Settlements and the World Bank – jointly with Bank of Canada and the Banca d'Italia – held the Seventh Public Investors Conference in October 2018 at the headquarters of the Banca d'Italia in Rome, Italy. The conference aimed to explore advances in the practice of public investment management on topics such as strategic asset allocation, reserve management frameworks, sustainability considerations for official institutions, portfolio construction, market inefficiencies and risk premia, and risk management.

This book covers many of the papers presented at the Conference, including some of the most up-to-date research and ideas on public asset management. This book is relevant to at least four categories of readers: (1) practitioners of public investment management, (2) investment consultants advising public managers, (3) academics/researchers, and (4) regulatory and oversight bodies of public investors. By engaging with readers on both the state-of-the-art research dealing with, and policies adopted by, public investors, this book aims to provide important context and perspective to a variety of interested audiences.

The book is organised into four parts, each covering one of the major topics covered during the Seventh Public Investors Conference:

- Part 1 focuses on central bank reserve management. The first chapter illustrates the methodological aspects of a strategic asset allocation process at a central bank that integrates the full balance sheet of the institution. Chapter two narrates how central bank reserve management procedures have evolved over the last two decades, and ponders the question of why central banks are so risk-averse.
- Part 2 presents sustainability considerations for official institutions. The first chapter provides an overview of environmental, social and governance (ESG) investing from the perspective of public investors, drawing on an informal survey conducted by the BIS among a number of central banks, international organisations and asset managers. The second chapter applies a machine learning algorithm to identify patterns between ESG profiles and financial performance for companies in a large investment universe, introducing new ideas to link ESG behaviour and the economy. Finally, the third chapter analyses whether institutional investors engage with the companies which they invest in to discuss corporate externalities such as greenhouse emissions – and if so, why – by analysing shareholder meeting votes by two major global investors.

- Part 3 considers quantitative tools for portfolio construction. The first chapter proposes a robust optimisation method used to obtain *near-optimal* portfolios, aiming to support rather than replace the investment decision-making process. The second chapter illustrates a related toolkit to conduct portfolio construction and selection using two decision-support tools: near-optimal portfolio analysis, and a tool to analyse the sensitivity of the ex-ante optimal portfolio to changes in the forecasts of key risk factors. Chapter three concludes by introducing a new algorithm to solve portfolio optimisation problems with hard-to-optimise objective functions based on simulated annealing.
- Part 4 presents risk management approaches and methodologies. The first two chapters discuss model risk, with the first arguing that model risk is not a new problem faced solely by financial market participants, but rather one for which the lessons learned across other disciplines can help to mitigate and manage financial model risk. The second proposes a holistic framework to reduce model risks. The third and final chapter proposes a measure to gauge liquidity risk – a feature required by central bank portfolios but hard to assess in practice.

This book would not have been possible without the contributions, first and foremost, of the presenters at the Seventh 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 Banca d'Italia and the funding provided by all sponsoring institutions to make the conference possible. We also thank the many participants, reviewers and staff from multiple institutions whose insightful comments and efforts contributed greatly to the preparation of this book, including Mario Barrantes, Pierre Cardon, James Chapman, Joachim Coche, Antonio Diez de los Rios, Nicola Faessler, Niloufar Khavari, Simone Letta, Arunma Oteh, Margarita Sanchez and Louisa Wagner.

Washington, DC	Eric Bouyé
Ottawa, ON, Canada	Narayan Bulusu
Rome, Italy	Marco Fanari
Rome, Italy	Patrizia Ferrauto
Basel, Switzerland	Mike McMorrow
Washington, DC	Shengting Pan
Rome, Italy	Tommaso Perez
Basel, Switzerland	Vahe Sahakyan
Basel, Switzerland	Omar Zulaica

Contents

Preface	i
Part 1. Central Bank Reserve Management	
The strategic asset allocation of the investment portfolio in a central bank Marco Fanari and Gerardo Palazzo, Banca D'Italia.....	1
Are central banks too risk-averse? Massimiliano Castelli, UBS Asset Management Stefan Gerlach, EFG International	23
Part 2. Sustainability Considerations for Official Institutions	
ESG investing: the role of public investors in sustainable investing Ulrike Elsenhuber and Adela Skenderasi, Bank for International Settlements.....	45
ESG investments: filtering versus machine learning approaches Carmine de Franco and Bruno Monnier, Ossiam Christophe Geissler and Vincent Margot, Advestis.....	57
BlackRock vs Norway Fund at shareholder meetings: institutional investors' votes on corporate externalities Marie Brière, Amundi Asset Management Sébastien Pouget, Toulouse School of Economics Loredana Ureche-Rangau, University of Picardie Jules Verne	81
Part 3. Portfolio Construction	
Robust optimisation by constructing near-optimal portfolios Martin van der Schans, Ortec Finance.....	111
Strategic asset allocation from theory to practice: new decision-support tools Golan Benita and David Hoffman, Bank of Israel.....	125
Portfolio optimisation problems with hard-to-optimise objective functions Martin Vesely, Czech National Bank	143
Part 4. Risk Management	
Stealing others' strategies: lessons on model risk David Jamieson Bolder, The World Bank	169
Model risk: a novel approach using a category-oriented framework Paulo Cacella, Banco Central do Brasil.....	183
An alternative approach to measuring the liquidity risk of public investors' investment assets David Doran, Steve Kilkenny, Šarūnas Ramanauskas, and Alex Shablov, Central Bank of Ireland.....	197

Part 1

Central Bank Reserve Management

The strategic asset allocation of the investment portfolio in a central bank

Marco Fanari and Gerardo Palazzo¹

Abstract

The balance sheets of central banks (CBs) have changed greatly since the financial crisis, due to the growth of domestic and foreign assets in response to economic and policy developments. As a result, the risks borne by CBs have increased, both quantitatively and qualitatively. Many CBs have begun or resumed dealing with less traditional assets, markets and counterparties that have significantly altered their risk profiles, raising important implications for the risk management of the CBs' investment portfolio. The purpose of this paper is to describe the methodological aspects of a strategic asset allocation (SAA) framework that takes into consideration the perspective of a CB's full balance sheet. Here we refer specifically to a hypothetical Eurosystem CB with no exchange rate policy and whose monetary policy actions reflect its contribution to the setting of monetary and financial conditions to maintain price stability over the relevant horizon. The financial risk profile of such a CB is therefore shaped by risks arising from a policy portfolio resulting from monetary policy operations and the domestic investment portfolio. In this setting, we argue that an effective SAA framework should have three main characteristics: (i) an integrated view of all CB assets and liabilities; (ii) a wide and detailed representation of the investment universe; and (iii) a tailored objective function and constraints. This is a completely different paradigm from the standard view, where foreign reserves and other investments are considered as isolated components of a CB's balance sheet. Under the view put forward in this paper, quantitative tools used to optimise financial portfolios that ignore interdependencies with core policy functions offer only a limited insight on real risks faced by a CB.

¹ The authors are staff members of the Bank of Italy. This version: June 2019.

1. Overview

Investment portfolio management in central banks (CBs) has evolved substantially since the global financial crisis. The balance sheets of CBs have changed greatly due to the growth of domestic and foreign assets resulting from CB policy actions taken in response to economic and policy developments. As a result, the risks borne have increased, both quantitatively and qualitatively. Many CBs have begun or returned to dealing with less traditional assets, markets and counterparties that have significantly altered their risk profiles.

These developments have had important implications for the risk management of the CB's investment portfolio. As CBs have become more aware of overall risks on their balance sheet, risk management techniques have evolved in response to growing demands for higher returns, while keeping concessions in terms of liquidity and safety to a minimum. This is often achieved through the embrace of a strategic asset allocation (SAA) process in which key portfolio characteristics are defined.²

The purpose of this paper is to describe the methodological aspects of an SAA framework that takes into consideration the perspective of a CB's full balance sheet. Of course, CBs are all different, given their particular mandates and institutional settings, and here we refer specifically to a hypothetical Eurosystem central bank with no exchange rate policy and whose monetary policy actions reflect its contribution to the setting of monetary and financial conditions to maintain price stability over the relevant horizon. The financial risk profile of such a CB is therefore shaped by risks arising from a policy portfolio resulting from monetary policy operations (ie provision of CB money against collateral, outright purchases of securities, bank deposits) and the domestic investment portfolio (ie holdings of foreign exchange reserves, gold and domestic financial assets unrelated to monetary policy).

In this setting, we argue that an effective SAA framework should have three main characteristics: (i) an integrated view of all CB assets and liabilities, (ii) a wide and detailed representation of the investment universe and (iii) a tailored objective function and constraints. This is a completely different paradigm from the standard view, where foreign reserves and other investments are considered as isolated components of a CB's balance sheet. Under the view put forward in this paper, quantitative tools used to optimise financial portfolios that ignore interdependencies with core policy functions offer only a limited insight on real risks faced by a CB. Indeed, this framework is based on the supposition that the primacy of institutional objectives gives rise to unavoidable risks that the investment portfolio can help mitigate.

While this framework is discussed in the context of a hypothetical Eurosystem central bank, it is flexible enough to potentially incorporate features of CBs with different policy objectives and/or economic uses of foreign reserves and domestic portfolios, depending on stakeholder preferences. This wide applicability could be attained, for example, by imposing the necessary constraints in the portfolio construction process (eg through currency and liquidity requirements).

Moreover, given that CBs are exposed to fluctuations in the global economy and financial markets, the determination of the SAA requires estimating returns of a wide range of financial asset classes. The modelling framework should take into account

² Borio et al (2008a, 2008b); Bakker and van Herpt (2007).

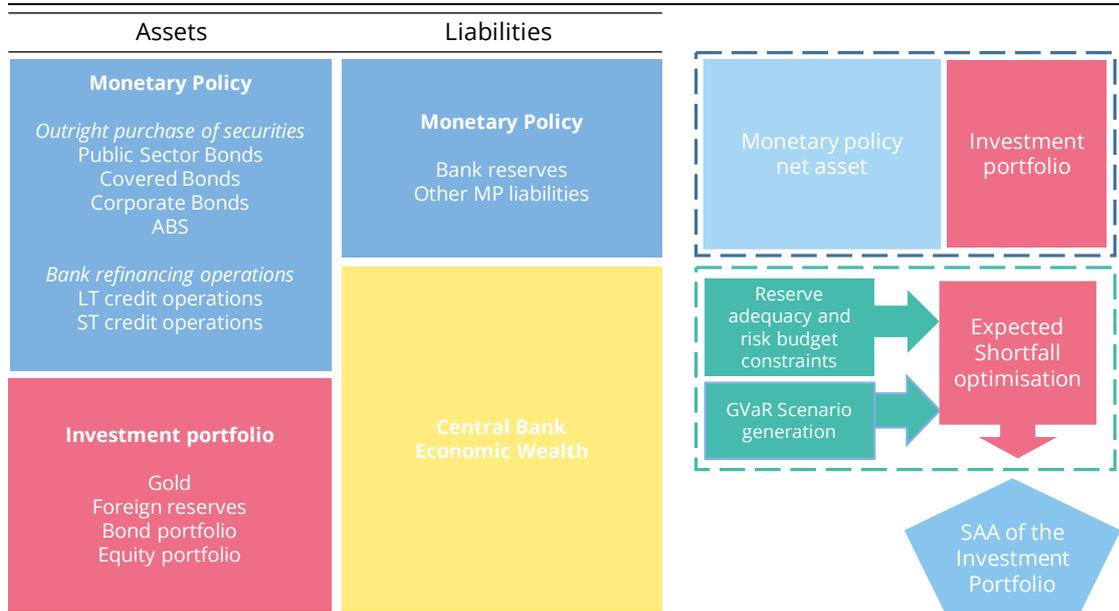
the co-movements that exist across markets and countries regarding the dynamics of output, inflation, foreign exchange rates, interest rates and equity prices. We use a global vector autoregressive (GVAR) model to simulate consistent scenarios for the economic and financial variables needed to evaluate the dynamics of the various components of the CB's balance sheet over the investment horizon.

CBs are also characterised by a peculiar asymmetric utility function: while risk-averse in normal times, they are much less so under extraordinary circumstances, when they must be prepared to take on significant risks. While many private investors might focus on performance in normal times, there is a broad consensus that CBs should be most concerned about portfolio performance during periods of stress. The rationale is that the health of the CB financial structure has to be robust especially in those adverse circumstances in which its institutional functions may lead to taking exceptional risks. It is during these times when, ideally, the investment portfolio should be a source of strength, not a factor that could further jeopardise the CB's financial position. The SAA problem of the investment portfolio is therefore solved by minimising the expected loss in the CB economic wealth (defined as the sum of investment and policy portfolios, both valued at market prices, net of bank reserves and other monetary policy liabilities) over the most adverse scenarios at the end of a long-term investment horizon. The optimisation is subject to a set of short-term constraints, which serve to temper the long-term nature of the SAA and to account for reputational risks. These constraints essentially control the risk associated with accounting losses and depletion of financial capital over a 12-month horizon. The overall scheme of our approach is represented in Figure 1.

Finally, we note that the SAA framework allows some extensions. For example, the CB balance sheet coverage could be broadened to include some implicit items, ie latent assets and/or contingent liabilities, that arise from a CB's policy mandates. These peculiar items are linked to a CB's institutional functions, such as banknote issuance and financial stability, implying an exposure to unavoidable risk factors, with potentially important implications for investment portfolio management and strategies. In this broader setting, economic wealth could be broadened to include the present value of future income from banknote growth (ie the interest costs saved because of the CB's power of issuing such a non-interest-bearing liability) and the contingent liability related to the lender of last resort function of the CB (which may imply the support to illiquid but solvent domestic banks). While these extensions are not explicitly included in the framework presented in this paper, we discuss them as exploratory ideas.

The remainder of the paper is organised as follows. In Section 2, we explain the importance of a CB having an integrated view of all financial risks in a CB's operations in order to define an effective SAA (eg we examine the role of foreign reserves holdings as a tool to hedge financial stability risks). In Section 3, we introduce the GVAR model used for the scenario generation process, and in Section 4 focus on the chosen portfolio optimisation technique. In Section 5, we present an approach for incorporating implicit assets and liabilities in the SAA analysis before concluding in Section 6 with a summary of the stylised characteristics of the optimal SAA under our framework.

Integrating the policy & the investment management functions of a central bank Figure 1



Source: Authors' example.

2. Integrated view

As in the case of other economic agents, the goal of the SAA in a CB is to find an optimal allocation for the investment portfolio, defined as foreign official reserves and domestic financial assets unrelated to monetary policy, across different asset classes in a way that reflects a specific combination of long-term risk-return preferences, general investment constraints and possibly the exposure to unavoidable policy-related risk-return factors.³ For CBs, important portfolio management implications arise from their unique policy mandates, organisational structure and eligible financial assets. A thorough assessment of these aspects then becomes crucial to provide a rational and consistent basis for determining the SAA.

Bindseil et al (2009) present a comprehensive and conceptually structured framework for the risk management in CBs. The authors explicitly address the need to overcome the widespread practice of segregating CB risk management tools between investment and policy portfolios; they review CB practices in the SAA area and briefly describe the European Central Bank's (ECB) approach based on a highly sophisticated optimisation model. Yet, they refrain from introducing a comprehensive quantitative framework, whereby the SAA is contingent on the core policy function of a CB. More recently, the importance of having a risk management framework that enables a comprehensive and consolidated view of all relevant risks in a CB has been

³ For example, in the case of private investors, a major exposure to a non-alienable risk-return factor is normally related to their "human wealth", defined as the present discounted value of their future expected labour earnings (Campbell and Viceira (2002)).

highlighted as a cornerstone principle of the IMF's revised *Guidelines for foreign exchange reserve management* (2013).⁴

An example of interrelation between policy functions and SAA: foreign reserves as a tool to hedge financial stability risks

In this section, we review the main theoretical motivations for holding foreign reserves for our hypothetical Eurosystem CB, and summarise an approach recently proposed in the academic literature that could be used to assess reserve adequacy. This adequacy calculation can then be formally integrated into the SAA framework by serving as a constraint in the objective function.

Obstfeld et al (2008) present a potentially useful framework for assessing reserve adequacy for our hypothetical Eurosystem central bank. While many reserve adequacy approaches emphasise negative balance-of-payments shocks (ie capital outflows), which can take place when the purchase of domestic assets by foreigners suddenly stops, Obstfeld et al suggest that similar shocks can also arise when purchases of foreign assets by domestic residents suddenly surge. As an example of this dynamic, some types of banking crisis feature domestic capital flight through a drain of domestic bank deposits, producing a dynamic that is line with the literature on the interaction between banking crises and currency crises (twin crises). In such a case, a flight from domestic bank deposits into foreign exchange puts the domestic banking system and the exchange rate under extreme pressure, and may eventually require the CB to step in as lender of last resort (double drain risk).⁵

The global financial crisis has highlighted the importance for commercial banks, including those in developed economies, to have access to adequate facilities for coping with foreign currency needs, ie foreign reserves held by the CB or through other currency arrangements.⁶ In the even broader view taken by Obstfeld et al, the need for more foreign reserves, beyond the trade and the external debt arguments, is mainly ascribed to the overall short-term liabilities of the domestic banking system, proxied by the M2/GDP ratio.⁷ The monetary aggregate M2 represents the bulk of the domestic assets that could be easily sold and reallocated in foreign assets by the domestic private sector (double drain scenario).

The approach proposed by Obstfeld et al assessing the adequacy of the amount of official reserves could be integrated in the SAA framework via an explicit lower

⁴ Item no 36 in the *Revised guidelines for foreign exchange reserve management* (IMF (2013)).

⁵ Calvo (1996), Calvo and Mendoza (1996) and Calvo (2006).

⁶ During the crisis, central banks cooperated to set large swap lines to manage a significant US dollar shortage. The credit facilities (currency swap lines) have been granted by the Federal Reserve to the major central banks to provide US dollar-denominated credit to domestic banks under funding pressure (Obstfeld et al (2009)).

⁷ Obstfeld et al assess reserve adequacy by estimating a panel regression in which the reserves to GDP ratio (expressed in logarithm) is explained by the following variables: (i) the M2/GDP ratio (expressed in logarithm) as a proxy of the size of the banking system and, more generally, of the financial development of the country; (ii) a measure of financial openness, based on a (properly scaled) index, defined by Edwards (2007); (iii) two (mutually exclusive) dummies for the exchange rate flexibility, soft pegging and hard pegging, as suggested by Shambaugh (2004); (iv) a dummy variable for the advanced economies; and (v) a measure of the original sin, ie the inability of a country to borrow abroad in its own currency, proxied by the internationally issued securities issued in foreign currency, as suggested by Eichengreen et al (2005).

bound in the optimisation process for our hypothetical Eurosystem central bank. Of course, alternative measures of reserve adequacy, such as the Guidotti-Greenspan rule that looks to a country's short-term external debt position, could be more appropriate for CBs outside the Eurosystem context.

3. Estimating return distributions for investment and policy portfolio assets

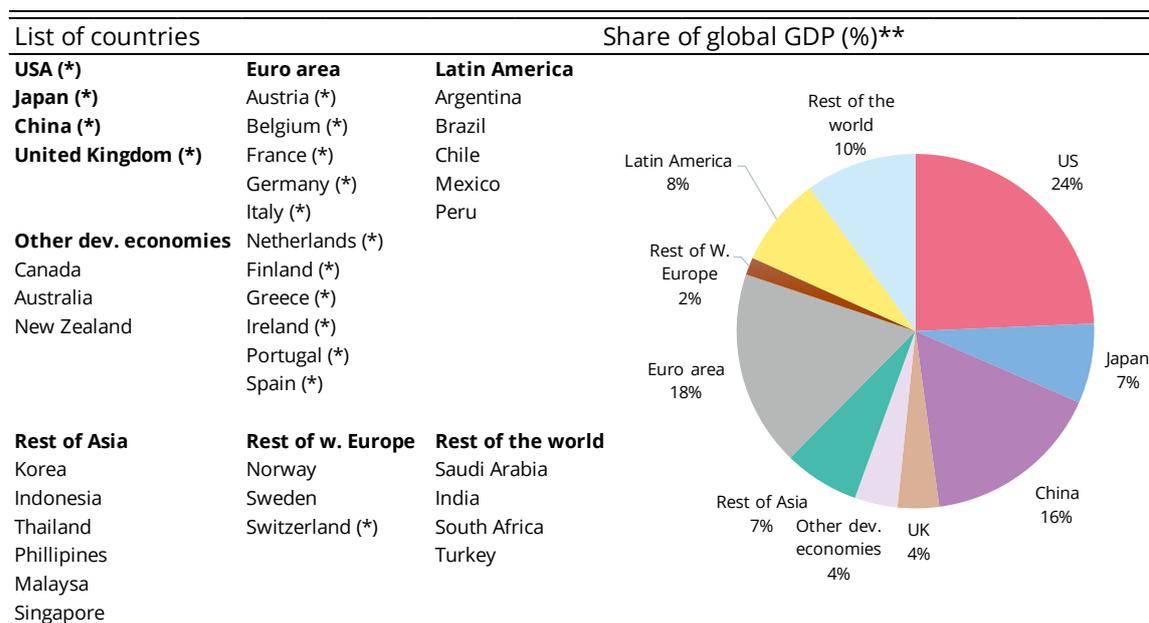
The SAA of the investment portfolio requires estimating the expected return distributions of a potentially wide range of financial asset classes across many currencies. This means that a modelling framework has to be employed that takes into account the interdependencies and co-movements across output, inflation, foreign exchange rates, interest rates, equity prices and any other relevant variables. This section discusses the modelling of asset return distributions across the central bank balance sheet, including foreign (marketable) assets, domestic marketable assets for policy or investment purposes, and domestic non-marketable assets for policy purposes.

The simulation of the financial asset returns

We employ the GVAR model, originally introduced in Pesaran et al (2004) and further developed in Dees et al (2007), to simulate financial asset returns. It is a compact global econometric model that is capable of generating simulations (ie density forecasts) for a core set of macroeconomic and financial variables across a large number of countries in a consistent manner.⁸

Building on the studies of Pesaran et al (2004) and Dees et al, we extend the model along two dimensions: (i) the geographic coverage is broadened to also include Ireland, Greece and Portugal, hence estimating a model with 36 countries (Figure 2); and (ii) corporate spreads and the gold price are added to the financial variables. At the end of 2011, the output from the 36 countries accounted for around 80% of world GDP. The US economy represented one quarter of the total GDP (in power purchase parity) from the 36 countries, followed by the euro area countries (18%), China (16%), Latin America countries (8%, with Brazil and Mexico as the main contributors). Details on the specification and estimation of the model are reported presented in Annex 1.

⁸ Implementation of this model by practitioners has been supported by the availability of code programmes for estimation and statistical testing. To estimate the model, we used the GVAR toolbox developed by the Centre for Financial Analysis & Policy of Cambridge Judge Business School (University of Cambridge, Smith and Galesi (2011)). We developed our own routines to perform scenario simulations.

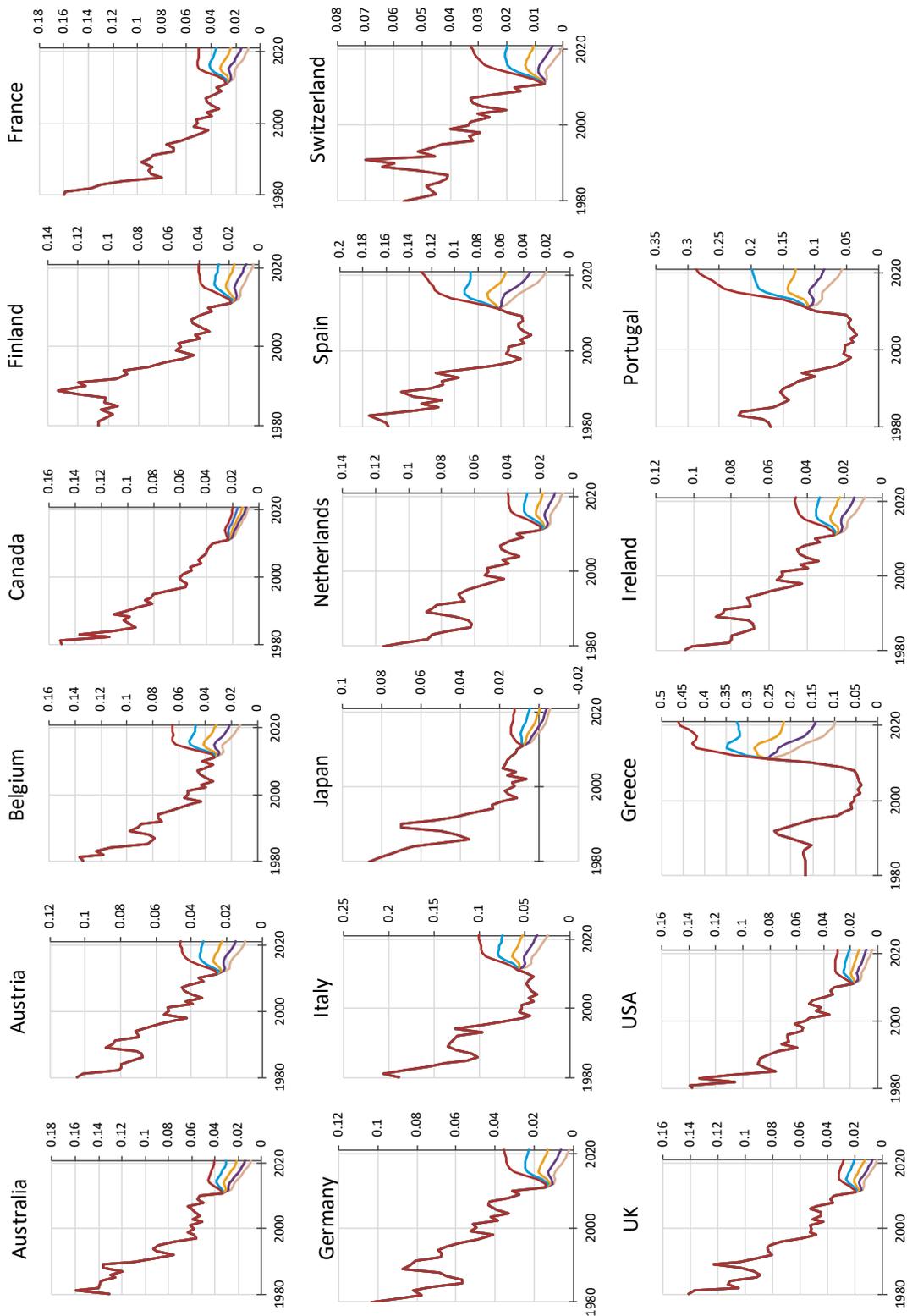


*Countries considered for the scenario-generating process. ** In USD, PPP-adjusted.

Source: Authors' calculations on World Bank data for 2011.

Our aim is to simulate the evolution of the foreign exchange rates, equity indices, money market interest rates, long-term government yields and corporate spreads needed to calculate the rate of returns for a plausible asset class universe for the investment portfolio of our hypothetical Eurosystem central bank. We also need to simulate the dynamics of the various monetary policy-related components of the economic wealth of the CB. For this purpose, we generate 10,000 scenarios over a 10-year horizon, drawing from the joint distribution of the residuals of the GVAR model.

Figure 3 shows the average and median value, the 75th and 90th percentile calculated on trajectories of 10-year government bond yields for a sample of selected countries. Based on the scenarios generated by the GVAR model, we calculate the financial asset rate of returns.



Note: Actual values from Q1 1980 to Q3 2012; simulated values from Q4 2012 to Q4 2022: average value, median value, 10, 25, 75 and 90th percentile.

Sources: Authors' calculations.

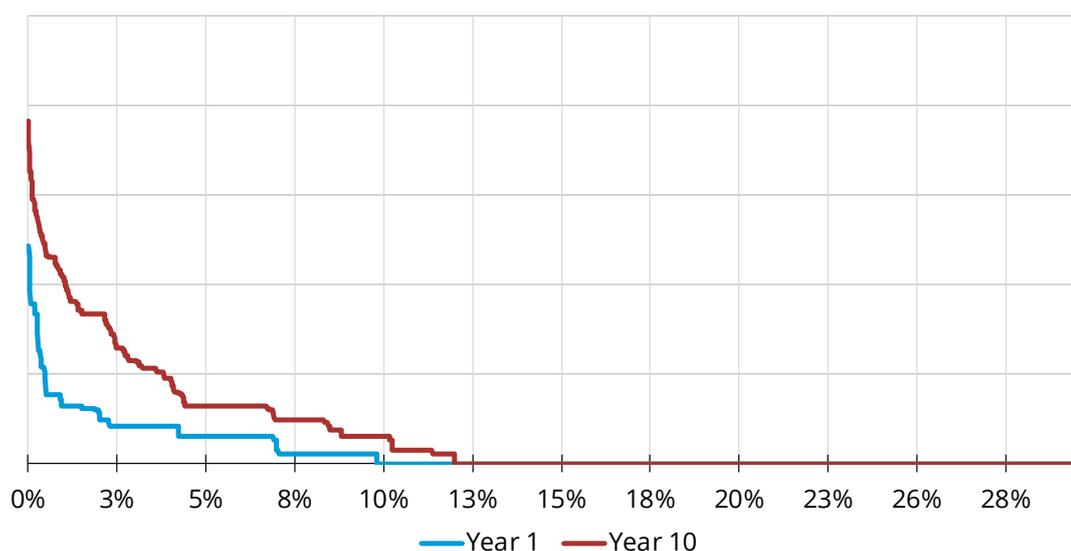
The simulation of losses on the monetary policy credit operations

The above-mentioned GVAR model covers all the market risk factors that are relevant for the dynamics of the investment portfolio and those marketable domestic assets held for monetary policy purposes. However, in the case of monetary policy credit operations (mainly collateralised loans to banks), market prices are not readily available. In fact, the interest rate earned by the CB is a policy rate that does not reflect the credit risk of the counterparty and the collateral quality (although these aspects are considered carefully in the collateral management framework through eligibility criteria, haircuts, margins and add-ons). Therefore, in order to estimate the loss distribution on the monetary policy credit operations, we directly simulate the event of default for counterparties and collateral issuers (*double-default* model).

For this specific purpose, we take the approach of Pesaran et al (2006), in which an event of default is assumed to occur when the value of firm (counterparty bank and collateral issuer) assets falls below the value of a specific threshold determined on the basis of the credit rating. The dynamics of the value of the firms' assets are simulated through the GVAR. The default distribution at a future point in time, for both counterparties and collateral issuers, is obtained by comparing the simulated asset values with thresholds in different scenarios. Assuming a constant loss given default, the final output is the estimation of portfolio credit losses (Figure 4).

Expected portfolio losses (EPL) at different time horizons and percentile values*

Figure 4



*EPL not disclosed on the vertical axis for confidentiality reasons.

Sources: Authors' calculations.

Further details on the integration of the expected monetary policy credit losses and the scenarios generated within the GVAR framework are reported in Box 1.

Credit risk in monetary policy refinancing operations

The approach is aimed at explicitly linking the loss distribution of the monetary policy credit portfolio (that mainly represents a portfolio of collateralised loans to euro area banks) to the evolution of the macroeconomic and financial variables generated by the GVAR model. This setting closely follows Pesaran et al (2006), and is essentially a Merton-type credit model where a firm is expected to default when the value of its assets (V) falls below a threshold value determined by its liabilities (D).

In most empirical applications, the default threshold is taken from balance sheet data (typically short-term debt plus a proportion of long-term debt) while the parameters governing the asset value dynamics are derived by jointly taking into account the default threshold, the market capitalisation of the equity of the firm and the volatility of equity returns. Pesaran et al (2006) apply an alternative, simplified, approach: (i) the default threshold is determined using credit ratings and the equity value dynamics (instead of balance sheet data) and (ii) the firm's asset value dynamics are entirely determined by its equity returns, modelled within the GVAR framework.

Default thresholds are expressed in terms of percentage ratios and indicate the maximum drop that a firm's equity value would be able to sustain before the firm defaults on its obligations. The likelihood of this event is driven largely by the firm's equity returns volatility. We calibrate default thresholds using the long-term historical default frequencies of different rating classes as well as the mean and volatility of equity returns.

We simulate the equity returns for each entity (counterparty and collateral issuer) using (i) equity market returns for each country through the GVAR model (systematic component) and (ii) a firm-specific shock (idiosyncratic component) for each counterparty and collateral issuer.

Defaults occur when the simulated return of the firm's equity value falls below the threshold-equity ratio. Thus, in each scenario, firms will be in a state of "default" or "non-default". Defaults are simulated for both bank counterparties and collateral issuers. Moreover, the collateral issuer state of default or non-default becomes relevant only when the counterparty defaults.

Applying this method to each entity, the default losses are calculated by obtaining the cumulative loss distribution of the credit portfolio over time.

4. Objective function and constraints

While there are no obvious criteria for establishing the appropriate trade-off between risk and return, there is a broad consensus on the desirability of constructing central bank portfolios over a long-term horizon and with a *conservative bias*. In some contrast with private investors, who typically focus on performance in normal times, CBs are most concerned about portfolio performance during periods of stress. This reflects the main objective of a CB's SAA framework: to preserve the value of financial resources required to effectively pursue public policy functions in an independent manner and in any circumstance, that is to say over the long run and especially in

adverse scenarios.⁹ This also helps eventually to entrench expectations among both the general public and financial market participants that the CB will not be unduly constrained in the pursuit of its monetary policy and financial stability objectives by concerns over financial resources.

Consequently, a portfolio optimisation approach that aims at minimising risks of losses in adverse scenarios has been preferred to the standard mean-variance framework, where the focus is the minimisation of symmetric risks around an expected outcome. We instead focus on minimising economic wealth over the adverse scenarios.

The objective function of the optimisation

More specifically, the SAA problem of the investment portfolio is solved by minimising the expected loss in the CB economic wealth (investment and policy portfolios, both valued at market prices, net of bank reserves and other monetary policy liabilities) over the most adverse scenarios at the end of a long-term investment horizon. The expected shortfall (ES) metric becomes the natural choice to serve as the primary risk measure under which to construct the SAA.

Although the optimal SAA has a long-term goal, investment policy considerations make it necessary to impose some constraints on portfolio risks in the short term. In this regard, a crucial issue, inevitably faced by CBs, is reputational risk; more specifically, CBs have an above-average sensitivity to capital depletion, to accounting losses and to the range of financial instruments that could be potentially held in the investment portfolio (investment universe). Consequently, the optimisation is subject to a set of short-term constraints that serve to temper the long-term nature of the SAA given these reputational risks. More specifically, these constraints control the risk associated with accounting losses and the depletion of financial capital over a 12-month horizon.

Formally, the SAA problem is specified as a multi-stage stochastic programming model (MSP).¹⁰ There are several reasons why we think the MSP approach is superior to standard mean-variance optimisation (MV). In the first place, the MV framework is based on variance, a measure of risk that cannot accurately handle the non-normal (fat-tailed and/or skewed) return distributions as might be empirically observed, especially for credit risk-sensitive instruments. Moreover, the quadratic utility function used in the MV framework may be inappropriate even with normal distributions, as it counts for positive and negative returns equivalently, while the objective function of a conservative investor should be more sensitive to downside risk. Value-at-risk (VaR) metrics have been widely used as alternative risk measures to overcome this problem, but they suffer from several drawbacks that limit their use in portfolio optimisation applications.¹¹ These shortcomings are not shared by the ES metric, which can be easily handled within the MSP approach.

⁹ For two different perspectives on why CBs need to preserve their financial strengths, see Sims (2003) and Cukierman (2011).

¹⁰ See Zenios (2007).

¹¹ Since VaR is defined as the expected loss at a given confidence level, it does not give information about the expected losses beyond that confidence level. Losses beyond the VaR threshold may be important especially for non-normal distributions. Theoretical shortcomings of VaR are: (i) it is not

More importantly, the MV framework is essentially myopic, in the sense that the investor groups all returns (which may represent various points in the future) into a single period and identifies an optimal buy-and-hold portfolio for this single period. This strategy simply specifies the proportions of initial wealth invested across the assets in the investment universe and these assets are then held till the end of the investment horizon under all scenarios. The MSP approach offers the significant advantage of considering the investment strategy in a truly dynamic and multi-period context: the portfolio is adapted over time (at specific dates in a discrete time setting) following clearly defined rules. For complex optimisation problems, such as defining the SAA of a CB, MSP may also be viewed as a superior alternative to dynamic programming, which is often associated with the “curse of dimensionality.”¹²

The flexibility of the MSP approach may also help to cope with the fact that, under the investment policy of the CB, the management style for certain assets might be more accurately represented by a buy-and-hold strategy. For context, the SAA of a CB is generally defined using the constant-mix rule, where the portfolio is rebalanced annually to the optimal SAA.¹³ For example, if the prices of the asset classes in the portfolio change (as would the corresponding weights in the portfolio), the portfolio would be rebalanced by selling (buying) the over (under) weighted ones to re-obtain the SAA mix. The MSP approach offers the advantage of being able to impose the constant mix rule only to a certain part of the portfolio, for example, the ex-gold portfolio.

The constraints of the optimisation

The risk-return preferences of the CB and other investment policy issues have been incorporated in the SAA optimisation process via several constraints:

1. A lower bound for the foreign reserves, including gold, that incorporates the considerations of the reserve adequacy analysis addressed in Section 2;
2. A constraint on the asset weights that aims to keep under control short-term financial risks (*financial risk constraint*);
3. A constraint to control the risk of the portion of the investment portfolio whose accounting valuation is based on market prices (*accounting risk constraint*).

The purpose of constraints 2 and 3, both calculated over a one-year horizon, is to define a risk tolerance for the SAA. The financial constraint requires that the expected losses on the investment and policy portfolios, measured with the ES metric at the 99th percentile over a 12-month horizon (one-year ES 99%), do not exceed the *financial capital allocated to risks* (Figure 5). This is calibrated by subtracting from the financial capital a desired amount that must be preserved to cover extreme risks, ie those arising from events beyond the 99th percentile of the occurrences (unallocated

sub-additive (diversification among financial assets may actually increase VaR rather than decrease as conventional portfolio theory would suggest); and (ii) it is, in general, non-convex, which causes great practical difficulties in optimisations applications due to possibly multiple local minima.

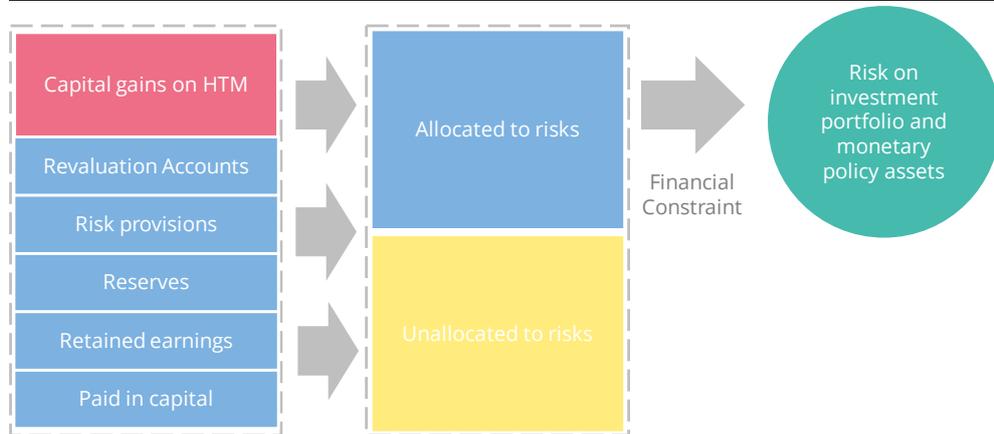
¹² The dynamic programming framework, both in discrete and continuous time specification, has been of limited practical value for institutional enterprise-wide risk management. The assumptions regarding the utility function and the asset price dynamics are usually restrictive. The framework may even ignore practical constraints such as trading restrictions that are typically imposed by corporate policies and operational requirements (Zenios (2007)).

¹³ See Zenios (2007) for a discussion of how the constant mix rule is consistent with the logic that, in a multi-stage context, the portfolio decisions must not depend on *clairvoyance* but only on current and past information (CD non-anticipativity constraint).

financial capital). Financial capital is defined as the difference between assets and liabilities valued at market prices or, equivalently, the accounting capital, including the revaluation accounts, plus the capital gains/losses on those assets that are valued at historical cost.

Financial risk control constraint

Figure 5



Source: Authors' example.

As far as the accounting risk constraint is concerned, CBs' typically high sensitivity to reputational risks calls for the minimisation of the probability of a negative result in the annual profit and loss. This constraint is specified by imposing the constraint that the difference between (i) the expected losses (one-year ES 99%) on the investment portfolio assets whose accounting valuation is based on market prices (such as gold, foreign exchange reserves, equities and marketable bond not held to maturity) and (ii) their respective revaluation accounts, must not exceed the general risk provisions.

5. A practical approach for incorporating CB off-balance sheet assets and liabilities

Important peculiarities arise from the unique policy mandates of CBs, fundamentally altering their risk profile relative to private sector institutional investors. Two important aspects include:

1. The contingent liability related to the *lender of last resort* (LoLR) function, ie the risk exposure that may arise in the future from providing liquidity to solvent banks in adverse scenarios;
2. The present value of *future income from banknote growth*, ie the interest saved by the CB on the potential increase of the stock of banknotes, which is an unremunerated and irredeemable liability.

These two peculiar implicit items – both arising from policy mandates – are extremely important for the SAA, as they define an exposure to unavoidable risk factors and influence the characteristics of the CB risk-return preferences. For these reasons, they could be considered as possible extensions to the SAA framework proposed in this paper: the balance sheet coverage – and the concept of the

economic wealth to be used in the optimisation process – could be consequently broadened to include these implicit items, in addition to the policy and the investment portfolios.

A practical approach could be as follows. For each implicit item, its economic value could be captured through a so-called replicating portfolio, ie a combination of asset classes whose marked-to-market value changes closely track those of the implicit item in different scenarios. This approach allows a CB to use portfolio-related metrics typically employed in risk management when integrating these unique risk exposures into the SAA framework.

Contingent liabilities related to the financial stability function

As LoLR to the domestic financial institutions, CBs have a responsibility for the provision of CB money and/or other assistance to solvent entities facing temporary liquidity problems that may lead to an increase in CB money (eg emergency liquidity assistance, or ELA, in the case of our hypothetical Eurosystem central bank). In a fully integrated risk-management framework, the unique risk-return profile potentially arising from these actions should be captured as part of the SAA process. Indeed, this is another reason why the financial stability function should be central to the SAA process for the investment portfolio.

A simple scheme may help to differentiate the emergency actions to support financial stability between: (i) standard lending facilities used for monetary policy operations (available to banks on their own initiative and requiring normal collateral); (ii) the conventional LoLR function of the CB, which would be against a possibly wider range of collateral and require a solvency test; and (iii) measures for likely insolvent banks (guarantees, capital injections etc). Measures under point (iii) fall clearly beyond LoLR support and outside the remit of a CB, whose involvement in financial stability issues is limited to the first two items.

As indicated previously, the risks related to the financial stability function arise exclusively from the LoLR function, whose aim is to reduce the systemic risk of a temporary liquidity crisis of a solvent bank. In these cases, a loss for the CB arises only if all following conditions jointly occur: (i) the illiquid bank becomes insolvent and does not return the full amount borrowed; (ii) the collateral value is less than the CB loan; (iii) the bank liquidation does not allow recovering the full loan amount. In a nutshell, at least in normal conditions, the risks associated with liquidity assistance may be considered close to zero. However, when facing a large bank whose capital strength is highly uncertain and the collateral quality is in the lower part of the acceptability range, risks of incurring into losses are not negligible anymore.

With this perspective in mind, the evaluation of the contingent liability related to expected losses on possible LoLR operations could be undertaken, for example, using the standard tools of contingent claim analysis (CCA). More specifically, one could identify a combination of financial options that would serve as a replicating portfolio, ie a combination of assets whose marked to market value tends to mimic the liability value under different scenarios (ie a financial stability replicating portfolio).¹⁴ This

¹⁴ The risk-free assets component of this portfolio could be modelled as a position in long-term government bonds with low and stable risk premiums, while the assets component could be

replicating portfolio would then be incorporated in the economic wealth portfolio that is the object of the optimisation procedure.¹⁵

The present value of future income from banknote growth

The CB is the public institution in charge of issuing banknotes, which are a non-interest-bearing and irredeemable liability. The amount of banknotes that the private sector wishes to hold changes over time; in the most popular money-demand functions (ie a Cagan-type function), it is positively correlated with nominal income (because of the increase in total transaction volume) and negatively with interest rates (because of the higher opportunity costs of holding banknotes).¹⁶

In an integrated SAA framework, we can take into account the present value of the expected stream of implicit revenues – the interest costs saved resulting from the issue of an unremunerated liability – that is associated with the time-varying stock of banknotes in circulation, and the impact on the overall risk profile of the CB.

The estimate of future interest savings poses a number of tricky challenges. First, the dynamics of future banknote growth are highly uncertain, and depend on the money demand function of the private sector and possibly on technological innovation. Second, an interest rate has to be selected to discount the stream of future increases in banknotes. Finally, a CB can legitimately count on this expected stream of revenues only insofar as it is compatible with price stability and the integrity of the monetary system; therefore the estimate has to be constrained to be consistent with the monetary policy objective of the central bank. These complexities are reflected in the wide range of estimates in the literature. For example, with reference to the euro zone, Buiter and Rahbari (2012) have a central estimate of around 20% of GDP (35% in 2015).

Including the present value of future income from banknote growth in an SAA framework also calls for its representation via a replicating portfolio. As in the previous case of the contingent liabilities related to the financial stability function, this would be a combination of asset classes whose mark-to-market value tends to mimic the systematic component of the present value of future income under different scenarios. For this purpose, a *style analysis* (or similar technique) could be employed in order to identify a suitable replicating portfolio.¹⁷ Thereafter, it could be fully incorporated in the economic wealth portfolio.¹⁸

modelled as a position in corporate bonds. For a general reference on how to use the CCA analysis for these practical purposes see Merton (1977) and Gray, Merton and Bodie (2007).

¹⁵ Using the scenario generation model (see Section 3) it would be possible to project the value of the financial stability replicating portfolio (proxy of the CB contingent liability) over a 10-year horizon and calculate its probability distribution at various future points in time.

¹⁶ See Cagan (1956).

¹⁷ The present value of future income of banknote growth could be expressed in terms of a combination of money market instruments, long-term bonds and equities (Sharpe (1992)).

¹⁸ Using the scenario generation model (see Section 3) it would be possible to project the value of this replicating portfolio (proxy of a CB latent asset) over a 10-year horizon and calculate its probability distribution at various points in time into the future.

6. Concluding remarks

In our full balance sheet approach, the optimal SAA of the CB's investment portfolio is constructed in light of its natural exposure to systemic and business cycle risk stemming from its core policy functions. In a sense, the SAA is a minimum risk portfolio adopting a long-run perspective and taking account of all the foreseeable factors that could determine the CB's wealth, including the risks stemming from institutional functions. The rationale is that the CB's financial structure has to be robust, especially in those adverse circumstances in which its institutional functions may lead to taking exceptional risks.

Ceteris paribus, this natural exposure to systemic and business cycle risk is the main reason that should lead a CB to consider countercyclical, low-credit risk hedging assets for inclusion in the SAA. Such a minimum-risk approach tends to produce a conservative SAA providing low returns in normal times. Although this aspect needs to be carefully considered, our view is that such an SAA is consistent with the *conservative bias* that should motivate risk management at CBs, where rewards from high returns carry a lower weight than the costs from reported losses, especially if they come from the management of the investment portfolio.

Moreover, the annual rebalancing rule involves a *contrarian strategy* (selling asset classes that have risen in price and buying those that have declined) which, provided that asset returns mean-revert, should enhance the risk-adjusted performance of the portfolio over the long run.

Finally, in taking a forward-looking and long-term (through-the-cycle) approach with clearly defined portfolio rebalancing rules, the CB's optimal SAA also plays a fundamental role in minimizing the procyclicality of CB portfolio management.¹⁹ The exposure toward countercyclical assets helps reduce selling pressure in time of crisis, when these assets tend to appreciate. According to research by IMF, at the beginning of the crisis in 2008, CB portfolio managers, concerned about increased credit and liquidity risks, lowered their risk exposure in the investment portfolio, especially by reducing the amount of money held in short-term deposits with commercial banks.²⁰ These reactions were inconsistent with the large volumes of liquidity provided to banking systems through monetary refinancing operations and may also have had unintentional signalling effects to market participants, exacerbating market turmoil. In any case, such procyclical behaviour sharply contrasts with what should be the typical CB conduct during a crisis.

¹⁹ Papaioannou et al (2013).

²⁰ According to Pihlman and van der Hoorn (2010), CB reserve managers joined the flight to quality and collectively pulled out more than USD 500 billion dollars of deposits and other investments from the banking sector from December 2007 to March 2009. McCauley and Rigaudy (2011) document how exposure to government-sponsored enterprises was reduced and securities lending programmes were scaled back. Reserve managers' investment in assets under market pressure was also reduced when credit rating downgrades breached the minimum level for inclusion.

References

- Bakker, A and I van Herpt (2007): *Central bank reserve management – new trends, from liquidity to returns*, Edward Elgar Publishing Limited.
- Bindseil, U, F Gonzales and E Tabakis (2009): *Risk management for central banks and other public investors*, Cambridge University Press.
- Borio, C, G Galati and A Heath (2008a): "FX reserve management: elements of a framework", *BIS Papers*, no 38, March.
- (2008b): "FX reserve management: trends and challenges", *BIS Papers*, no 40, May.
- Buiter, W and E Rahbari (2012): "Looking into the deep pockets of the ECB", *Global Economics View*, Citi Economics, 12 February.
- Cagan, P (1956): "The monetary dynamics of hyperinflation", in M Friedman (eds) *Studies in the quantity theory of money*, University of Chicago Press.
- Campbell, J, A Low and A MacKinley (1997): *The Econometrics of Financial Markets*, Princeton University Press.
- Campbell, J and L Viceira (2002): *Strategic asset allocation – portfolio choice for long-term investors*, Oxford University Press.
- Calvo, G (1996): "Capital flows and macroeconomic management: tequila lessons", *International Journal of Finance and Economics*, vol 1, no 3, July, pp 207–23.
- (2006): "Monetary policy challenges in emerging markets: sudden stop, liability dollarization, and lender of last resort", *NBER Working Papers* no 12788.
- Calvo, G and E Mendoza (1996): "Mexico's balance-of-payments crisis: a chronicle of a death foretold", *Journal of International Economics*, vol 41, no 3–4, November, pp 235–64.
- Cukierman, A (2011): "Central bank finances and independence – how much capital should a central bank have?", in S Milton and P Sinclair (eds) *The capital needs of central banks*, Routledge Publishing House.
- Dees, S, F di Mauro, M Pesaran and L Smith (2007): "Exploring the international linkages of the euro area: a global VAR analysis", *Journal of Applied Econometrics*, no 22.
- Edwards, S (2007): "Capital controls, sudden stops, and current account reversals", in S Edwards (eds) *Capital controls and capital flows in emerging economies: policies, practices, and consequences*, University of Chicago Press.
- Eichengreen, B, R Hausmann and U Panizza (2005): "The pain of original sin", in B Eichengreen and R Hausmann (eds) *Other peoples' money*, University of Chicago Press, pp 13–47.
- Glick, R and A Rose (1999): "Contagion and trade: why are currency crises regional?", *Journal of International Money and Finances*, no 18, pp 603–17.
- Gray, D, R Merton and Z Bodie (2007): "New framework for measuring and managing macrofinancial risk and financial stability", *NBER Working Paper*, no 13607, November.

International Monetary Fund (2013): *Revised guidelines for foreign exchange reserve management*, Washington, DC.

McCauley, R and J-F Rigaudy (2011): "Managing foreign exchange reserves in the crisis and after", *BIS Papers*, no 58, October, pp 19–47.

Merton, R (1977): *An analytic derivation of the cost of deposit insurance and loan guarantees: an application of modern option pricing theory*.

Obstfeld, M, J Shambaugh and A Taylor (2008): "Financial stability, the trilemma, and international reserves", *NBER Working Papers*, no 14217, August.

——— (2009): "Financial instability, reserves, and central bank swap lines in the panic of 2008", *NBER Working Papers*, no 14826, March.

Papaioannou, M, J Park, J Pihlman and H van der Hoorn (2013): "Procyclical behavior of institutional investors during the recent financial crisis: causes, impacts, and challenges", *IMF Working Paper*, no 193.

Pesaran, M, T Schuermann and S Weiner (2004): "Modeling regional interdependencies using a global error correcting macroeconometric model", *Journal of Business and Economic Statistics*, vol 22, no 2, April.

Pesaran, M, T Schuermann, B-J Treutler and S Weiner (2006): "Macroeconomic dynamics and credit risk: a global perspective", *Journal of Money, Credit and Banking*, vol 38, no 5.

Pihlman, J and H van der Hoorn (2010): "Procyclicality in central bank reserve management: evidence from the crisis", *IMF Working Paper*, no 150.

Scherer, B (2007): *Portfolio construction and risk budgeting*, Risk Books, 3rd edition.

——— (2008): *Portfolio Management*, Risk Books.

Shambaugh, J (2004): "The effects of fixed exchange rates on monetary policy", *Quarterly Journal of Economics*, vol 119, no 1, pp 301–52.

Sharpe, W (1992): "Asset allocation: management style and performance measurement", *Journal of Portfolio Management*, vol 18, no 2, winter, pp 7–19.

Siegel, J (2004): "The long-run equity risk premium", *CFA Institute Conference Proceedings*.

Sims, C (2003): *Fiscal aspects of central banks independency*, Department of Economics, Princeton.

Smith, L and A Galesi (2011): *GVAR Toolbox 1.1 User Guide*, July.

Zenios, S (2007): *Practical financial optimization: decision making for financial engineers*, Blackwell.

Annex 1 – The specification and estimation of the global vector auto regressive model

This annex describes the approach we used to specify and estimate the global vector auto regressive (GVAR) model. The GVAR is specified by combining 36 country-specific error correction models (ECMs). For each country model, the domestic economic and financial variables are linked to the corresponding country-specific foreign variables, constructed by weighting the domestic variables of the other 35 countries using bilateral trade weights. The GVAR modelling approach provides a general yet practical global modelling framework for the scenario generation process.

Data

Domestic variables

Most country specific models include the following domestic variables:

Real GDP	$y_{it} = \ln(GDP_{it} / CPI_{it})$	Inflation rate	$\pi_{it} = \ln(CPI_{it} / CPI_{it-1})$
Real equity index	$q_{it} = \ln(EQ_{it} / CPI_{it})$	Short-term (ST) interest rate	$\rho_{it}^S = 0.25 \times \ln(1 + R_{it}^S / 100)$
Long-term (LT) interest rate	$\rho_{it}^L = 0.25 \times \ln(1 + R_{it}^L / 100)$,	Real exchange rate	$e_{it} = \ln(E_{it} / CPI_{it})$
Corporate spread	$cs_{it} = 0.25 \times \ln(1 + CS_{it} / 100)$		

where GDP_{it} is the nominal gross domestic product, CPI_{it} the consumer price index, EQ_{it} the nominal equity price index, E_{it} the exchange rate in terms of US dollars, R_{it}^S the short rate, and R_{it}^L the long rate of interest, for country i during the period t . Due to insufficient data availability, and the fact that not all countries have well developed capital markets, not all countries contain the same set of domestic variables.

Country-specific foreign variables

The country-specific foreign variables (starred variables) are obtained by applying a weighting scheme to the foreign variables: for each country i , the foreign variable, for example the long-term rate, is obtained by averaging the long rate for all the other economies considered in the model. The weighted average is obtained by taking as weights the flows of imports and exports of the i -th country with respect to all other countries.²¹ The foreign variables are:

Real GDP	$y_{it}^* = \sum_{j=0}^N w_{ij} y_{jt}$	Inflation rate	$\pi_{it}^* = \sum_{j=0}^N w_{ij} \pi_{jt}$
Real equity index	$q_{it}^* = \sum_{j=0}^N w_{ij} q_{jt}$	ST interest rate	$\rho_{it}^{S*} = \sum_{j=0}^N w_{ij} \rho_{jt}^S$
LT interest rate	$\rho_{it}^{L*} = \sum_{j=0}^N w_{ij} \rho_{jt}^L$	Real exchange rate	$e_{it}^* = \sum_{j=0}^N w_{ij} e_{jt}$
Corporate spread	$cs_{it}^* = \sum_{j=0}^N w_{ij} cs_{jt}$		

where w_{ij} is the share of country j in the trade (exports plus imports) of i -th country. For a specified time period (eg one year or longer), these data may be collected in a 36x36 matrix, where for each country in a row, the trade shares with respect to all the other countries are displayed in columns. This matrix shows the degree to which a country depends on the remaining ones.

For example, in the case of Italy, about 70% of its international trade is with the first 10 commercial partners. With the exception of China and the United States, all major partners are European countries (euro area countries, the United Kingdom and Switzerland), with the two core economies, France and Germany, representing one third of Italy's total trade. Applying these trade weights to the domestic variables of Italy's trading partners, we obtain the Italian foreign variables.

In addition to domestic and foreign variables, oil and gold prices are considered as exogenous variables (global variables) in each country specific model, except the United States. Following Pesaran et al (2004) and Dees (2007), given the role of the United States in the global economy, the model for this country has been specified as follows: (i) in addition to the usual core variables, oil and gold prices are included as endogenous variables; and (ii) foreign output, inflation rates and exchange rates are the only exogenous variables (interest rates and equity indices are not included).

The GVAR is estimated on quarterly data from Q2 1979 to Q3 2012. A quarterly frequency is considered appropriate for the prediction of long-term returns. Data on monthly frequency contain significant noise; on the other hand, the time series of annual data are not particularly deep for the euro area.²²

²¹ Research on determinants of business cycle co-movements highlights that bilateral trade is one of the most important sources of inter-country business cycle linkages (Glick and Rose (1999)). A possible alternative could be the use of weights based on capital flows; however this information may be of a lower quality and more volatile than weights based on trade data (Dees et al (2007)).

²² Scherer B (2007, 2008).

The G-VAR model has been estimated using the G-VAR toolbox developed by the Centre for Financial Analysis & Policy of Cambridge Judge Business School (University of Cambridge). The toolbox was originally launched in December 2010 with the release of version 1.0, sponsored by the European Central Bank. Version 1.1, the one we used, was released in July 2011 and is available to download, free of charge, from the Judge Business School website.²³

Assumptions

For each series, we imposed a mean consistent with the consensus expectations or, in cases in which these are not available, with average historical values. For real GDP, inflation and long-term interest rates, we used the long-run forecasts by Consensus Forecasts, covering a period of 10 years.²⁴

For equities, mean returns are obtained assuming a convergence of the price-to-earnings ratios (PE) in different markets to their respective average historical values (taken as equilibrium values).

For short-term rates, we used the estimate for the current year, available from the most recent survey and we assume an unchanged level over the next two years. From the third to the fifth year, we assume a linear convergence to a long-run level of the short rate, obtained as the difference between the consensus forecasts of the 10-year interest rate and the term spread (long rate minus short rate) calculated over the period 1999–2012.

Average values for US and euro area investment grade corporate spreads are taken by forecasts produced by two prominent investment banks.

For the gold price, we imposed a mean real rate of return equals to zero (in US dollars). Historical evidence shows that over the last 200 years (from 1802 to 2003) the gold rate of return has closely followed the US inflation rate.²⁵ Over the period 2004–12, the price of gold has more than tripled, partly due to the flight to quality observed during the financial crisis.

The specified mean values are consistent with a macroeconomic scenario characterised by a sub-par growth rate, inflation rates and short-term interest rates anchored to policy targets and sovereign spreads (and gold price) consistent with a general normalisation of global economic and financial conditions.

From the GVAR simulations to the asset returns

Based on the scenarios generated by the GVAR model, we calculate financial asset rates of return. For money market instruments, we used the returns on a three-month deposit, for which the interest rate is observed at the beginning of each period. Bond

²³ The toolbox is available at <http://www-cfap.jbs.cam.ac.uk/research/gvartoolbox/index.html> and consists of a set of Matlab procedures executed via an Excel-based interface (Smith and Galesi (2011)).

²⁴ With regard to the yield forecasts for the 10-year government bonds issued by Greece, Ireland, Portugal, Austria, Belgium and Finland, some ad hoc hypothesis have been specified taking into account the spreads of these countries with respect to the yield of the German Bund over the period 1999–2009 (pre-sovereign debt crisis).

²⁵ Over the period 1802–2003, the gold price, deflated by the US inflation rate, provided an overall return of 40%, corresponding to a 0.15% return on an annual basis (Siegel (2004)).

returns are calculated from the simulated long-term yield; the duration of the bond, assuming a constant maturity of 10 years, is calculated via the approximation formula described in Campbell, Lo and McKinlay (1997).

For corporate bonds, yields to maturity are obtained adding the simulated credit spreads to the corresponding government yields. As the typical maturity for corporate bonds is five years, the corresponding five-year government bond yield is obtained from the short-rate time series and the five-year government term spread. This term spread is not directly available in the data set and is obtained by linear interpolation between the short and the long government rate time series. Corporate bond returns are then calculated from the simulated five-year corporate bond yield.

Stocks returns are directly calculated from simulated values of the equity index levels. Bilateral nominal exchange rates for each country in the GVAR model with respect to the US dollar are calculated from real exchange rates and consumer price indices. For euro area countries (EA), nominal exchange rates against the US dollar are weighted by PPP-GDP shares (to total EA GDP). Finally, as we use the euro as *numeraire*, we calculate its value in terms of the other currencies that are considered relevant for investment purposes (UK pound, Swiss franc, Japanese yen, Canadian dollars and Australian dollars).

Are central banks too risk-averse?¹

Massimiliano Castelli² and Stefan Gerlach³

Abstract

This paper asks whether central banks are too conservative investors. Since reserves are held for foreign exchange intervention, central banks have prioritised holding safe assets that are liquid in episodes of market turmoil. Moreover, reserves were historically small and have only recently become so large that they exceed what could plausibly be needed for intervention. Several governance factors that bias central banks toward being too conservative are identified. These include a principal-agent problem between the central bank and the Ministry of Finance; the need to ensure sufficient asset management experience among board members and the senior management; and a bias towards a steady stream of profits arising from the profit distribution rules. To offset these problems, governance changes may be necessary.

¹ Paper prepared for the Seventh Public Investors Conference, organised by the BIS, the World Bank, the Bank of Canada and the Bank of Italy, Rome, 22–23 October 2018. The views expressed in this article are those of the authors and do not represent those of the institutions with which they are affiliated. This version: 20 August 2019.

² UBS Asset Management, Managing Director, Global Sovereign Markets.
Email: massimiliano.castelli@ubs.com.

³ EFG Bank, Chief Economist. Email: stefan.gerlach@efgbank.com.

1. Introduction

As a consequence of the Great Financial Crisis (GFC), central banks are becoming increasingly important and sophisticated as investors. Before the crisis, their balance sheets were typically small and their significance stemmed largely from their role as monetary policymakers. However, the large increases in their foreign exchange (FX) reserves over the last two decades have made them important market participants.

Central banks' prominence as investors has also been amplified by the fact that they have increased the diversification of their assets and are now active in many more market segments than before the GFC. This development has been driven by the size of their reserves, which in many cases far exceed what is necessary for monetary policy purposes and therefore has induced central banks to think of broadening the range of assets in which they invest.

Another factor underpinning the diversification is the global search for yield, as asset managers seek to escape the very low levels of interest rates introduced by central banks in an effort to boost economic activity and stave off deflation. Nevertheless, despite these developments, central banks remain conservative investors, compared with pension funds and insurance companies in the private sector and other public sector investors.

The increased size of central banks FX reserves and their growing diversification suggests that some central banks allocate their portfolios in ways increasingly similar to those of sovereign wealth funds (SWFs). However, they do so without formally changing their investment processes, focusing instead on holding short-term liquid instruments and not taking advantage of long-term return opportunities of alternative asset classes.

The objective of this paper is twofold: (i) to discuss how central bank reserve management procedures and practices have evolved over the last two decades; and (ii) to question whether central banks are highly risk-averse as investors and, if so, which factors may have caused them to be so.

The remainder of this paper is organised as follows. Section 2 analyses the evolution of central bank asset allocation over the last 20 years. A central finding is that central banks have diversified their portfolios since the crisis struck. Section 3 compares central banks' asset allocations with those of other institutional investors. It concludes that while central banks have, like many other investors, "reached for yield" in recent years, they remain conservative. Section 4 focuses on whether central banks should diversify their portfolios further in the post-quantitative easing (QE) era. Here, the key finding is that the case for central banks to diversify more away from fixed income is strong and valid across a range of scenarios for the global economy. Section 5 turns to whether the current level of reserves is adequate given the current macroeconomic environment in which floating exchange rates are common. The main conclusion of this section is that many countries have a level of reserves that exceeds what traditionally has been considered as adequate. Section 6 explores some possible reasons as to why central banks are so risk-averse, focusing on principal/agent problems, the relative lack of asset management expertise at the board and senior management levels that are responsible for determining risk tolerance, and profit distribution rules. Section 7 looks at ways to overcome this bias through various institutional changes. Finally, Section 8 concludes.

Before proceeding, it is useful to note that central banks manage three different pools of assets: (i) FX reserves (which have risen sharply in emerging markets following the Asian financial crisis at the end of the 1990s), (ii) pension fund assets and (iii) assets (largely fixed income assets) acquired as a consequence of QE. These pools of assets serve different purposes, have varied investment styles and are associated with different liabilities. This paper considers FX reserves management where central banks have considerable room in terms of diversification and investment across asset classes, regions and currencies.

2. Changes in central bank reserve allocations in recent decades

Central bank FX reserve management practices have evolved considerably in recent decades. Twenty years ago, central banks had zero (or negligible) allocations to asset classes such as corporate bonds, asset-backed securities, emerging market debt and equity. The bulk of their assets was held in cash, bank deposits and short-duration government bonds from advanced economies. On the investment side, they were largely concerned about liquidity as FX reserves were considered as an instrument to intervene in FX markets when episodes of market pressures occurred and to provide liquidity to the financial sector during periods of domestic financial stress.

Two decades later the picture has changed. Most notably, FX reserves managed by central banks have grown to unprecedented levels: with more than USD 11 trillion of reserves, they constitute one of the largest institutional investor segments together with insurance and pensions funds. Given that the bulk of these assets are held in the fixed income markets of a few advanced economies, central banks' investment decisions can move markets. Furthermore, via QE central banks from advanced economies have been intervening heavily in global markets with the explicit goal of keeping interest rates low and to dampen volatility, which makes risky assets such as equities more attractive, with the intention of boosting economic activity. A few central banks (including the Bank of Japan) also hold domestic equities in addition to foreign government and corporate bonds.

Today, central banks invest across a wider range of asset classes, including spread products in fixed income markets such as corporate bonds and asset-backed securities, and increasingly outside the fixed income space such as listed equity. According to the *UBS Annual Reserve Manager Survey*, which has been tracking central banks asset allocation since 1998, more than half of central banks can invest in corporate bonds and asset-backed securities (ie MBS and ABS) and more than a third can invest in emerging market debt and equity (Graph 1).

Trends in approved asset classes
Percentage of respondents

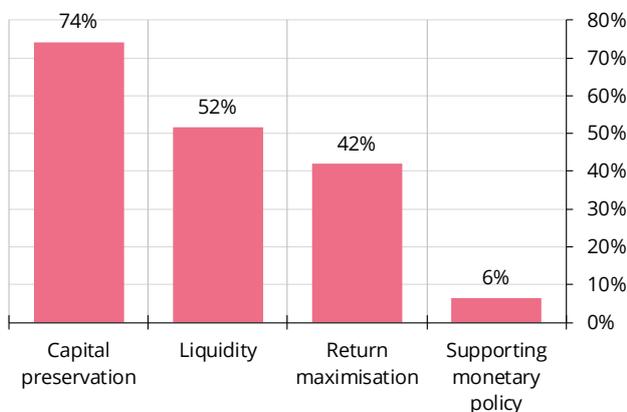
Graph 1

Asset class	2019	2018	2017	2016	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000	1999	1998
Supranationals	94	95	93	90	89	94	82	87	76	70	68	70	72	70	74	63	62	60	62	54	63	60
Sovereign eurobonds	85	86	89	93	92	90	78	73	64	68	65	66	66	64	60	60	58	58	61	60	66	66
US agencies	85	76	74	83	74	49	54	69	55	65	71	84	86	84	82	76	78	75	71	62	68	54
Inflation-protected bonds	73	62	76	70	61	49	48	40	44	47	45	38	33	28	16	9	na	na	na	na	na	na
Covered bonds	45	43	48	53	34	43	40	45	30	38	40	53	58	50	48	44	38	35	37	34	28	12
Bank debt	45	43	48	40	42	41	26	33	20	31	29	48	44	48	41	21	24	21	26	20	16	4
Corporates	61	57	56	48	42	43	34	33	31	29	26	38	41	40	38	38	32	28	22	20	15	10
MBS / ABS	58	57	52	40	39	25	20	22	37	27	38	46	52	44	39	39	27	22	17	19	12	2
Emerging market debt	36	24	26	30	21	18	12	11	22	24	16	16	10	na								
Equities	39	24	30	33	26	18	16	24	18	19	14	18	22	18	5	3	2	na	na	na	na	na
Private equity	18	5	19	na																		
Hedge funds	15	0	11	na																		

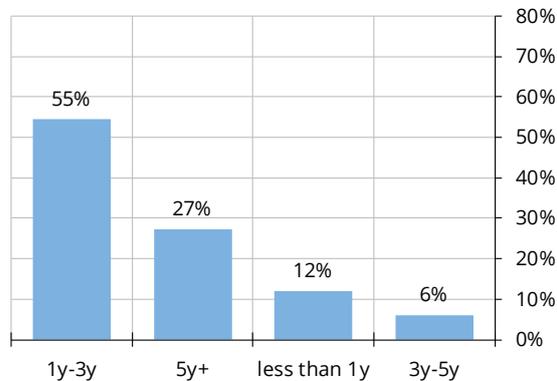
Source: UBS Annual Reserve Manager Seminar Survey, various years.

The primary objectives of central banks when investing their reserves are capital preservation, liquidity and return. Capital preservation and liquidity dominate, with the return objective considered important but only as long as the two primary targets are fulfilled. This prudent asset allocation is considered appropriate by central banks given their mandate: an ample reserve of liquid funds might be needed should crises occur, and large losses could have serious reputational costs. In addition to that, central banks have a relatively short investment horizon, which generally implies a pro-cyclical investment behaviour with potentially negative effects on long-term returns.

What are your primary investment objectives?



What is your investment horizon?



Source: UBS Annual Reserve Manager Seminar Survey, 2019.

The increasing diversification of the last two decades reflects several factors, including the dramatic rise in FX reserves, particularly in emerging market economies. It was, however, the sharp drop in interest rates following the start of the GFC in 2008 that prompted central banks to join other institutional investors in the “search for yield”. The bulk of FX reserves has traditionally been invested in government bonds denominated in USD, EUR, JPY and a few other currencies. Most of these markets were affected by the extraordinary monetary policy measures adopted by central banks to stave off the risk of deflation and support recovery. As a result, since 2009 nominal returns on cash and short duration government bonds across these markets have been very low and, in some cases, negative. In 2009–18, cash generated an average yearly return of 0.7%; the return on a short-duration government bond portfolio was around 1%. A central bank’s portfolio invested across the major currencies in a 50/50 cash/short duration bond portfolio has generated a return of around 1% since 2009, well below the level of the previous decade (of around 4%).

Capital preservation can be defined in various ways but, at a minimum, investors want to protect capital in real terms to maintain its purchasing power. Since 2009, investing only in cash and short government bonds – assuming an inflation rate of around 2% or slightly lower – has not allowed central banks to protect the real value of their reserves.

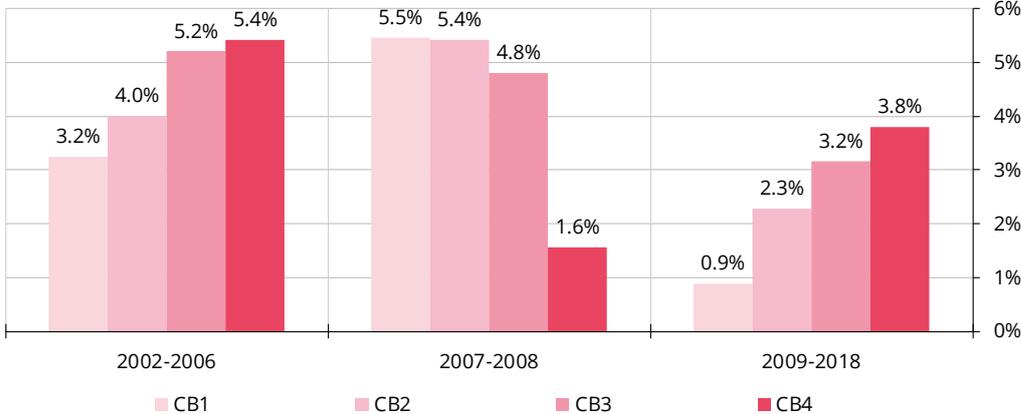
Reducing cash and increasing duration in government bonds boosted returns. Long government bonds have generated good returns over the last few decades as a result of the secular fall in long-term interest rates. However, since 2009 their returns have also been falling: about 3.2% compared with nearly 6% before 2008 and with the additional risk of investing in an expensive asset class. According to data provided by the US Treasury Department, duration into spread products and US Treasuries held by official institutions has been rising since 2009, providing empirical evidence for such a trend (Jones (2018)).

By broadening the investable universe, central banks have been able to capture higher returns and gain diversification benefits. Since 2009, investment-grade corporate bonds generated an annual return of nearly 5% and emerging market

bonds in hard currency more than 5%. Global equities rewarded investors with an annual return of more than 9%. Over this period, a central bank portfolio with 60% invested in cash and short duration bonds and 40% diversified in (investment grade) corporates, asset-backed securities, supranationals and long-duration bonds generated a return of above 2%, more than doubling the return of the cash/short duration bond portfolio and protecting the real value of the reserves. By diversifying even further in more risky asset classes, such as emerging market debt and equities, central banks were able to generate returns above 4%, thus fulfilling all their investment objectives of capital protection, liquidity and return enhancement.

Returns of CB1 to CB4 portfolios over 2009–18
Percentage

Graph 3



Source: UBS Asset Management. Please refer to Table 1 “Central banks’ portfolio with allocation to alternatives” for more information on the composition of sample central bank portfolios CB1-4. 2019 YTD includes data up to May 2019.

Central banks have often increased duration when diversifying into equities. Traditionally the returns on (long-term) government bonds and equities have been negatively correlated, thus extending duration on fixed income has been considered as a way to protect the portfolio in case of falling equity prices. The negative correlation between equity and fixed income returns provides important diversification benefits to central banks’ portfolio, which have large allocations to fixed income assets and are heavily exposed to interest rate risk. Central banks have simply adopted the same investment strategy as many other institutional investors such as pension funds. In this sense, over the last decade FX reserve management has evolved towards common practices among institutional investors.

3. How does central bank asset allocation compare with that of other institutional investors?

Central banks’ asset allocation varies considerably across regions and institutions, reflecting different levels of reserves and different economic, financial and institutional conditions. Broadly speaking (and based on anecdotal evidence), central banks in the Americas and Africa are the most conservative, with some limited

diversification within fixed income. They only rarely venture into riskier asset classes such as emerging market debt or equity.

At the other end of the risk spectrum, Asian central banks have been diversifying aggressively over the last several years. This is largely a reflection of a high level of reserves, which are often well above the level considered as adequate for precautionary reasons, and the high level of national saving prevailing in these economies. They often diversify into equities and a few also invest into illiquid alternative asset classes such as real estate.

European and Middle Eastern central banks stand in the middle between America and Asia, but with a recent increasing trend towards more diversification, particularly into equities. There are some notable exceptions – for instance, the Swiss National Bank, which has a very high level of reserves and a relatively high allocation to equities.

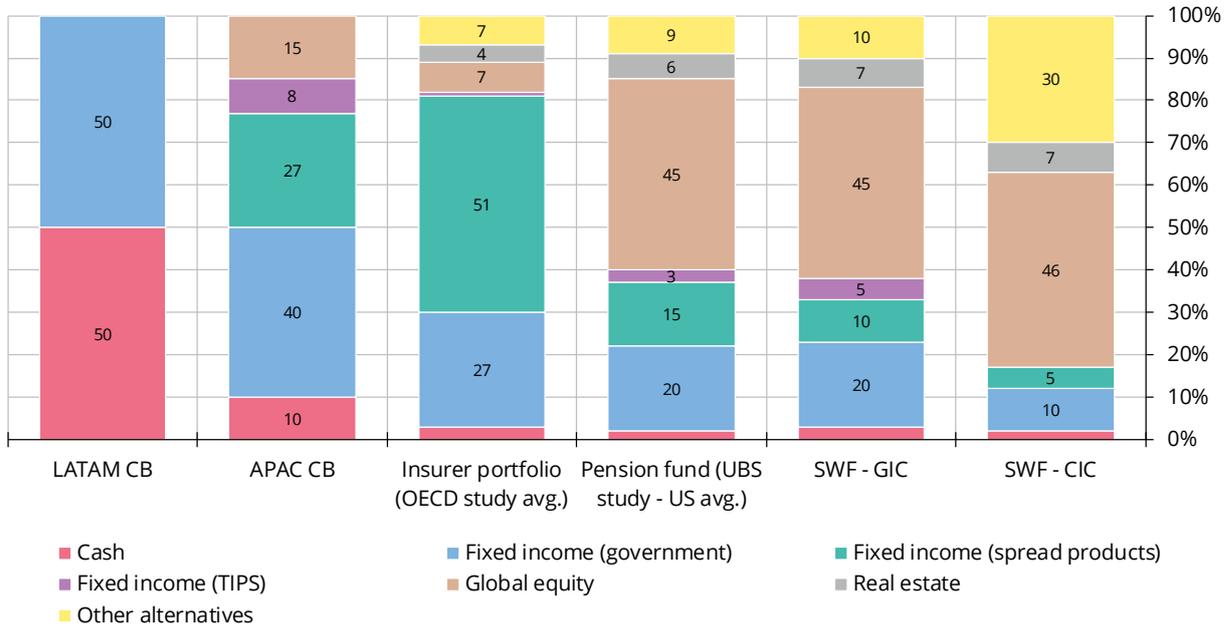
Despite the trend toward increased diversification, in general the investment profile of central banks remains conservative when compared with other institutional investors such as pension funds or insurance funds. Two main differences stand out: (i) central banks still have a relatively small allocation to equities; and (ii) have an almost zero allocation to alternative asset classes such as hedge funds, real estate and infrastructure.

Pension funds generally invest 40% into equities and increasingly invest into alternative asset classes. Insurance funds invest the bulk of the portfolio into fixed income largely because of regulations. But they also have equities and, more importantly, alternative asset classes.

When compared with SWFs, the asset allocation of central banks appears to be even more risk-averse, which is not surprising given the long-term investment horizon of these institutions and the fact that they have been created precisely to diversify reserves aggressively in global capital markets. However, it is also worth noting that some SWF's assets are sometimes run by central banks (eg Norway), showing that central banks have the ability and skills to manage more diversified portfolios when they are given such a mandate from their sponsoring governments.

Asset allocation of selected institutional investors
Percentage of total assets

Graph 4



Source: UBS Asset Management on OECD data. GIC is the SWF in Singapore in charge of diversifying excess reserves in global markets. China Investment Corporation (CIC) is the largest Chinese SWF.

4. Should central banks increase diversification further in the post-QE era?

Diversification paid off in the QE era as credits and equities produced good returns with low volatility due to the secular fall in long-term interest rates and extraordinarily expansionary monetary policy. Central banks that diversified 10–20% of their reserves into equities enjoyed very good risk-adjusted returns.

An intriguing question is whether this trend will continue as the world moves into the post-QE era. As global growth remains steady and interest rates are gradually normalised, cash and short-duration government bonds are expected to generate returns in excess of 2% over the next five years. Global government bonds are expected to be the worst-performing asset class in this scenario as long-term interest rates rise gently from historically lows. Credits and other spread products also generate lower returns than in the past decade as interest rate rises and spreads start from historical lows. Despite current high valuations in certain markets, global equity is expected to generate good returns, well above those in fixed income assets but slightly lower than during the last decade.

In these market conditions, a portfolio including only fixed income assets is likely to generate returns below 2%. While this is better than the returns experienced in the QE years due to higher policy rates, it is below the expected rate of inflation. In such a scenario, diversifying into equities will boost returns above 3%, thus beating inflation and providing reserve managers with real returns.

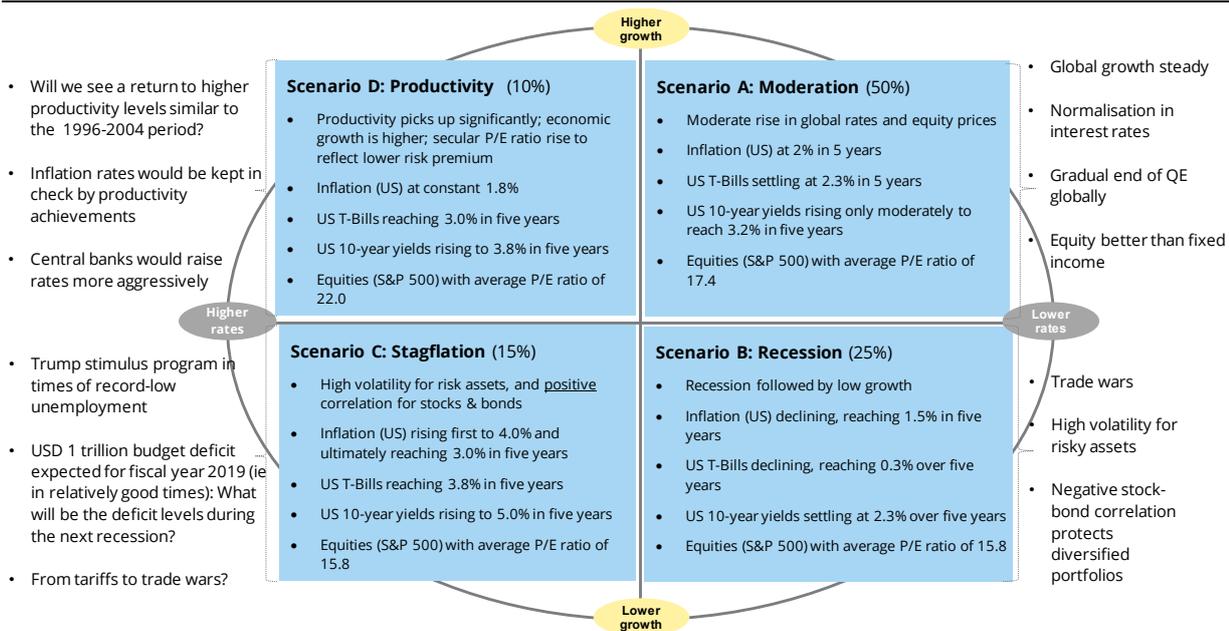
Should the global economy suffer a slowdown or a recession, interest rates would likely fall again to (close to) zero and QE would be resumed. Central banks would be once again challenged to protect the real value of their reserves given their large allocations to government bonds issued by advanced economies. The search for yield would resume.

Should central banks therefore further expand their investment universe as has been done by other institutional investors? From a purely asset allocation perspective, the case for adding alternatives to further diversify reserves is strong. Listed equity is currently considered attractive not in absolute terms but relative to fixed income. There is much uncertainty over the future pace of the global economy following 10 years of (weak) recovery, the potential negative impact of the gradual normalisation in policy rates and the gradual withdrawal of QE. The likelihood of a slowdown or even recession in the United States is rising.

Graph 5 shows possible scenarios for the global economy over the medium to long term with different combinations of growth levels and yields. Furthermore, Table 1 shows the hypothetical returns of different asset classes and portfolios across different scenarios based on UBS valuation models.

Inflation, growth, yields & valuation: four scenarios

Graph 5



Source: UBS Asset Management.

The table indicates how central banks' portfolios, including selected alternative asset classes (ie real estate, hedge funds and infrastructure), would perform across these scenarios. In the base case, maintaining more than 50% of the portfolio to fixed income assets and adding 15% of alternative asset classes would boost returns in both absolute and risk-adjusted terms. These selected alternative asset classes (private equity is excluded to reduce the reputational risk arising from investing in specific private companies) not only provide a source of additional returns but also improve the risk-return profile of the portfolio. In a recession scenario, returns are lower but still positive as the fixed income component of the portfolio compensates for the losses experienced in the riskier equity and alternatives.

Portfolios diversified into alternatives would also perform better in a stagflation scenario. This is the worst scenario for those central banks that have diversified into equities, as the traditional negative correlation between the returns on fixed income and listed equity would turn positive.

Overall, the case for further diversification, including alternatives for central banks, is strong, particularly as the global economy is likely to shift to a new regime characterised by a lower return on fixed income and higher volatility.

Central banks' portfolio with allocation to alternatives*

Table 1

		CB1	CB2	CB3	CB4	CB5a	CB5b	CB5c	CB5d
Cash		50%	10%	10%	10%	10%	10%	10%	10%
GGB 1-3		50%	50%	30%	30%	20%	25%	25%	25%
GGB				10%	10%	5%	5%	10%	5%
Corporate Bond			10%	10%	8%	8%	8%	5%	6%
TIPS			10%	10%	8%	8%	8%	5%	6%
Securitisied			10%	10%	8%	8%	8%	5%	6%
Supranationals			10%	10%	8%	8%	8%	5%	6%
EMD Hard Curr				10%	3%	3%	3%	3%	3%
EMD Local Curr									
Global Equity					15%	15%	10%	15%	15%
Real Estate					0%	5%	5%	7%	8%
Private Equity									
Hedge Funds					0%	5%	5%	5%	5%
Infrastructure					0%	5%	5%	5%	5%
Gold									
Commodity									
Total		100%							
		CB1	CB2	CB3	CB4	CB5a	CB5b	CB5c	CB5d
Historic	Yearly Return	2.20%	3.02%	3.81%	4.21%	5.11%	4.78%	5.09%	5.18%
	Std Dev	0.74%	1.86%	2.57%	2.92%	4.58%	3.96%	4.68%	4.92%
	Return/Std. dev	2.95	1.62	1.48	1.44	1.12	1.21	1.09	1.05
	Minimum rolling yearly Return	0.22%	-1.06%	-1.86%	-7.53%	-15.81%	-12.99%	-16.60%	-17.62%
BaseCase	Yearly Return	2.36%	2.73%	3.18%	3.84%	4.39%	4.04%	4.36%	4.45%
	Std Dev	1.2%	2.6%	3.0%	4.0%	4.7%	4.0%	4.6%	4.7%
	Return/Std. dev	1.97	1.04	1.05	0.96	0.94	1.00	0.95	0.95
Recession	Yearly Return	1.6%	2.4%	2.9%	2.5%	2.5%	2.6%	2.4%	2.5%
	Std Dev	1.2%	2.0%	2.2%	3.5%	4.8%	3.9%	4.9%	5.1%
	Return/Std. dev	1.34	1.19	1.32	0.73	0.53	0.66	0.50	0.48
Stagflation	Yearly Return	2.3%	1.8%	2.0%	1.7%	2.2%	2.2%	2.0%	2.2%
	Std Dev	1.2%	3.1%	3.7%	5.8%	6.8%	5.8%	6.7%	6.8%
	Return/Std. dev	1.81	0.58	0.55	0.30	0.32	0.37	0.30	0.32
Productivity	Yearly Return	0.8%	1.3%	2.2%	2.8%	3.7%	3.3%	3.8%	3.9%
	Std Dev	1.2%	2.2%	2.4%	3.0%	3.5%	3.0%	3.4%	3.5%
	Return/Std. dev	0.73	0.60	0.93	0.96	1.05	1.08	1.11	1.11

*GGB 1-3 is a short-duration government bond portfolio invested into USD (65%), EUR (25%), JPY (5%) and GBP (5%); GGB is a long-duration government bond portfolio with the same currency breakdown. Both portfolios hedged in USD. Historical returns are for 2002–19 May. Expected returns are for 2019–24.

Source: UBS Asset Management.

5. Central banks' capacity to bear risk

As noted above, central banks are more conservative investors than pension funds and insurance companies but also more than other public sector entities. Nevertheless, since the onset of the financial crisis, they have diversified their portfolios to include higher-yielding but riskier assets.

Why might central banks be so conservative? Traditionally, central banks operated with fixed exchange rates and needed FX reserves to sell against domestic currency in periods of speculative outflows. Since pressures could develop very quickly and unexpectedly, it was essential to hold the reserves in a highly liquid form so that they quickly could be used for intervention in currency markets. In practice, this meant that central banks held safe short-term USD treasury debt.

Over time, however, two factors have reduced the relevance of this consideration. First, few central banks now operate monetary policy with a fixed exchange rate. With inflation targeting and similar strategies, in which the exchange rate only matters to the extent it risks pushing inflation away from the target, central banks' need to hold FX reserves for intervention purposes has been sharply reduced. Nevertheless, some commodity exporting economies, for instance those in the Gulf, still maintain fixed exchange rates with the US dollar; this has not prevented them from aggressively diversifying accumulated reserves via SWFs. In the Gulf, central banks hold a relatively low level of reserves.

Second, FX reserves are now in many cases far larger than what could plausibly be needed for intervention. Following the Asian financial crisis, central banks in emerging economies decided to increase their FX reserves to be able to better withstand occasional episodes of market pressures. This was made all the easier to achieve following the onset of the GFC after the collapse of Lehman Brothers in September 2008, when the Federal Reserve and central banks in other advanced economies cut interest rates to unprecedentedly low levels and adopted unconventional monetary policy to provide further stimulus. This led to large inflows in emerging economies, which central banks often absorbed to mitigate the upward pressure on their exchange rates.

The consequence has been large increases of their FX reserves that are now exceptional by any standard. The literature on what is the adequate level of reserves is vast and different approaches can be adopted (see Castelli and Scacciavillani (2012)). As a rule of thumb, a level of reserves at around 10% of GDP is considered as adequate to cover the FX transaction needs of the government and private sectors and to have a buffer against unforeseen events such as a sudden drop in cross-border capital flows or stress in the financial sector. As one can see from Graph 6, many countries have reserves well above this threshold.

The reduced need to hold foreign reserves for exchange market intervention, coupled with the huge increase in reserves, means that many central banks by now have become in all but name sovereign wealth funds. Nevertheless, they often retain their past investment strategies and focus on short-term liquidity rather than long-term capital returns.

FX reserves as a proportion of GDP in selected economies
Percentage

Graph 6

Country	GDP (current \$bn)	FX reserves (\$bn)	FX reserves as % of GDP
Switzerland	704	762	103%
Lebanon	56	43	81%
Singapore	361	277	79%
Saudi Arabia	782	487	65%
Algeria	180	114	58%
Czech Republic	242	145	58%
Thailand	487	193	40%
Bulgaria	65	26	40%
Botswana	19	7	39%
Mauritius	14	5	38%
Cambodia	25	8	34%
Jordan	42	14	33%
Israel	370	112	30%
Malaysia	354	96	26%
Peru	225	61	26%
Croatia	61	16	26%
Oman	82	20	25%
Japan	4,972	1,202	23%
Kuwait	141	31	23%
Korea	1,619	379	22%
China	13,407	3,140	22%
Iceland	26	6	21%
Philippines	331	70	21%
UAE	425	85	21%
Russia	1,631	346	20%
Angola	107	24	20%
Poland	586	108	18%
Brazil	1,868	359	17%
Qatar	192	31	17%
Vietnam	241	37	15%
Tunisia	40	6	15%
Colombia	333	45	14%
Norway	435	60	14%
Mexico	1,223	165	14%
Chile	298	38	14%
India	2,717	385	14%
Azerbaijan	45	6	13%
Latvia	35	4	12%
South Africa	7	43	12%
Indonesia	1,022	124	11%
Bangladesh	288	32	11%
Ghana	65	6	11%
Uganda	28	3	11%
Kazakhstan	171	17	10%
Turkey	766	83	9%
Sweden	551	53	9%
New Zealand	203	19	8%
Argentina	518	50	8%
Nigeria	397	29	7%
United Kingdom	2,829	127	4%
Canada	1,711	77	4%
Australia	1,418	58	4%
Spain	4	52	3%
Italy	2,072	38	2%
Belgium	533	10	2%
Austria	458	7	1%
Germany	4,000	60	1%
France	2,775	38	1%
Estonia	30	0	1%
Greece	219	2	1%
Netherlands	913	5	1%

Source: UBS Asset Management, OECD. Figures for GDP from the International Monetary Fund and figures for FX reserves from the World Bank.

To conclude this section, it is worth noting that central banks are better able to shoulder risk than private sector asset managers. This is because in an emergency they can operate temporarily with little or even negative capital. A financial firm that experienced large losses would soon find counterparties disengaging and customers withdrawing business. If the losses were sufficiently large, the firm would find itself frozen out of the markets and unable to survive. Central banks, by contrast, are monopoly suppliers of domestic currency and are always in position to execute payments. Moreover, there are no legal reasons why insolvent central banks cannot continue to operate (as several have done), in contrast to private institutions that would be resolved in such a situation.

6. Why are central banks so risk-averse?

Nevertheless, central banks seem to be excessively risk-averse. Three factors may play a role: (i) a principal/agent problem may exist, (ii) the board responsible for setting the investment strategy and risk tolerance may lack investment expertise and experience, and (iii) profit distribution rules may reduce the scope for central banks to withstand losses.

6.1 Principal/agent problems

Excessive caution may reflect a principal/agent problem. This arises when the agent (in this case the central bank) makes decisions on behalf of the principal (in this case the treasury), but is motivated to act in its own best interests, which may be contrary to those of its principal. Such problems can arise when the two parties have different interests and the principal cannot be sure that the agent, who holds more information, is acting in the principal's best interest.

Since central bank profits reduce the needs for treasuries to finance government expenditures by raising distortionary taxes, treasuries are naturally keen for the central bank to generate returns similar to those of private sector asset managers, which requires it to assume similar risks.

For the central bank, however, the risk-return trade off may appear different. It may release quarterly profit/loss statements to the public, which often are regarded as newsworthy by the media. Reports of losses, even if rare and also experienced by many private sector investors, may be used by the press to argue that the central bank is incompetent. Furthermore, many central banks believe that the effectiveness of their monetary policy depends on their announcements being credible and they being seen as competent. Excessive media attention to occasional short-term losses is therefore undesirable. As a consequence, central banks may structure their portfolios to avoid such losses, at the cost of lower long-term profits.

6.2 Lack of investment expertise in central bank boards and senior management

Reputational considerations may also be important to a central bank's board and senior management. The board typically must approve the central bank's investment strategy and determine its risk tolerance on the advice of the central bank's senior

management. Central bank boards often consist of prominent members of society, drawn from the legal and accounting professions, academia, labour unions, or from retired politicians or civil servants. However, they are rarely drawn from those with expertise in asset management. Similarly, members of senior central bank management also rarely have extensive professional experience in asset management. Since the main tasks of central banks are to conduct monetary policy and, in some cases, to supervise banks, those that rise to the top of these institutions typically do so through these routes rather than through managing the FX reserves.

Given the private reputational risks associated with presiding over a central bank experiencing losses on its investment portfolio, and the associated risk to the credibility of the institution, it is not difficult to see that board members and members of the senior management will err on the margin of safety in establishing investment guidelines. As a consequence, the central bank may not seek to earn a market return on its FX reserves even if the treasury wishes it to do so.⁴

6.3 Profit distribution rules

Another factor that can induce central banks to be excessively cautious is the profit distribution rules. These are typically set in central bank legislation and therefore not always easy to change. Uncertainty about the size of future profit transfers to the treasury is a problem for the fiscal authorities that must prepare long-term projections of government revenues. With distributions paid from current profits, central banks can come under pressure to ensure steady income and to avoid losses that may prevent the central bank from transferring profits every year. This can tilt the central bank's investment strategy in the direction of holding relatively safe assets that generate sure, but small, returns.

7. Dealing with excessively cautious central banks

To overcome the problems mentioned above, governance changes may be necessary.

7.1 Clarifying the central bank act

Central bank boards and senior management may attach greater weight to the effects on their reputation and credibility of negative publicity arising from occasional large losses on their portfolios, and may therefore adopt more conservative investment strategies than desired by their principals. This may be all the easier since central bank acts often provide no guidance about how funds should be invested.

Turning to monetary policy for a moment, there is much agreement that inflation has been lower and more stable in recent decades because central bank acts were rewritten to clarify that price stability was a primary objective of policy. With a clear legal remit and political backing, central banks have been free to focus on this objective, raising the likelihood of success.

⁴ This raises the issue of the appropriate degree of transparency in central bank asset management, an area in which practices vary across regions and countries. See IMF (2014) for a review of central banks' transparency standards.

This suggests that spelling out central banks' investment objectives in legislation might also be desirable. Of course, it may not be possible to make clear in detail what the objectives are. However, phrases such as "reserves not necessary for foreign exchange market purposes should be invested for long-term capital gain" or that "in managing its investments, the central bank should adopt principles broadly similar to those of other long-term asset managers in the public and private sectors" might be helpful. Clarifying society's preferences in legislation provides legitimacy to the central bank's asset management decisions and reduces the risk to its reputation and credibility of occasional capital losses.

From this point of view, the recent *Revised guidelines for foreign exchange management* (IMF (2014)) was too cautious to heighten the focus on long-term returns. The revised guidelines also do not adequately address pro-cyclical portfolio behaviour, which is one implication of excessive risk aversion by central banks. Given the large pool of assets managed by central banks, their investment behaviour is likely to amplify market movements, particularly during periods of falling asset prices (Jones (2018)).

There are several examples of such investment behaviour by central banks. For instance, following the GFC, central banks reduced by half their deposits with commercial banks, thus amplifying the liquidity crunch of the international banking system. During the euro area fiscal crisis, several central banks cut their exposure to peripheral government bonds as credit rating agencies reduced the sovereign debt ratings of these countries. Other official institutions with a much more diversified asset allocation and low risk aversion, for instance SWFs, bought equities when other investors were selling, thus contributing to provide liquidity to the market.

In these cases, there was a clear conflict between the goal of central banks to maintain stability in the financial system during periods of financial stress and the procyclical behaviour of central banks.

This conflict could resurface in the future. For instance, should inflation surprise on the upside and force central banks to tighten monetary policy by more than expected, the impact on the government bond market could be large, tempting central banks to cut losses and shift assets into cash to protect their portfolios.

7.2 Governance and investment committees

Another way to overcome the risk of an excessively cautious investment approach is to delegate decision-making to an investment committee. Again, a comparison with monetary policy is warranted. While monetary policy decisions historically were taken by the central bank governor (unless the central bank lacked independence, in which case they were set by the minister of finance), in the last two decades monetary policy committees (MPCs) have been increasingly adopted. The hallmark of these is that the members – who are appointed for three- to five-year terms – comprise a combination of senior central bank staff and "outsiders" who are selected based on their expertise. As a consequence, monetary policy decisions are not in the direct control of the central bank.

The benefit of outsiders is that they may be less prone to group think, which is always a risk in a central bank. While differences in remuneration levels between the public and private sectors may make it difficult to attract experts with asset management experience from the private sector to senior positions in central banks,

a temporary appointment may be more attractive. Indeed, the Bank of England's MPC often attracts private sector economists.

In this regard, the experience of SWFs is important. The best run SWFs are those where the responsibilities of the sponsoring government, the principal and the SWFs' management, the agent, are clearly defined and separate. The sponsoring government is in charge of defining the risk and return expectations. This does not require it to be prescriptive in terms of individual asset classes, regions etc, but rather by indicating what level of risk is considerate acceptable. According to this governance model, the sponsoring government defines a "reference portfolio", ie a simple combination of listed liquid equity and fixed income. The SWFs' management will be in charge of defining the investment policy within those parameters, maintaining a certain discretion in portfolio decisions including the selection of asset classes. The SWF's management could also include illiquid asset classes if this improves the risk-return trade-off when compared with the reference portfolio.

7.3 Profit rules

Changing profit rules may also be helpful. Currently, central banks typically repatriate some fraction of their annual profit to the ministry of finance. If the ministry values a highly predictable stream, the result may be that the central bank aims for secure, but low, profits.

To avoid this, some device that allows the central bank to smooth profits over time might be helpful. For instance, the central bank could pay the annual profits into a buffer, held by the central bank (or perhaps even by the treasury itself or a new institution), from which profits could be paid at a regular rate. An alternative might be for the central bank to pay out a fixed fraction – such as the expected return – of the portfolio. The pay-out rate could then be adjusted regularly in light of some objective criteria. This is effectively the logic followed by foundations, university endowments and SWFs.

Needless to say, any change of this type could have important legal and accounting implications that would need to be considered.

8. Concluding remarks

This paper has sought to shed light on whether central banks are too conservative as investors. Central banks' risk aversion, which exceeds that of most private sector investors (and other public sector investors), is seen largely as reflecting historical circumstances. In particular, since reserves have been held to permit FX intervention, central banks have put a premium on holding assets that are highly liquid even in circumstances of market turmoil. Central banks' reserves have also in many cases been relatively small and have only recently grown to such an extent that they exceed what could plausibly be needed for intervention purposes.

A number of institutional and governance factors that may bias central banks toward being conservative investors are identified. These include a principal/agent problem that can arise between the central bank and the ministry of finance; the need to ensure that board members and senior central bank management have sufficient asset management experience; and a potential bias towards a steady stream of profits

arising from the rules determining the distribution of profits. The main conclusions are that central banks are plainly in a position to raise returns by assuming more risk, that some changes in governance may make them more prone to do so, and that their ability to shoulder risks is greater than that of many private sector investors.

Overall, it seems appropriate for central banks to reconsider whether their current risk-return trade-offs are appropriate in light of the size of their reserves and to address any low risk bias identified, for instance by adding more asset management expertise at the board and senior management level, set better incentives to resolve any principal-agent problem and improve governance on the investment side. The experience of SWFs, institutions created to accelerate diversification of reserves, particularly with regards to governance, incentives and the ability to attract and retain financial expertise, should be looked at by central banks.

References

Bank for International Settlements (2011): "Portfolio and risk management for central banks and sovereign wealth funds", *BIS Papers*, no 58, October.

Berkelaar, A, J Coche and K Nyholm (2010): *Central bank reserves and sovereign wealth management*, Palgrave Macmillan.

Castelli, M and F Scacciavillani (2012): *The new economics of sovereign wealth funds*, Wiley Finance, United Kingdom.

IMF (2014): *Revised guidelines for foreign exchange management*.

Jeanne, O and R Rancière (2006): "The optimal level of international reserves for emerging market countries: formulas and application", *IMF Working Paper*, vol 6, no 229.

Jones, B (2018): "Central bank reserve management and international financial stability – some post-crisis reflections", *IMF Working Paper*, vol 18, no 31.

Part 2

Sustainability Considerations for Official Institutions

ESG investing: the role of public investors in sustainable investing

Ulrike Elsenhuber¹ and Adela Skenderasi²

Abstract

This paper provides an overview of environmental, social and governance (ESG) investing from the perspective of public investors. In an effort to evaluate the role of public investors in sustainable investing, the Bank for International Settlements in April–May 2018 conducted an informal survey among a number of central banks, international organisations and asset managers. The results indicate that public investors are increasingly being pressed to play a role in sustainable investing, but they face a number of challenges in this process. Among these, the most critical are the lack of a commonly adopted ESG taxonomy, and also limitations on the application of the various ESG approaches in some of the portfolios they manage. At present, central banks have introduced ESG factors mainly for their pension fund investments, with the aim of further integrating sustainable investing into their own funds and in foreign exchange reserves portfolios, given that the strategic asset allocation of the latter tends to be less diverse and focused on asset classes that currently do not have a conventional ESG approach. Given the growing relevance of ESG investing for public investors, this paper identifies the motivations for more sustainable investing among public investors and draws attention to the underlying constraints that influence the decision-making of public investors with respect to sustainable investing.

¹ Bank for International Settlements, Asset Management Unit. Email: Ulrike.Elsenhuber@bis.org.

² International Monetary Fund, Investment Office. Email: Adela.Skenderasi@imf.org. The author was employed by the BIS Asset Management Unit during the drafting of this research paper.

1. Introduction

Over the last decade, sustainable investing has gained considerable momentum as investors acknowledge the benefits of environmental, social and governance (ESG) investing from a risk management perspective (both for the portfolio and the reputation of the institution), as well as from the viewpoint of the long-term impact on society and the environment.

At the same time, given the nascent stage of sustainable investing, there is no commonly accepted definition or standard for what constitutes a “sustainable” investment, causing a great deal of uncertainty. Instead, there are many different names and taxonomies, each emphasising a particular aspect and different sets of motives. Nevertheless, all of these are related and overlap to a significant extent. While the focus of this paper is ESG investing, we introduce some of the other main concepts used to incorporate sustainability considerations in the investment process to highlight how they differ, namely:

1. **Ethical or values-based investing:** an approach where specific ethical considerations or religious values are taken into account in the investment process. Ethical investing is also commonly referred to as negative screening, as it excludes investments in specific companies or industries that are considered objectionable due to their involvement in activities that are in conflict with particular values or social norms. For example, the exclusion of controversial industries, such as tobacco, alcohol or weapons, from portfolio construction would be referred to as ethical investing.
2. **Impact investing or corporate engagement:** a practice whereby impact investors actively seek to influence corporations by encouraging better practices and might engage in shareholder meetings to address ESG shortcomings. Impact investors allow investments in products with specific features, such as green bonds or bonds from issuers with a low carbon footprint, and they consider social or environmental benefits alongside financial return.
3. **Positive screening:** a process that targets the selection of best-in-class companies to form the desirable investment universe. An example of a positive screening approach would be to favour investments in global companies with the highest ESG scores relative to sector peers.
4. **ESG investing:** an approach that seeks to incorporate ESG factors, alongside financial considerations, in the investment process to evaluate respective risks and opportunities and to measure the sustainability of the investments. ESG metrics may be applied in the form of an overlay strategy³ or they may be included in all aspects of the investment process.⁴

ESG considerations span:

- **Environmental issues** such as climate change, energy and water usage, CO₂ emissions, pollution, waste etc.

³ Eg exclusion of issuers with the lowest ESG ratings or applying an ESG momentum tilt by over/underweighting issuers that experienced ESG up- or downgrades. The integration of ESG into existing investment portfolios in the form of an overlay strategy allows a present alpha strategy to be maintained, while seeking to use ESG ratings and ESG momentum for risk-return enhancement.

⁴ That is, in security valuation, formation of expected returns, risk analysis and portfolio construction.

- **Social issues** such as workplace diversity, labour laws, health and safety considerations etc.
- **Governance issues** such as business ethics, board structure and independence, executive compensation, accounting, anti-fraud and anti-corruption policies etc.

ESG investing is based on the notion that ESG factors are drivers of a company's long-term value, risk and return, and that they signal how sustainable the company is over the long term. Including ESG factors in the analysis of a company extends the assessment beyond traditional financial risk metrics and offers a more holistic approach towards risk management, as it includes an analysis of intangible and "soft" factors to capture the non-financial risks to which a company is exposed.

In the past two decades, several ESG ratings providers have emerged. They aim to assess the key sustainability risks of different entities (sovereigns, corporations or others). ESG ratings are relative in nature: they rank individual issuers within relevant peer groups, such as a particular industry or sector and may thus help to identify best-in-class peers. Each ESG ratings provider uses its own proprietary methodologies to rate entities on ESG measures, leading at times to differences in their assessments of an issuer. Despite this, a number of studies⁵ using ESG ratings from both MSCI and Sustainalytics (two of the primary ESG ratings providers), revealed broadly similar results about the impact of ESG factors on portfolio performance. As such, portfolios with high ESG scores outperformed portfolios with low ESG scores, with the governance factor showing the strongest link to outperformance.

In this paper, we document the rationale for ESG investing among public investors, the fundamental challenges they face, the role of public initiatives in ESG investing, and finally we present through the survey results why and how public investors implement ESG investing.

2. Rationale for ESG investing

Over the last decade, as ESG investing has shifted from a regional, niche initiative to a global trend, public investors' conversation around ESG considerations has also evolved. A growing number of them is considering implementing one of the ESG investing approaches in their portfolios (pension funds, own funds or foreign exchange reserves). There have been several motivations leading to the consideration of ESG investing, chief among them are the following:

a. Addressing ESG-related concerns

The public and private sectors recognise that failure to mitigate climate risk and extreme weather patterns present some of the most pressing risks to the global macroeconomy. As such, public institutions face significant pressure to address the climate risk both through a strengthened regulatory environment that incorporates such risks in the oversight of the financial sector, but also by setting a public example of how they conduct their business and manage their investment portfolios. In addition to climate risk, public investors at times face pressure to exclude companies involved in controversial practices from their

⁵ See Barclays (2016).

portfolios, based both on local norms and values but also legislation. The financial impact on portfolio performance of implementing ESG considerations has in some instances taken a secondary role to targeting the harmful activities, as public investors also need to consider the adverse impact their inaction may have on public perception.

b. Better expected portfolio performance in the long term

Among investors, there is still concern about the potential negative effects on performance from implementing an ESG investment policy. The unease is fuelled by anecdotal evidence, such as the case of the sovereign wealth fund of Norway, which suggests that restrictive negative screening, which significantly shrinks the investable universe, typically results in lower returns.⁶ However, academic research demonstrates that other approaches to ESG investing, such as portfolio optimisation towards higher-rated ESG issuers, ESG integration in portfolio construction, or shareholder engagement, may have positive effects on financial performance in the long term. While ESG investing is still in the early stages of development, and performance analyses may be inconclusive, several studies have shown that bond issuers with higher ESG ratings are associated with better risk-adjusted performance compared with peers with lower ESG ratings over the long term. In their analysis of more than 2,000 empirical studies on the relationship between ESG factors and corporate financial performance, Friede et al⁷ demonstrate that around 90% of studies reveal a non-negative correlation between ESG factors and corporate financial performance, with the majority of studies revealing a positive correlation. The positive impact of ESG factors on corporate financial performance also seems to be stable over time.

3. The role of public initiatives in ESG investing

Several policy initiatives to promote sustainable investments have helped to promote their rapid growth over the past decade. Statistics of the Global Sustainable Investment Alliance suggest that 26% of professionally managed assets in ex-Japan Asia, Australia, Canada, Europe, Japan, New Zealand and the United States qualified as sustainable investments at the beginning of 2016.⁸

The most prominent policy initiative to promote responsible investments internationally is the United Nations-endorsed **Principles for Responsible Investment (PRI)**. Launched in 2006, the PRI is an investor-sponsored initiative in partnership with the UN Environment Programme (UNEP) Finance Initiative and UN Global Compact, which sets forth six voluntary and aspirational principles for incorporating ESG issues into investment practice. At the outset, the principles were undersigned by approximately 100 signatories, who collectively managed USD 6.5

⁶ The loss of over 150 basis points annualised over 12 years from the performance of Norway's sovereign wealth fund was attributed to its negative screening approach.

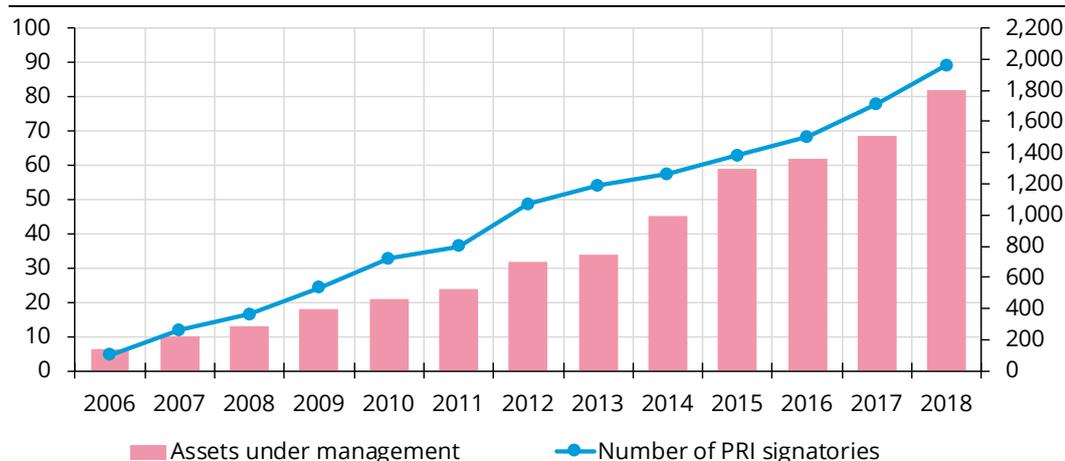
⁷ See eg Friede et al (2015).

⁸ The Global Sustainable Investment Alliance (GSIA) is an international collaboration of membership-based sustainable investment organisations, which publishes a review of sustainable, responsible and impact investing on a biennial basis. The review covers sustainable investment data from ex-Japan Asia, Australia, Canada, Europe, Japan, New Zealand and the United States and was last conducted in 2016.

trillion in assets. Since then, the number of signatories has grown to 1,961, representing a total of USD 82 trillion assets under management in April 2018 (Graph 1). Anecdotal evidence suggests that most asset managers operating globally today are PRI signatories, underscoring their commitment to sustainable investing. Additionally, being a PRI signatory is perceived as a standard for independent third-party verification and transparency of the investments by investors. While becoming a PRI signatory is less relevant for public investors today, public pressure and widespread acceptance of the PRI may result in central banks and international organisations signing the principles in the future.

ESG assets under management and signatories to the PRI
USD trillions; number

Graph 1



Source: PRI. Data as of end-April.

Another, more recent initiative, established in December 2015 by the Financial Stability Board (FSB) at the request of the G20, is the **Task Force on Climate-related Financial Disclosures (TCFD)**.⁹ The TCFD followed the agreements made at the 2015 Paris Climate Conference, where more than 190 countries committed themselves to keep global warming below 2°C and to channel financial flows towards low-carbon and climate-resilient developments.¹⁰ The TCFD was tasked with identifying the information needed by investors, lenders and insurance underwriters to adequately assess and price climate-related risks and opportunities, considering physical, liability and transition risks associated with climate change, and to issue recommendations for voluntary and consistent disclosure of climate-related financial risks in financial reporting. The result of this undertaking was the issuance of four recommendations on climate-related financial disclosures, which are widely adoptable across different industries and jurisdictions globally. Given the involvement of the public sector in this initiative and the fact that users and providers of financial capital increasingly recognise the risks and opportunities inherent in a changing climate¹¹ with effects that could even become a financial stability issue, a more widespread adoption of

⁹ See Task Force on Climate-related Financial Disclosures, www.fsb-tcfid.org/.

¹⁰ See Sustainable Innovation website, www.cop21paris.org/about/cop21/.

¹¹ See N Stern (2016).

climate-related financial disclosures and increasing recognition of the global financial sector's role in relation to global climate change may be expected.

At the European level, the European Commission (EC) launched an initiative in December 2016 that encourages the financial sector to re-orient investments towards more sustainable technologies and businesses, to finance growth in a sustainable manner over the long term and to contribute to the creation of a low-carbon, climate-resilient and circular economy.¹² For this purpose, it established a **High-Level Expert Group (HLEG) on Sustainable Finance**, which advised the EC on the steering of public and private capital flows towards sustainable investments, identified the steps that financial institutions and supervisors should take to protect the stability of the financial system from risks related to the environment and advised the EC on how to deploy these policies on a pan-European scale. The HLEG published an action plan on sustainable finance, which was adopted by the EC in March 2018. Among key actions, the HLEG recommended the establishment of a clear EU classification system (taxonomy) for environmentally sustainable economic activities, an EU Green Bond Standard, the introduction of measures to clarify asset managers' and public investors' duties regarding sustainability and the strengthening of transparency with regard to companies' ESG policies. The establishment of a common taxonomy and the development of a Green Bond Standard may be regarded as particularly valuable, as these standardised classifications might solve one of the above-mentioned deficiencies connected with ESG investing, which is the lack of a commonly accepted definition of what constitutes a sustainable investment. Once these standards are established on a European scale, they might also gain a worldwide relevance.

In emerging market economies (EMEs), the **Sustainable Banking Network (SBN)**, which is a voluntary market-based network of EME financial sector regulatory agencies and banking associations (including central banks), has taken the lead in promoting sustainable finance and addressing climate change in line with international practices.¹³ The SBN is supported by the International Finance Corporation (IFC) and was launched in September 2012 with members from 10 EMEs. By February 2018, the SBN had grown to a network of members from 34 EMEs, representing USD 42.6 trillion in banking assets (about 85% of total banking assets in EMEs). The focus of the SBN is to support the private sector in adapting to developments linked to environmental and social sustainability and to contribute to national sustainable development agendas. As of February 2018, 15 countries had launched national policies, guidelines, principles or roadmaps focused on sustainable banking. They include Bangladesh, Brazil, China, Colombia, Indonesia, Kenya, Mexico, Mongolia, Nigeria, Peru, Turkey and Vietnam. Nineteen others (classified as being at the "initiating stage"), had committed themselves to taking sector-wide action to promote sustainable finance.

Several central banks are also active in ESG investing and have taken action within their jurisdictions in order to ensure that the financial sector begins to address challenges resulting from the economic and financial impact of climate change and/or to promote sustainable finance. A recent initiative in this regard is the

¹² See European Commission website, Sustainable investing, https://ec.europa.eu/info/business-economy-euro/banking-and-finance/sustainable-finance_en/.

¹³ International Finance Corporation, IFC Environmental and Social Performance Standards www.ifc.org/wps/wcm/connect/topics_ext_content/ifc_external_corporate_site/sustainability-at-ifc/company-resources/sustainable-finance/sbn.

Network of Central Banks and Supervisors for Greening the Financial System (NGFS).¹⁴ The NGFS was established in December 2017 by eight central banks and supervisors. Since then, the network has added several other members plus a number of observers, among them the Bank for International Settlements (BIS). The network's purpose is to help strengthen the global response required to meet the goals of the Paris agreement and to enhance the role of the financial sector to manage risks and to mobilise capital for green and low-carbon investments in the broader context of environmentally sustainable developments.

4. Main findings of the informal ESG survey conducted in April–May 2018

In April–May 2018, the BIS Asset Management unit conducted an informal survey of 26 peer institutions and select asset managers that have been at the forefront of sustainable investing, in order to assess sustainable investment practices at other institutions. The sample of surveyed institutions is not representative of the central bank and international organisation communities, but it provides relevant information from the perspective of early adopters of ESG investing strategies. Table 1 summarises the responses across the following seven ESG investment concepts:

1. **ESG investment policy:** the formalised commitment to integrate ESG factors in the investment criteria;
2. **Negative screening:** the exclusion of certain sectors or companies based on specific ESG criteria (values or social norms);
3. **ESG integration:** the systematic inclusion of ESG factors into traditional financial analysis;
4. **Shareholder engagement:** the use of shareholder power to influence corporate behaviour, including through direct engagement with management and/or boards, filing shareholder proposals, and proxy voting that is driven by ESG guidelines;
5. **PRI signatory:** the signing of the UN-endorsed PRI, which is perceived as a public commitment to sustainable investing;
6. **Green bonds:** investments in debt securities that are issued to raise capital specifically to support climate-related or environmental projects; and
7. **Carbon footprint measurement:** the measurement of greenhouse gases produced to directly and indirectly support human activities.

¹⁴ See Bank of France website, Network for greening the financial system, <http://www.banque-france.fr/en/financial/stability/international-role/network-greening-financial-system>.

ESG trends among public institutions and asset managers
In per cent

Table 1

	Central banks	International organisations	Asset managers
ESG investment policy	73%	33%	88%
Negative screening	55%	33%	13%
ESG integration	64%	50%	75%
Shareholder engagement	36%	33%	50%
PRI signatory	0%	33%	88%
Green bonds	91%	67%	0%
Carbon footprint measurement	0%	33%	25%

Source: BIS.

a. ESG investing trends among central banks

In an effort to better understand the evolving ESG investing practices among central bank investors, we contacted a number of institutions that have, in recent years, not only expressed an interest in sustainable finance, but have also made progress in implementing it. The majority of the central banks that have already established ESG investment policies have done so in recent years primarily for their pension funds and own funds, but in some cases also for their foreign reserves. Some central banks also participate in initiatives to promote green finance and to increase transparency on climate risks, such as the above-mentioned Network for Greening the Financial System and the Task Force on Climate-related Financial Disclosures. Our discussions confirmed that sustainable investing policies among central banks not only vary by country and institution but are also dependent on a country's public commitment to the UN's Sustainable Development Goals (SDGs).¹⁵ Most of the surveyed central banks that had established an ESG investment policy were in countries that were actively promoting the UN's SDG agenda.

Public pressure and national laws demanding certain exclusions from investments played an important role in determining whether or not central banks pursued a policy of negative screening. More than half the central banks in the survey had chosen to exclude a number of investments from their portfolios based on negative screening. Some of the banks had taken a norms-based approach in negative screening, establishing their lists of non-investable companies based on national laws, guidelines or international conventions. Others had constructed lists of excluded securities focusing on reputational risks to the central bank. Central banks that had established ESG investment charters for their pension funds focused primarily on the ESG integration approach, as this endorsed a sustainable investment portfolio without reducing investment return expectations. Prior to drafting their charters, they had consulted a number of other public investors, and sought to ensure that the resulting statement emphasised the public good to society. One central bank remarked that their charter had been designed to be implemented within three years, giving them sufficient time to consider any side effects of the sustainable investment policy on the portfolio. Another institution had adopted an ESG investment policy across all asset classes in response to the national financial regulator's requirement.

¹⁵ The UN has established 17 global SDGs to align the interests of investors, companies and society on sustainability matters.

Academic literature advocates that shareholder engagement could be a beneficial instrument in fostering sustainable investing, but most of the surveyed central banks suggested that this remains a costly approach to implement and monitor.

A number of central banks that use external asset managers mentioned that they require fund managers to be PRI signatories and that they use this requirement for screening the asset manager universe. One of the institutions had even considered the possibility of becoming a PRI signatory itself, but the disclosure requirements appeared burdensome and made the application process unappealing.

Most of the central banks we surveyed were also using green bonds in their portfolios. However, in some cases, the use of green bonds was not based on any ESG investment policy, but was motivated primarily by their financial attractiveness. A few of the central banks said that, while green bonds were a suitable first step in establishing an ESG investment policy, they currently focused on a limited number of sectors and represented a very narrow share of the fixed income universe, thereby making large-scale green bond investments unfeasible.

At this stage, central banks had not implemented a carbon footprint measurement process as there was no well established methodology and it remained unclear how the outcome of any measurements should be interpreted. Additionally, there was uncertainty whether the burden of the carbon footprint measurement remained with the company, ie the issuer of a security, or the investor in the company, and whether the reporting by the investor might lead to double-counting of the carbon footprint within the same security.

b. ESG investing trends among international organisations

In addition to central banks, we also held discussions with a number of international financial organisations that are considered peers of the BIS, as well as others who have been pioneers in sustainable investing. While a few of them have yet to adopt a sustainable investment policy, all of these organisations are currently at various stages of considering ESG investments for their pension funds and other investment portfolios. Those that do not yet have an ESG investment policy expect to put one in place in the near future.

The international organisations confirmed that, while ESG investment approaches could take various forms – from negative screening to ESG integration and shareholder engagement – ESG integration was the preferred approach by both those that had already introduced a policy and those that were looking to draft one. There was a general agreement that ESG investing was still in the early stages; hence, even institutions that had adopted a policy early had deliberately moved slowly with implementation, allowing themselves to observe the enhancements around sustainability and modifying their course of action accordingly.

In several of our discussions, these institutions noted that becoming a PRI signatory not only sent a strong public signal of support for responsible investing, but also put pressure internally on the investor to take concrete steps to move the investment portfolio towards more sustainable investment practices.

The international organisations that had invested in green bonds noted that the use of these instruments was viewed as a first step in making the fixed income allocation more sustainable. However, green bonds represented a very minor portion of the overall fixed income market and efforts needed to be made to encourage higher green bond issuance across various sectors.

An additional observation made during the discussions with the pension funds of international organisations was that a number had already started measuring the carbon footprint of their investment portfolios (primarily by using the relevant MSCI module). However, some of them considered this exercise to be part of their long-term portfolio monitoring, and not a reporting tool that gave meaningful information to stakeholders when presented as an isolated measure at a point in time.

c. ESG investing trends among asset managers

In an attempt to better understand ESG investing practices, we also surveyed the asset managers of the BIS Pension Fund. Most have already established firm-wide policies on ESG issues. These ESG investment policies reflect the growing interest in sustainable investing among global investors. They recognise that companies that made an effort to improve their ESG practices also tended to enhance their long-term profitability by becoming more sustainable. By adopting a sustainable investment policy, the asset managers expected to better identify the companies that could deliver higher returns to investors in the long term. Furthermore, asset managers viewed the development of ESG investment policies as part of their fiduciary responsibility to identify potential intangible risks (ie the costs of energy, climate change impact, board representation, alignment of management and shareholder interests etc) that could impact investments.

Most of the managers noted that, unless requested by investors, they would not implement negative screening, as the approach tended to shrink the investable universe and could consequently erode performance. Another evolving trend among asset managers was the establishment of teams and committees in charge of ESG issues, because they recognised that sustainable investing was no longer only an isolated investment style, but instead ESG factors needed to be integrated into all investment processes. These teams served various functions in different organisations, but the overarching role and responsibility was to define the firm's policy and direction in respect of ESG matters. They were responsible for the oversight of the ESG investment policy implementation, engagement with invested companies on ESG issues, ESG integration analysis into the investment process, and working alongside policymakers in developing a global ESG taxonomy. The asset managers' approach to ESG investing had also evolved over time as their knowledge and awareness of ESG topics advanced. They noted that companies were also under more pressure, from both regulatory bodies and shareholders, to increase sustainability reporting. As a result, asset managers had introduced sustainability scoring systems for their portfolios, taken a more active role in dealing with companies represented in their portfolios by focusing on engagement and voting, and, in some cases, included an ESG component in their employee compensation packages.

The asset managers we reviewed assigned different weights to ESG factors, with some of them putting more emphasis on governance issues, particularly in relation to the judicious exercise of shareholder engagement through proxy voting. The asset managers with significant assets under management recognised that shareholder engagement was a powerful tool in influencing companies on ESG matters and had subsequently designated significant resources (both employees and systems) to focus on proxy voting.

Most of the asset managers were PRI signatories and, in some cases, were measuring the carbon footprint of the asset management business. However, they were steering away from measuring the carbon footprint of the portfolios in which they were invested.

Asset managers and public investors used a number of tools to capture ESG data. The majority of them mentioned MSCI ESG data analytics as one of the leading tools in an evolving space. In addition, they made use of Aladdin (which has incorporated the MSCI ESG module), Reprisk (a provider of global business intelligence on ESG risks), ISS for their proxy voting services and the Bloomberg ESG analysis tools.

5. Practical considerations in implementing ESG investing for public investors

While public investors embrace the philosophy of ESG investing, they struggle with the execution of an ESG investment policy, as the implementation costs remain unclear and the path is relatively new and untested. Given that ESG investing addresses different concerns for different investors, it is important to determine at an early stage which goals are being pursued with ESG investing (eg whether it is to avoid reputational risks, or to pursue political priorities, or those of marketing, performance, risk management or beneficiary interests). It is important for central banks to differentiate among the various portfolios (pension funds, own funds, and foreign exchange reserves) when implementing ESG investing. At this stage, central banks enjoy the most flexibility in implementing ESG investing within their pension fund portfolios. This is driven primarily by the pension funds' very diverse strategic asset allocation, which not only allows for investments in fixed income and equities, but also into alternative investment strategies, where ESG investing approaches can be implemented without major obstacles. Similarly, central banks' own funds may be slightly more limited by their strategic asset allocations in terms of the eligible asset classes, but there is still enough scope for allocating to ESG investments in a meaningful way. ESG investing in foreign exchange reserves remains challenging due to the restrictions on the strategic asset allocation, which focuses primarily on short-duration sovereign fixed income. However, the unanimous adoption of the UN SDGs by 193 countries is expected to be followed by the issuance of SDG bonds not only by the private sector and development banks but also by sovereigns. This, along with the evolving ESG ratings for sovereigns, may facilitate the implementation of ESG investing in the future strategic asset allocation of foreign exchange reserves.

Conclusion

As evidenced by the rise in policy initiatives and also the results of the informal BIS survey, ESG investing is becoming a mainstream topic for public investors. However, significant work remains to be done to achieve standardisation in both the ESG lexicon and practices for a smooth implementation of sustainable investing by public investors. As such, it is important for investors to not only determine at an early stage the goals that are being pursued with respect to ESG investing, but to also consider the viability of the practices shared by early adopters. The practical considerations suggested in this paper may help investors achieve a more sustainable investment portfolio in a feasible manner, while leaving space to consider the next stages in ESG investing.

References

Baldinger, M, M Castelli, C Greenwald and P Salman (2017): *Long-term value creation: sustainable investing for sovereign institutions*, UBS Asset Management.

Barclays (2016): *Sustainable investing and bond returns*, <http://www.investmentbank.barclays.com/our-insights/esg-sustainable-investing-and-bond-returns.html>, 31 October.

Barkó, T, M Cremers and L Renneboog (2017): "Shareholder engagement on environmental, social and governance performance", *Finance Working Paper*, European Governance Institute.

Barkó, T, M Cremers and L Renneboog (2017): "Shareholder engagement on environmental, social, and governance performance", *CentER Discussion Paper*, no 2017-040, Tilburg: CentER, Center for Economic Research.

Bernow, S, B Klempner and C Magnin (2017): *From 'why' to 'why not': sustainable investing as the new normal*, McKinsey Consulting, October.

CFA Institute (2018): *Guidance and case studies for ESG integration in equities and fixed income*.

Clark, G, A Feiner and M Viehs (2015): *From the stockholder to the stakeholder: how sustainability can drive financial outperformance*, University of Oxford and Arabesque Partners, March.

Czerwinska, T and P Kazmierkiewicz (2015): "ESG rating in investment: risk analysis of companies listed on the public market in Poland", *Economic Notes*, vol 44, no 2, July, pp 211–48.

DB Climate Change Advisors, Deutsche Bank Group (2012): *Sustainable investing: establishing long-term value and performance*, June.

Deutsche Asset Management (2017): *Measuring physical climate risk in equity portfolios*, November.

Ehlers, T and F Packer (2017): "Green bond finance and certification", *BIS Quarterly Review*, September.

Friede, G, T Busch and A Bassen (2015): "ESG and financial performance: aggregated evidence from more than 2000 empirical studies", *Journal of Sustainable Finance and Investment*, vol 5, no 4, pp 210–33.

Giese, G, L Lee, D Melas, Z Nagy and L Nishikawa (2018): *Foundations of ESG investing, integrating ESG into benchmarks*, MSCI, May.

Global Sustainable Investment Alliance (2016): *Global Sustainable Investment Review*.

Inderst, G and F Stewart (2018): *Incorporating environmental, social and governance (ESG) factors into fixed income investment*, World Bank Group, April.

JP Morgan Perspectives (2018): *ESG investing goes mainstream*, 9 May.

Khan, M, G Serafeim and A Yoon (2016): "Corporate sustainability: first evidence on materiality", *The Accounting Review*, vol 91, no 6, pp 1697–724.

Stern, N (2016): *The economics of climate change: the stern review*, 30 October.

Swiss Sustainable Finance (2017): *Handbook on Sustainable Investments*.

ESG investments: filtering versus machine learning approaches*

Carmine De Franco,[†] Christophe Geissler,[‡] Vincent Margot[§] and Bruno Monnier^{**}

Abstract

We designed a machine learning algorithm that identifies patterns between environmental, social and governance (ESG) profiles and financial performance for companies in a large investment universe. The goal of the algorithm, which falls in the category of supervised machine learning, is to predict the (conditional) excess return of each company over the benchmark, given the specific values taken by some of its ESG indicators (the features). In other words, the algorithm identifies regions in the high-dimensional space of ESG features that are statistically related to financial outperformance or underperformance. The final aggregated predictions are transformed into scores, which allow us to design simple strategies that screen the investment universe for stocks with positive scores. By linking ESG features with financial performance in a non-linear way, our strategy is shown to be an efficient stock picking tool, outperforming classic strategies that screen stocks according to their ESG ratings, such as the popular best-in-class approach. Our paper introduces new ideas into the growing field of financial literature investigating the links between ESG behaviour and the economy. We show, indeed, that there is clearly some form of alpha in the ESG profile of a company, but that this alpha can be accessed only with powerful, non-linear techniques such as machine learning.

JEL classification: D83, G10, G11, G34.

Keywords: best-in-class approach, ESG, machine learning, portfolio construction, sustainable investments.

* Papers in this volume were prepared for a conference co-hosted by the BIS, the World Bank, the Bank of Canada and the Bank of Italy, in Rome, Italy on 22–23 October 2018. 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.

[†] Carmine De Franco PhD, Head of fundamental research at Ossiam (Paris, France), carmine.de-franco@ossiam.com.

[‡] Christophe Geissler CEO at Advestis (Paris, France), cgeissler@advestis-conseil.com.

[§] Vincent Margo, Analyst at Advestis (Paris, France), vmargot@advestis-conseil.com.

^{**} Bruno Monnier, CFA, Head of operational research at Ossiam (Paris, France), bruno.monnier@ossiam.com.

1. Introduction

The relationship between corporate social performance (CSP) and corporate financial performance (CFP) is a fairly old theme in economic research. In its earlier stages, CSP was met with scepticism among critics: Nobel prize-winning economist Milton Friedman wrote in *The New York Times Magazine* in 1970 that "... there is one and only one social responsibility of business – to use its resources and engage in activities designed to increase its profits so long as it stays within the rules of the game, which is to say, engages in open and free competition without deception or fraud...." (Friedman (1970)).

As time went on, however, the number of studies highlighting the positive, or at least non-negative, relationship between CSP and CFP has grown significantly, probably beginning with the initial work by Bragdon and Marlin (1972) on the link between environmental virtue and financial performance. Fifty years later, the number of proponents of CSP and, more broadly, environmental, social and governance (ESG) integration in both corporate management and investors' choices has grown exponentially. As has the number of financial products, funds and exchange-traded funds that offer ESG versions of a large panel of investment strategies (mainly on equity and bonds). The current investment approach now seems in complete contrast to that of Friedman's, with the most recent empirical literature highlighting the link between ESG performance and alpha (Chong and Phillips (2016), Giese et al (2016), Zoltan et al (2016)).

Nevertheless, the question regarding the relationship between CSP and CFP remains largely unanswered. Reviews of published papers (meta-analysis) highlight that most empirical studies published on this theme report a non-negative or weak positive relationship between CSP and CFP (see eg Orlitzky et al (2003), Allouche and Laroche (2005), Wu (2006), Van Beurden and Gössling (2008), Margolis et al (2009), Friede et al (2015)). Other researchers take a more optimistic view and report either a significant relationship between CSP and CFP (Peiris and Evans (2010), Filbeck et al (2014), Indrani and Clayman (2015)) or, at the least, that CSP is not detrimental to CFP as long as one manages to build the portfolio with care, even if there is no clear value added in ESG integration (Kurtz and Di Bartolomeo (2011)).

Although we do not share the very optimistic and mostly overstated enthusiasm about the direct relationship between ESG and financial performance, we do believe a strong relationship exists between ESG and the sustainability of corporate businesses. Specifically, we believe ESG has an impact on financial performance and risks, but not linearly. We welcome the efforts that investors are undertaking to include ESG criteria in their portfolio choices, and hope this will trigger economic and cultural changes in corporate management. At the same time, however, we remain sceptical regarding the far-too flaunted capability of basic ESG ratings to act as an alpha generator in a portfolio. It remains true, however, that ESG data, reports and analysis can contain useful information related to the strengths and weaknesses of corporations. Unfortunately, ESG ratings are, by construction, a composite measure that dramatically reduces this rich set of information.

Our contribution to the growing literature on this topic is to show that, empirically, there is no value added in portfolios based on simple ESG screenings. Although it usually results in no harm to the performance, we do not find any alpha in such approaches. However, by recognising the intrinsic value of the large panel of ESG indicators that are aggregated to form the ESG ratings, we show that it is possible

to extract value from them, which, in turns, translates into real alpha. By exploring large data sets of specific ESG indicators, we are able to identify those that significantly impact corporate financial performance. In a simplified example, we can agree that for a company in the utility sector, the environmental performance can, most likely, be a discriminating criterion for financial performance; at the same time, governance can play an important role if we compare a utility company in an advanced market economy in Europe, for example, with one in an emerging market economy. Similarly, direct carbon emissions for banks are probably not as relevant to them as would the exposures of these banks, through loans, to highly polluting companies. In short, aggregate measures such as ESG ratings lose valuable information contained in the ESG indicators, which therefore lower their predictive power.

Searching for interesting patterns between specific ESG indicators and financial performance for a large set of companies remains out of reach for the standard tools available to econometricians. This search takes place in a high-dimensional space and is not oriented by previously derived information relating to these ESG features. To deal with this complexity, we developed a machine learning algorithm that allows us to identify features and patterns that are relevant to explain the link between CSP and CFP. The algorithm maps the regions in our high-dimensional space of ESG features that have been consistently associated with outperformance or underperformance. In the econometric parlance, we look at those regions for which the conditional expectation of each stock's forward return is statistically positive (or negative), given that its relevant ESG features fall in these regions. We say that these relevant ESG features "activate" the region. By observing the ESG features, we then obtain a significant signal regarding the future financial performance of the stock.

This identification is done with a set of rules that take the form of *If-Then* statements. The *If* statement identifies the region in the ESG space: ie the values that some ESG features must take in order to activate the rule. The *Then* statement produces a prediction of the excess return, over the benchmark, that we can expect from a stock whose ESG features fall in that region. The final prediction is the aggregation of the predictions made by these rules and is transformed into a score [-1, 0, +1]. We therefore focus on the sign of the prediction of excess return rather than on its value. This usually makes the estimation more robust.

The aggregation method mimics a panel of experts, each of whom specialises in an ESG feature (eg environment, independence of the board, ESG reporting verification, employee incidents etc) and makes a prediction given the ESG behaviour of the company. When the aggregated prediction is close to zero, ie the panel of experts is split between optimistic and pessimist forecasters, the final prediction is set at zero. The algorithm is regularly trained over time so that it can react and readjust to the new observed data. The algorithm is used to design a very simple strategy that screens the investment universe and selects all stocks with a positive score. The resulting portfolio is compared with a classic ESG best-in-class portfolio, which consists of all stocks in the investment universe whose ESG ratings are above a given threshold within their peer groups. Our empirical results show that the simple machine learning screened portfolio significantly outperforms the ESG best-in-class approach and the benchmark.

This is in line with the economic belief that ESG data are valuable in assessing financial performance, but also confirms that aggregated ESG ratings are not suited to distinguishing between outperformers and underperformers over the long run.

Even if a perfect distinction is out of reach, our results clearly confirm that there is alpha in the granular ESG data, but the relationship between ESG and financial performance is definitely not linear. Furthermore, the predictive power of the scores vanishes with time. We prove, indeed, that regularly training the algorithm over time and producing up-to-date sets of rules are key components of the superior performance of machine learning when it comes to stock screening.

2. Data

The analyses in this paper are carried out on portfolios based on the investment universe defined by the market capitalisation-weighted MSCI World Index USD, which consists of the largest corporations by market capitalisation listed in the United States, Canada, western Europe, Japan, Australia, New Zealand, Hong Kong SAR and Singapore. Portfolios are calculated in USD and net dividends are reinvested in the portfolio itself. Stock prices and dividends are taken from Thomson Reuters/Datastream. We reconstruct a proxy of the MSCI World Index by using end-of-month compositions as well as proxies for benchmarks in the United States, Europe⁶ and developed economies in Asia.⁷

We also consider sector portfolios derived from the MSCI World Index and the regional benchmarks by filtering stocks that belong to the same sector: consumer staples (CS), consumer discretionary (CD), energy (EN), financials (FI), health care (HC), industrials (IN), information technology (IT), materials (MA), telecommunication services (TL) and utilities (UT).

For each company in the investment universe, we collect ESG ratings from Sustainalytics.⁸ An ESG rating is a comprehensive measure based on three pillars – environment, social and governance – that assesses the strengths and weaknesses of a company along these three directions. The pillars are themselves based on a large set of specific indicators. For the purposes of this study, the composite ESG rating is the arithmetic average of the three ratings – environment (E), social (S) and governance (G) – each of which is itself the combination of roughly 50 narrower indicators. Finally, for each company, we consider its relative peer group, which consists of all companies with a similar business, hence comparable from a sustainability point of view. ESG data are available from 2009 onwards, collected with a relatively stable methodology and uniform coverage. All portfolios presented in the following sections are rebalanced on a monthly basis, at the end of every month, with a four-day lag between data extraction and a portfolio's implementation. Portfolios are benchmarked against classical cap-weighted, liquid and investable portfolios, a standard practice in the financial industry. Through the entire analysis, the term "alphas" refer to CAPM-alphas unless stated otherwise.

⁶ Stocks in the MSCI World Index domiciled in Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom.

⁷ Stocks in the MSCI World Index domiciled in Australia, Hong Kong SAR, Japan, New Zealand and Singapore.

⁸ One of the largest providers of ESG ratings.

3. The best-in-class approach

One of the most popular approaches to embed ESG criteria in the portfolio construction process is the so-called best-in-class approach.

Given a threshold x , one excludes the stocks whose ESG ratings belong to the lowest x -quantile. The exclusion is usually carried across peer groups, ie groups of stocks with very similar characteristics. The reason behind this is twofold:

- Removing stocks with low ESG ratings within peer groups ensures that the final economic mesh of the filtered universe remains similar to the initial investment universe.
- ESG ratings have a structural, sector-driven bias that usually favours specific sectors (eg IT or health care sectors) while penalising others (eg energy or utilities). Given this bias, the filtering of peer groups makes comparisons of ESG ratings independent of the sectors.

For the purpose of this study, an ESG best-in-class portfolio derived from a market capitalisation-weighted portfolio removes, within each peer group, the stocks whose ratings belong to the lowest x -quantile. The portfolio is finally scaled to sum up to one. This approach, quite popular among investors, should not be thought of as a way to enhance performance. As Tables 1–4 show, ESG best-in-class filters applied to standard market capitalisation-weighted indexes do not lead to outperformance.

Except for Europe and relatively low threshold levels, we find small but negative excess returns and negative information ratios for the ESG best-in-class portfolios over their benchmarks with almost unchanged risks. Although the approach does not create outperformance per se, it does not carry structural underperformance either. Optimistically, one could accept the fact that embedding ESG objectives in a portfolio does not significantly modify its risk/return profile.

Our findings are not in contradiction with the large literature that finds positive links between ESG and financial performance. But the consistency and durability over time of the ESG factor has been questioned since the very beginning. Aupperle et al (1985) finds no significant relationship between social responsibility and corporate profitability, and similar results were obtained in Capelle-Blancard and Monjon (2012) and Humphrey and Tan (2014). Griffin and Mahon (1997) report that a correlation between financial performance and social performance depends on the measure used to distinguish between high and low social performers.

Table 1: World Developed

	ESG best-in-class			
	Bench	10%	30%	50%
Ann performance	10.07%	10.01%	9.93%	9.51%
Ann volatility	13.34%	13.31%	13.44%	13.82%
Sharpe ratio	0.73	0.73	0.72	0.67
Max drawdown	-21.91%	-21.79%	-22.02%	-22.57%
Information ratio	-	-0.27	-0.25	-0.41

Table 2: United States

	ESG best-in-class			
	Bench	10%	30%	50%
Ann performance	13.45%	13.25%	13.46%	13.2%
Ann volatility	14.61%	14.54%	14.49%	14.4%
Sharpe ratio	0.9	0.89	0.91	0.9
Max drawdown	-18.99%	-18.87%	-18.71%	-18.04%
Information ratio	-	-0.71	0.02	-0.18

Table 3: Europe

	ESG best-in-class			
	Bench	10%	30%	50%
Ann performance	6.37%	6.55%	6.47%	6.31%
Ann volatility	19.25%	19.19%	19.19%	19.29%
Sharpe ratio	0.32	0.33	0.32	0.31
Max drawdown	-30.25%	-30.21%	-30.2%	-30.54%
Information ratio	-	0.43	0.22	-0.11

Table 4: Asia-Pacific

	ESG best-in-class			
	Bench	10%	30%	50%
Ann performance	6.83%	6.71%	6.41%	5.75%
Ann volatility	15.54%	15.71%	16%	16.2%
Sharpe ratio	0.42	0.41	0.38	0.34
Max drawdown	-24.8%	-24.95%	-25.27%	-25.8
Information ratio	-	-0.21	-0.36	-0.46

Key performance indicators of the MSCI World Index and three market capitalisation-weighted regional benchmarks, together with ESG best-in-class filtered portfolios with different thresholds: 10%, 30% and 50%. Data are shown in USD from August 2009 to March 2018.

Source MSCI, Datastream, Sustainalytics.

Our results are more in line with Revelli and Viviani (2015), for which "... the consideration of corporate social responsibility in stock market portfolios is neither a weakness nor a strength compared with conventional investments...". It should be noted that many fund managers and institutional investors surveys report that ESG is mostly viewed firstly as a risk mitigation tool (Van Duuren et al (2016)) and eventually as a long-term performance driver. We share the optimistic view of Nobel prize-winning economist Robert Shiller, for which both society and the financial community would find the use of socially responsible practices mutually beneficial (Shiller (2013)). At the same time, we also believe that short- to medium-term financial performance is, at best, weakly correlated to ESG ratings, at least for such broad investment universes as the MSCI World Index (which contains more than 1,600 companies). We find this to be the case for several reasons:

- i. The investment universes are relatively large and the aggregated ESG ratings have too low a signal-to-noise ratio to allow for an efficient selection of outperforming stocks.
- ii. ESG ratings are global metrics that embrace environmental, social and governance criteria. As such, they may be too reductive, and we may lose a significant amount of information from the single indicator to the aggregated scores.
- iii. Granularity is key: as an example, it is likely that companies in specific sectors (eg energy) react differently to changes in the E score compared with the S score.
- iv. In the search for a rational economic theory behind ESG, some argue that by divesting low ESG-rated companies, investors raise their cost of capital and, in turn, the return these companies have to offer to attract new investors. As such, in the short run, they may show higher performance, but over time, the level of return they have to offer becomes unsustainable. In other words, the action of divesting may take time to materialise in both investors' portfolios and low ESG-rated companies (see eg Asness (2017)).
- v. The period considered in this study spans from the earlier stages of the recovery in 2009 to March 2018. Therefore, we consider key performance indicators over

a period of strong equity market, characterised by high returns and historically lower levels of volatility. This market regime can potentially affect the overall strength of ESG filtered portfolios.

To illustrate item (iii), we consider sector portfolios derived from the MSCI World Index and from the three regional benchmarks (the United States, Europe and Asia-Pacific) and we apply both ESG and single pillars E, S and G, at 30% best-in-class filtering. Tables 5–8 show the results. For the sake of simplicity, we only show annualised excess returns over the relative benchmark sector portfolios and information ratios.

Overall, it is not straightforward to detect clear patterns between excess returns and ESG metrics conditionally to the regional benchmarks. But we can definitely detect specific triplets sector/region/metric that produce significant positive excess returns. Clearly, integrating ESG criteria in the utilities sector enhances in-sample performance. But the right metric to use clearly depends on the geography: in the world developed region (Table 5) the best excess return for the utilities sector is achieved when one uses the G score only at 0.82%; in the United States (Table 6) it is better to look at the S rating under which utilities achieve 0.7%. In Europe (Table 7) it is with the E score that utilities obtain the best result with 3.25%, while in Asia-Pacific (Table 8) it is, once again, the composite ESG rating that achieves the highest excess return at 3.07%.

More generally, there is no sector nor metric for which the excess return of the best-in-class filtered sector achieves a positive excess return in all the regions. Similarly, there is no region nor sector for which all metrics produce positive excess returns. Finally, no sector achieves positive excess returns across all regions and metrics. In other words, finding performance drivers when integrating ESG criteria in a best-in-class fashion is out of reach.

From Tables 5–8, only 12 out of 40 sector/metric portfolios in the world developed region turn out to have a positive excess return, and seven of them are obtained when one considers the G score. In the United States, we find positive excess returns in 12 out of 40, with no clear indication on the best metric to use. We notice, however, that all the metrics seem to work in the utilities sector. In Europe, we count 22 out of 40 sector/metric pairs with positive excess returns. For four sectors (consumer discretionary, materials, telecommunication services and utilities) all metrics work accurately. In Asia, we have 16 out of 40 portfolios with positive excess returns, with no clear patterns between sectors and metrics, except for the energy sector, for which all metrics produce positive excess returns, even if their magnitudes are relatively small.

Table 5: World Developed

	ESG	E	S	G
CD	-0.33% (-0.3)	-0.46% (-0.46)	-0.65% (-0.4)	-1.22% (-0.65)
CS	-0.26% (-0.26)	-0.45% (-0.47)	-0.65% (-0.59)	0.62% (0.53)
EN	-0.24% (-0.13)	-0.41% (-0.19)	-0.26% (-0.17)	0.86% (0.31)
FI	-0.51% (-0.43)	-0.46% (-0.42)	-0.87% (-0.63)	0.19% (0.13)
HC	-0.39% (-0.35)	-0.5% (-0.3)	-0.47% (-0.4)	-0.53% (-0.43)
IN	-0.31% (-0.28)	-0.32% (-0.27)	-0.32% (-0.31)	-0.33% (-0.19)
IT	-0.13% (-0.12)	-0.44% (-0.45)	-1.53% (-0.52)	0.1% (0.05)
MA	-0.09% (-0.05)	-0.09% (-0.03)	-0.17% (-0.08)	0.32% (0.18)
TL	1.26% (0.64)	1.17% (0.74)	0.15% (0.06)	0.73% (0.21)
UT	0.16% (0.1)	-0.94% (-0.39)	0.72% (0.43)	0.82% (0.32)

Table 6: United States

	ESG	E	S	G
CD	-1.95% (-0.82)	-0.48% (-0.39)	-0.91% (-0.55)	-2.92% (-0.98)
CS	-0.94% (-0.84)	-0.46% (-0.43)	-0.46% (-0.32)	-0.26% (-0.12)
EN	-0.09% (-0.06)	-0.66% (-0.44)	0.01% (0.01)	0.55% (0.16)
FI	-0.22% (-0.19)	-0.11% (-0.11)	-0.2% (-0.17)	1.02% (0.67)
HC	-0.23% (-0.22)	0.19% (0.15)	-0.55% (-0.5)	-0.58% (-0.37)
IN	0.4% (0.54)	0.25% (0.32)	0.01% (0.01)	-0.6% (-0.56)
IT	-1.36% (-0.96)	-0.7% (-0.75)	-1.76% (-0.51)	-0.37% (-0.11)
MA	0.34% (0.18)	0.72% (0.32)	-0.76% (-0.32)	-0.14% (-0.07)
TL	-0.38% (-0.37)	-0.32% (-0.38)	-0.4% (-0.2)	1.32% (0.28)
UT	0.63% (0.56)	0.47% (0.38)	0.7% (0.7)	0.13% (0.12)

Table 7: Europe

	ESG	E	S	G
CD	0.51% (0.32)	0.45% (0.29)	0.18% (0.14)	1.07% (0.68)
CS	0.16% (0.14)	0.03% (0.03)	-0.01% (-0.01)	0.39% (0.32)
EN	-1.2% (-0.3)	-2.01% (-0.38)	-0.5% (-0.12)	-1.05% (-0.27)
FI	0.31% (0.18)	0.04% (0.03)	0.33% (0.17)	-0.28% (-0.21)
HC	-0.38% (-0.49)	-0.44% (-0.55)	-0.53% (-0.67)	-0.11% (-0.18)
IN	0.03% (0.03)	0.03% (0.03)	-0.39% (-0.35)	-0.67% (-0.44)
IT	-0.63% (-0.29)	-1.13% (-0.49)	0.01% (0.08)	-0.71% (-0.43)
MA	0.12% (0.03)	0.23% (0.06)	0.51% (0.14)	0.6% (0.18)
TL	1.81% (0.67)	1.72% (0.63)	1.82% (0.62)	2.06% (0.71)
UT	1.79% (0.54)	3.25% (0.87)	0.09% (0.03)	0.41% (0.16)

Table 8: Asia-Pacific

	ESG	E	S	G
CD	-0.29% (-0.19)	-0.11% (-0.11)	0.09% (0.04)	0.75% (0.37)
CS	0.6% (0.52)	0.37% (0.12)	-0.3% (-0.28)	-0.18% (-0.09)
EN	0.12% (0.04)	0.34% (0.11)	0.03% (0.01)	0.25% (0.1)
FI	-0.27% (-0.16)	-0.09% (-0.06)	-0.71% (-0.46)	-0.08% (-0.05)
HC	0.29% (0.19)	-0.67% (-0.27)	0.14% (0.08)	-0.03% (-0.02)
IN	-0.76% (-0.39)	-0.41% (-0.22)	-0.42% (-0.22)	-0.61% (-0.35)
IT	-1.32% (-0.59)	-0.92% (-0.4)	-1.01% (-0.43)	-0.96% (-0.28)
MA	-0.29% (-0.26)	0.04% (0.03)	-0.58% (-0.44)	-0.07% (-0.05)
TL	0.17% (0.04)	-0.18% (-0.11)	-0.61% (-0.09)	1.83% (0.24)
UT	3.07% (0.65)	-3.57% (-0.87)	0.16% (0.04)	2.6% (0.55)

Annualised excess returns (information ratios) between market capitalisation-weighted sector portfolios and their ESG best-in-class filtered versions for the MSCI World Index and the derived regional benchmarks. In bold pairs sector/indicator for which the excess return is positive. Best-in-class filters are performed with the ESG rating together with the single pillars environment (E), social (S) and governance (G) ratings. Data are shown in USD from August 2009 to March 2018. Source: MSCI, Datastream and Sustainalytics. The rows correspond to the standard GICS sectors: consumer staples (CS), consumer discretionary (CD), energy (EN), financials (FI), health care (HC), industrials (IN), information technology (IT), materials (MA), telecommunication services (TL) and utilities (UT).

In conclusion, our empirical findings confirm that simple ESG filtering does not result in better performance. Rather, it behaves as a negative factor, reducing performance. Given the short period we consider, and the market regime that equity markets have experienced since 2009, we share the view that ESG best-in-class integration is, most likely, neutral to financial performance. Nevertheless, our results highlight the fact that geographies and sectors do not react to ESG criteria in the same way. But finding interesting and statistically significant patterns between ratings, pillars, their underlying narrow indicators (*features*) and financial performance, for more than 150 indicators on more than 1,600 companies in the MSCI World Index, over a roughly 10-year period, is out of reach for both human and linear statistical tools. The next section introduces other techniques that can overcome this complexity and exploit this huge set of data.

4. Machine learning

In this section, we introduce a deterministic, easily understandable machine-learning prediction algorithm, aimed at finding consistent and statistically significant patterns between ESG ratings and financial performance. The algorithm explores a high-dimensional data set of ESG granular indicators for all the companies in our investment universe. The goal of the algorithm, which falls in the category of supervised machine learning, is to predict the (conditional) excess return of each company over the benchmark, given the specific values taken by some of its ESG indicators (the features). In other words, the algorithm identifies regions in the high-dimensional space of ESG features that are statistically related to financial outperformance or underperformance. Features include raw and derived ESG indicators,⁹ sector and country classifications, company size and controversy indicators.

The regions are characterised by *rules* in the form *If-Then*, so that the algorithm finally consists of a set of such rules. The *If* statement is a list of conditions on the features $x_t \in X = X_1 \times X_2 \dots \times X_d$, where X_i is the set of possible outcomes of the feature i and $d=447$ is the total number of features.¹⁰ Therefore, a rule defines a hyper-rectangle of X . The *Then* statement is the prediction of the three-month forward excess return conditional to the *If* statement. Since the rules correspond to hyper-rectangles in the feature space, we finally obtain relatively simple and understandable regions. Furthermore, to avoid overfitting, the algorithm only selects a finite number of such rules. At each time t , the predictions of each rule are aggregated into one prediction, \hat{y} , through convex combination. The algorithm is calibrated (*trained*) on the training set and the rules are used out-of-sample. The learning process works at two independent levels:

⁹ For each raw indicator, as for example the E score, we also look at the derived indicator relative to the peer group and the sector. All these transformations can potentially contain useful information. On the other side, the use of both raw and derived indicators rapidly increases the dimension of the feature space d .

¹⁰ We use 164 ESG raw indicators, from which we derive peer group and sector relative indicators and three valuation indicators. In total $164 \times 3 + 3 = 495$. From these indicators we remove 48 indicators for which either the sector or the peer group derived indicators are too close, or for which historical data are missing.

- At the end of year $N+1$ we train the algorithm on an expanded data set of features and stock total returns that contain the data set used at the end of year N augmented by all the new observed data (features and stock total returns) from the end of year N to the end of year $N+1$. To initialise the algorithm, we train it over three years of data (from 2009 to 2012). By expanding the data set, the algorithm is able to access new data and explore new patterns, so that it can strengthen or nuance some rules that were previously discovered.
- Daily, the algorithm can update the weights used to aggregate each rule's prediction, by overweighting rules with a good prediction rate and underweighting the others. Therefore, following day predictions will benefit from the *experience* the algorithm is gaining on the rules and their predictive power. The weight of each rule can be viewed as a confidence index. Of course, this is possible because the algorithm is able to assess the goodness of its predictions by looking at the realised three-month return.

To avoid threshold effects, we transform the final prediction for each stock into a score $S \in \{-1, 0, +1\}$, where $+1$ stands for significantly positive excess return prediction, -1 for negative prediction and 0 for an uncertain prediction. The case where $S=0$ is usually related to stocks for which some of their ESG indicators would eventually signal financial outperformance, while other ESG indicators rather signal potential underperformance. The picture is then nuanced, and the algorithm cannot make a precise prediction. This is a very common situation in finance, where different indicators can yield different forecasts, so that, in aggregate, the forecast turns out to be uninformative. The learning process is divided into two steps. Following Nemirovski (2000) and Tsybakov (2003), the training set D_N at the end of year N is divided into two sub-data sets: D_n the learning set and D_t the aggregation set, with $t \gg n$ and $n + t = N$. The learning set D_n is used to design and select the set of rules used by the algorithm to make predictions. The aggregation set is used to fit the coefficients of the convex combination, in line with the *expert aggregation theory* of Cesa-Bianchi and Lugosi (2006) and Stoltz (2010).

Independent suitable rules. Let $D_N = ((x_1, y_1), \dots, (x_N, y_N)) \in (X \times \mathbb{R})^N$ be the training set. Here y_i denotes the three-month return for some stock and x_i is the d -dimensional vector of its ESG features. The training set consists of a large but finite number of $(d+1)$ -vectors spanning all stocks in the investment universe and all available dates. The training set $D_n \subseteq D_N$ includes the first n data points in D_N and $D_t = ((x_{n+1}, y_{n+1}), \dots, (x_N, y_N))$ the order being induced by the time.

Definition 4.1. For any set $E \subset X$, we define

$$\mu(E, D_n) := \frac{\sum_{i=1}^n y_i \mathbf{1}_{x_i \in E}}{\sum_{i=1}^n \mathbf{1}_{x_i \in E}}$$

where, by convention, $0/0 = 0$.

The set-valued map μ represents the conditional excess return of a stock over the benchmark, given that its ESG features x belong to E .

Definition 4.2. Let \mathbf{r} be a hyper-rectangle on X : $\mathbf{r} = \prod_{k=1}^d I_k$ where each I_k is an interval of X_k . A rule f is a function defined on $\mathbf{r} \times (X \times \mathbb{R})^N$ as

$$f(x, D_n) =: \mu(\mathbf{r}, D_n), \forall x \in \mathbf{r} \quad (4.1)$$

The hyper-rectangle \mathbf{r} is called the **condition** and $\mu(\mathbf{r}, D_n)$ is called the **prediction** of the rule f . The event $\{x \in \mathbf{r}\}$ is called the **activation conditions** of the rule f .

A rule f is completely defined by its condition \mathbf{r} . So, with an abuse of notation, we do not distinguish between a rule and its condition. We define two crucial numbers for a rule:

Definition 4.3. Let f be a rule as in Definition 4.2 defined on $\mathbf{r} = \prod_{k=1}^d I_k$.

a. The **number of activations** of f in the sample D_n is

$$n(\mathbf{r}, D_n) := \sum_{i=1}^n \mathbf{1}_{x_i \in E}$$

b. The **complexity** of f is

$$cp(\mathbf{r}) := d - \#\{i \leq k \leq d \mid I_k = X_k\}$$

The algorithm does not consider all the possible rules, but only those with a given *coverage* and *significance*. We call these rules *suitable*, and their definition is given below.

Definition 4.4. A rule f , defined on \mathbf{r} , is a **suitable rule** for the training set D_n if and only if it satisfies the two following conditions:

a. **Coverage condition**

$$C_{min} \leq \frac{n(\mathbf{r}, D_n)}{n} \leq C_{max} \quad (4.2)$$

with C_{min} and C_{max} suitably chosen in the calibration step.

b. **Significance condition**

$$|\mu(\mathbf{r}, D_n) - \mu(X, D_n)| \geq z(\mathbf{r}, D_n, \alpha) \quad (4.3)$$

for a chosen $\alpha \in [0, 1]$ and a function z .

The coverage condition (4.2) excludes rules that are activated only on small sets (ie with a low coverage rate, C_{min}) and rules that are too obvious (ie with a high coverage rate, C_{max}). The threshold in the significance condition (4.3) is set such that the probability of falsely rejecting the null hypothesis $\mu(\mathbf{r}, D_n) = \mu(X, D_n)$ is less than α . The parameter α permits to control the number of suitable rules. The higher α , the higher the number of suitable rules. In what follows, we generate rules of complexity $c \geq 2$ by a *suitable intersection* of rules of complexity 1 and rule of complexity $c-1$.

Definition 4.5. Two rules f_i and f_j defined on r_i and r_j respectively, form a **suitable intersection** if and only if they satisfy the two following conditions:

a. Intersection condition

$$\begin{aligned} r_i \cap r_j &\neq \emptyset, \\ n(r_i \cap r_j, D_n) &\neq n(r_i, D_n), \\ n(r_i \cap r_j, D_n) &\neq n(r_j, D_n) \end{aligned} \quad (4.4)$$

b. Complexity condition

$$cp(r_i \cap r_j) = cp(r_i) + cp(r_j) \quad (4.5)$$

The intersection condition (4.4) avoids adding a useless condition for a rule. In other words, to define a suitable intersection, r_i and r_j must not be satisfied by the same points in D_n . The complexity condition (4.5) means that r_i and r_j have no marginal index in common.

Designing suitable rules. The design of suitable rules is made recursively on their complexity. It stops at a complexity c if no rule is suitable or if the maximal complexity $c = cp_{max}$ is achieved.

Complexity 1: The first step is to find suitable rules of complexity 1. First notice that the complexity of evaluating all rules of complexity 1 is $O(ndm^2)$. Rules of complexity 1 are the base of the algorithm search heuristic. So, all rules are considered, and only suitable ones are kept, ie rules that satisfied the coverage condition (4.2) and the significance condition (4.3). Since rules are considered independently, the search can be parallelised.

Complexity c: Among the suitable rules of complexity 1 and $c-1$, we select M rules of each complexity (1 and $c-1$) according to a chosen criterion. Then it generates rules of complexity c by pairwise *suitable intersection* according to Definition 4.5. The complexity of evaluating all rules of complexity c , obtained from their intersections, is $O(nM^2)$. Here again, since rules are considered independently, the evaluation can be parallelised. The parameter M helps to control the computing time.

Selecting suitable rules. We select a subset S from all suitable rules which maximises the gains expected from rule in D_n and such as their conditions form a covering of X .

Algorithm. The calibration of the algorithm is structured in two parts: in the first one, it finds all suitable rules, and in the second one it retains only an optimal subset of it. To avoid threshold effects, overfitting and to manage the numerical complexity, we discretise each feature in X into m classes with empirical quantiles (modalities).¹¹ Thus, each modality of each variable covers about $100/m$ percent of the sample. In practice, m must be inversely related to d : The higher the dimension of the problem, the smaller the number of modalities.

¹¹ Of course, such procedure is performed only on real-valued features with more than m different values. Categorical features are left unchanged.

The parameters of the algorithm are:

- m , the sharpness of the discretisation;
- $\alpha \in [0, 1]$, which specifies the false rejecting rate of the test;
- z , the significance function of the test;
- C_{max} and C_{min} the coverage bounds;
- cp_{max} the maximal complexity of a rule; and
- $M > 0$, the number of rules of complexity 1 and $c-1$ used to define the rules of complexity c .

Aggregation. Let $D_t = ((x_{n+1}, y_{n+1}), \dots, (x_N, y_N)) \in (X \times \mathbb{R})^N$, where $n + t = N$ be the aggregation set and let \mathcal{S} be the set of R rules selected by the algorithm. At each time t , the predictions of each rule are aggregated into one prediction \hat{y}_t as follows:

$$\hat{y}_t = \frac{\sum_{i=1}^R \pi_{i,t} f_i(x_t, D_n)}{\sum_{i=1}^R \pi_{i,t} \mathbf{1}_{x_t \in r_i}} \quad (4.6)$$

with $\pi_{i,t} = 1/R$. When the realised value y_t is known, the weights $\pi_{i,t+1}$ are updated with the following formula:

$$\pi_{i,t+1} = \pi_{i,t} \frac{\exp(-\eta l(f_i(x_t, D_n), y_t))}{\sum_{k=1}^R \pi_{k,t} \exp(-\eta l(f_k(x_t, D_n), y_t))} \quad (4.7)$$

with $\eta > 0$ and l a convex loss function.

Remark 4.6. One can notice that $f_i(x_t, D_n)$ is not defined if $x_t \notin r_i$. In (4.6), \hat{y}_t is well defined for all x_t , since the set \mathcal{S} is a covering of X . In (4.7) we follow the methodology of the **sleeping expert aggregation** from Devaine et al (2013). Once trained, the machine learning algorithm produces predictions of the excess returns, which are transformed into a scores $S \in \{-1, 0, +1\}$, given the out-of-sample ESG features x_t for each company. Table 9 shows some examples of rules taken from the learning process of the algorithm. The table lists three rules associated with positive predictions (opportunities) and five rules with negative predictions.

Opportunity rules: positive excess return

Table 9

Feature	Relative to	Activation set	Rule description
Business ethics incidents	Sector	[5.9]	WHEN business ethics incidents is high relative to sector AND board remuneration disclosure is high relative to sector THEN opportunity
Board remuneration disclosure	Sector	[5.9]	
Board independence	All	[9.9]	WHEN board independence is at the maximum AND board remuneration disclosure is high relative to sector THEN opportunity
Board remuneration disclosure	Sector	[5.9]	
Board independence	All	[9.9]	WHEN board independence is at the maximum AND business ethics incidents is high relative to sector THEN opportunity
Business ethics incidents	Sector	[5.9]	

Risk rules: negative excess return

Feature	Relative to	Activation set	Rule description
Verification of ESG reporting	Sector	[0.7]	WHEN verification of ESG reporting is not high relative to sector AND board remuneration disclosure is low relative to sector THEN risk
Board remuneration disclosure	Sector	[0.4]	
Quantitative performance	All	[5.9]	WHEN quantitative performance score is high AND board remuneration disclosure is low relative to sector THEN risk
Board remuneration disclosure	Sector	[0.4]	
Verification of ESG reporting	All	[0.6]	WHEN verification of ESG reporting is not high AND quantitative performance score is high THEN risk
Quantitative performance	All	[6.9]	
Gender diversity of board	Peer group	[0.8]	WHEN gender diversity of board is not high relative to peer group AND employee incidents is very low relative to peer group THEN risk
Employee incidents	Peer group	[0.2]	
Green logistics programmes	Delta score	[0.2]	WHEN green logistics programmes delta score is very low AND qualitative performance delta score is very low THEN risk
Qualitative performance	Delta score	[0.2]	

Some rules from the learning process of the algorithm at end 2012, 2013 and 2016. All features are discretised over 10 modalities (zero to nine) except for qualitative performance, which is discretised over six modalities (zero to five). High values for the features correspond to good ESG performance.

Each rule consists of two features and two intervals. The “relative to” properties indicate whether the feature must be calculated over all stocks in the universe (all), over a sector, over a peer group, or whether we should look at the variations of the feature over time (delta score).

Whenever the values taken by the features for a given company fall in the given intervals (we say that the stock activates the rule) the algorithm makes a prediction on its excess return. It is important to remark that we aggregate all the predictions, and we transform the final aggregated prediction into a score $S \in \{-1, 0, +1\}$, so that in the end we mainly look at the sign of the prediction rather than at its magnitude. We also remark that, while the set of rules remains unchanged for one year (until the next learning process), the output of the rules can change over time, because raw indicators can change and because the aggregated weights of the rules change over time.

Finally, the use of granular, rich ESG data are a key element of the power of the machine learning algorithm: indeed, it works quite poorly if one only considers aggregated E, S and G scores.

5. Machine learning application

We now compare the predictive power of the machine learning algorithm developed in Section 4 with the classical best-in-class approach. More precisely, we try to assess whether filtering stocks over scores derived from the algorithm outperforms the standard filtering over ESG ratings (best-in-class).

For the sake of simplicity, we only present the “world developed” universe and, among the strategies presented in Section 3, we only consider the 30% best-in-class, as it is very close to what investors look at for their ESG portfolios.

We recall that this strategy excludes, at each monthly review, the stocks whose ESG ratings are in the lower tercile within each peer group, and finally scale the weights so that their sum is one. To insure replicability of the portfolio, the ESG ratings are taken four days before the review date (which is end-of-month). At the monthly review, we also build three portfolios based on the scores calculated with the machine learning algorithm, with the rules calculated at the end of the year that precedes the review:

Positive ML screening: The portfolio selects all stocks in the investment universe whose scores are $+1$. The weights are finally scaled to sum up to one (maintaining the market capitalisation-weighting scheme of the benchmark)

Positive ML screening sector-matched: Same selection as for the positive ML screening portfolio, but the scaling of the weights is done in such a way that the final sector breakdown of the portfolio is matched to the benchmark’s one.

Negative ML screening: The portfolio selects all stocks in the investment universe whose scores are -1 . The weights are finally scaled to sum up to one (maintaining the market capitalisation-weighting scheme of the benchmark)

As before, the scores are taken four days before the review date. We consider the sector-matched portfolio because the absolute screening usually introduces significant sector deviations with respect to the benchmark.

It should be noticed that market capitalisation-weighted portfolios have some drawbacks: they are trend-following and show sector concentrations. However, this is relatively limited in our case as the benchmark is a large and relatively well diversified portfolio, where even large cap stocks rarely exceed 3% of the index. We do not report results relative to the equally-weighting scheme (1/N) as they are very similar to the cap-weighted scheme in relative terms.

Alternatively, one could use the standard mean-variance approach. We do not consider it in the research process because for this approach one should design a covariance estimation procedure, which may introduce noise and subjectivity in the portfolio construction, and define a procedure to calculate expected return, either model-based, or estimated from past data or again in a Black-Littermann fashion. In both cases, the sensitivity of the strategy to the set of procedures (and corresponding parameters) will play a key role in the outcome. And associated turnover will be very high when compared to cap-weighted selections. Our choice of a market capitalisation-weighted scheme is therefore the best way to assess the power of our ML algorithm as it is simple, stable and produces low-turnover portfolios.

Table 10 collects the main results for these portfolios using data since January 2013. The sample length is driven by the availability of good and uniform quality ESG

data (since 2009) and the initial three years of data needed for the first training of the ML algorithm. Although we recognise that the period over which we can test the machine learning algorithm is relatively short (five years and three months), the results we obtain contain some interesting insights.

Table 10

	Machine learning screening				ESG best-in-class
	Bench	Positive	Positive sect matched	Negative	30%
Ann performance	10.32%	13.07%	11.66%	8.31%	10.13%
Ann volatility	10.50%	11.14%	10.96	10.95%	10.57%
Sharpe ratio	0.94	1.14	1.03	0.72	0.92
Max drawdown	-18.07%	-14.99%	-16.46%	-22.47%	-17.91%
Information ratio	-	1.01	0.58	-0.54	-0.32
Ann CAPM alpha	-	2.47%	1.15%	-1.81%	-0.24%

Key performance indicators of the MSCI World Index (bench), the market capitalisation-weighted selection filtered over positive scores from the ML algorithm, the one with the sector allocation matched to the benchmark, the one screened over negative scores and the 30% ESG best-in-class filtered portfolios. Data are shown in USD from January 2013 to March 2018.

Source: MSCI, Datastream, Sustainalytics.

First, the positive ML screening outperforms all the other portfolios: the benchmark on an annualised basis by 2.76%, the ESG best-in-class portfolio by 2.94% and the negative ML screening by 4.77%. And while the realised annual volatilities remain in the range 10.50% to 11.14%, there are significant differences in the realised maximum drawdowns: the negative ML screening shows a -22.47% loss from its peak, while the positive ML screening loss from its peak accounts for -14.99%.

These two combined results show that the machine learning algorithm is clearly able to distinguish between opportunity stocks (the ones with positive scores) from risky stocks (negative scores). Figure 1 shows the historical behaviour of these two portfolios and the benchmark. We notice that the positive ML screening outperforms the negative ML screening over time, with the benchmark in between.

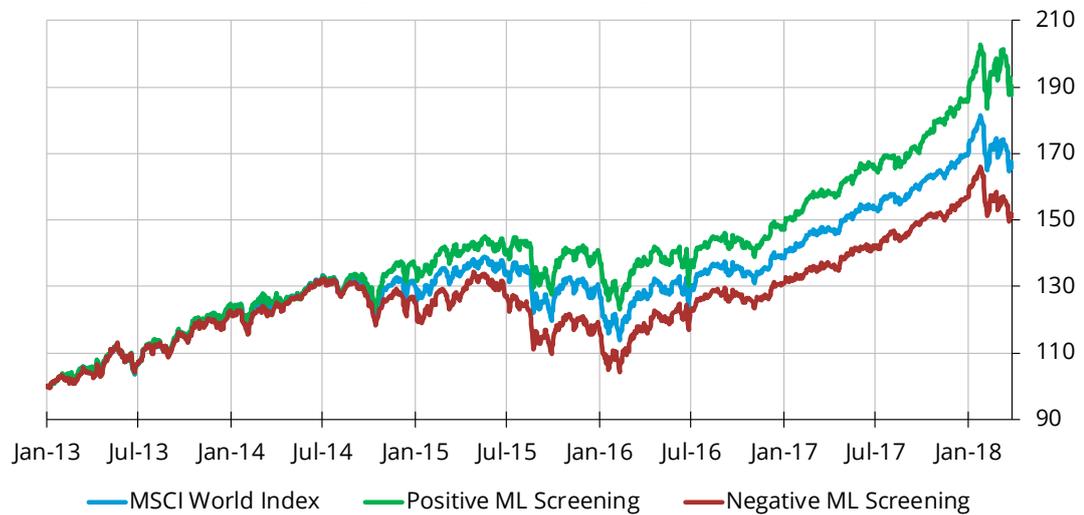
Furthermore, in years when the benchmark shows very high performance with very low volatility, typically in bull market regimes, the differences between the two strategies are less pronounced. On the contrary, when the market is in a bear regime or has no clear trend, the positive ML screening clearly outperforms its negative counterpart, as shown in Table 11.

In years when the benchmark performance is very significant (2013 or 2017), the positive ML screening is still able to achieve some outperformance, but the spread with the negative ML screening is somehow lower in years when the market performance is negative or low (in 2014, 2015 and, most recently, 2018).

Simulated strategy levels for the benchmark MSCI World Index, the selection filtered over positive scores and the selection filtered over negative scores

Figure 1

Base level = 100



Simulated strategy levels for the benchmark MSCI World Index, the market capitalisation-weighted selection filtered over positive scores from the ML algorithm (positive ML screening) and the one screened over negative scores (negative ML screening). Data in USD from January 2013 to March 2018.

Source: MSCI, Datastream, Sustainalytics.

Interestingly, the excess return of the sector-matched version is also positive, even if lower in magnitude when compared to the positive ML screening. By neutralising the sector component (because their weights are the same in the benchmark), the outperformance essentially comes from the stock-picking.

Table 11

Year	Bench	Machine learning screening			ESG best-in-class
		Positive	Positive sect matched	Negative	30%
2013	23.95%	0.72%	0.12%	-1.39%	-0.14%
2014	4.97%	3.88%	3.65%	-2.75%	-0.17%
2015	-0.89%	3.79%	1.89%	-5.30%	0.07%
2016	7.40%	-2.14%	-1.91%	3.73%	-0.44%
2017	22.44%	3.82%	0.61%	-2.57%	-0.02%
2018	-1.37%	3.92%	2.24%	-1.63%	-0.24%

Calendar year performance for the MSCI World Index (bench) and the excess returns for the positive, positive sector-matched and negative ML screening as well as for the ESG 30% best-in-class portfolio. Data are shown in USD from January 2013 to March 2018.

Source: MSCI, Datastream, Sustainalytics.

For the negative ML screening, the excess return is always negative except for 2016. Finally, the best-in-class portfolio shows almost systematically small but negative excess returns, except in 2015 when it managed to outperform by 0.07%. Once again, our findings confirm that for very large and diversified universes, the

simple ESG filtering does not bring alpha, although it does not significantly reduce the performance with the best-in-class approach.

Table 12 collects the results of standard factor regressions of the portfolios over the classic Fama and French four-factor model. The positive ML strategy delivers strong, positive and statistically significant alpha, with no specific exposure to size or momentum factors, while it is slightly exposed to the growth factor and a market beta close to one. The same behaviour exists for the sector-matched version, although the resulting alpha is now smaller. R^2 are close to one for all specifications. Evidence from these factor exposures point to the strong picking ability of the ML algorithm, as the outperformance is not coming from unintended factor bets.

	Machine learning screening			ESG best-in-class
	Positive	Positive sect matched	Negative	30%
Ann alpha	2.041%**	1.128%*	-0.142%	-1.088%
Market beta	1.006***	0.989***	1.003***	0.99***
Size	-0.052	0.043	-0.029**	0.095
Value	-0.148***	-0.096**	0.008	0.163**
Momentum	0.042	0.009	-0.002	-0.078
R ²	96.64%	97.77%	99.83%	93.32%

Four-factor model regressions. Returns are sampled at a monthly frequency. We have replaced Kenneth French's Market Factor (MKT) with the MSCI World Index, to force a more intuitive market factor for these strategies. Stars refer to statistical significance: *** = 99% significant, ** 95% significant, * = 90% significant, no star = not significant. Data from January 2013 to March 2018.

Source: MSCI, Datastream, Sustainalytics, Kenneth French's website.

Finally, Figure 2 shows the one-year excess return of the ML strategies against the benchmark. The positive ML screening delivers consistent positive excess return over time, relatively regular, excepted between Q2 and Q4 2016. At the same time, the negative ML screening shows consistently negative and highly volatile excess returns over time. Said otherwise, the ML algorithm achieves its objective of identifying stocks, from their ESG profile, that are indeed able to deliver superior returns.

Simulated one-year rolling excess returns of positive ML screening and the negative ML screening over the benchmark MSCI World Index

Figure 2

Percentage in USD



Data in USD from January 2014 to March 2018.

Source: MSCI, Datastream, Sustainalytics.

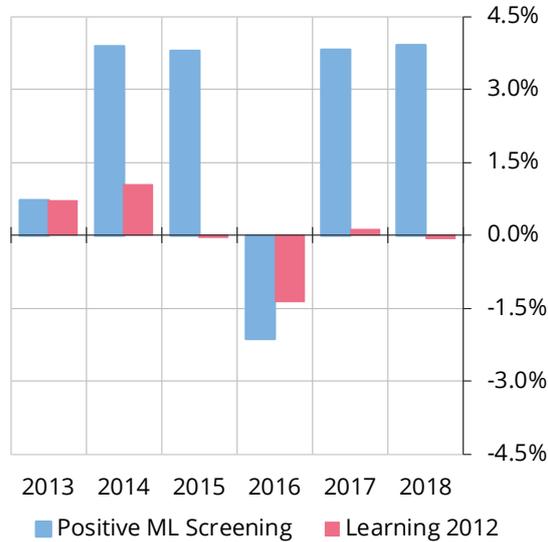
The effects of learning. The machine learning algorithm is initially trained over three years of data and then updated yearly. During these regular updates, the algorithm learns from the new flow of data it can access: it can test its rules to confirm, nuance or remove some of them, and selects new rules linked to statistically significant patterns. This learning process is key in the final performance of the model (and for the positive ML screening portfolio built upon it). To measure this effect, we form four portfolios named LEARNING Y, where Y = 2012, 2013, 2014, 2015 as follows:

- For each year Y, we consider the set of rules related to the learning at the end of the year Y.
- We calculate the scores for all stocks in the universe from the end of year Y to March 2018 with this set of rules.
- LEARNING Y is built as positive ML screening, except that the underlying scores are calculated with the same, not updated set of rules calibrated at the end of year Y.

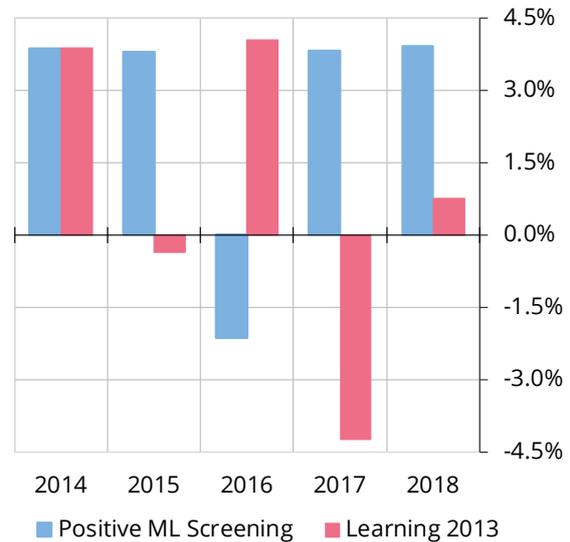
LEARNING Y uses a unique, static set of rules that is never updated (no learning). By construction, the portfolios positive ML screening and LEARNING Y coincide over the period *1 January, Y+1 to 31 December Y+1*, because, over this period, they use the same set of rules (hence the same scores) to screen the investment universe. Figure 3 shows the calendar excess returns of these portfolios together with the positive ML screening portfolio over the benchmark MSCI World Index. Since we only show out-of-sample results, the time frame of each LEARNING Y portfolio is different. In most cases, we see that positive ML screening outperforms the LEARNING Y portfolios after the first year (since they are the same on the first year). Indeed, the excess return for the LEARNING Y portfolios usually shrinks to zero and becomes even negative over time. In other words, the predictive power of the scores vanishes over time, so that it is important to train the algorithm on the new observed data to update the set of rules.

Percentage in USD

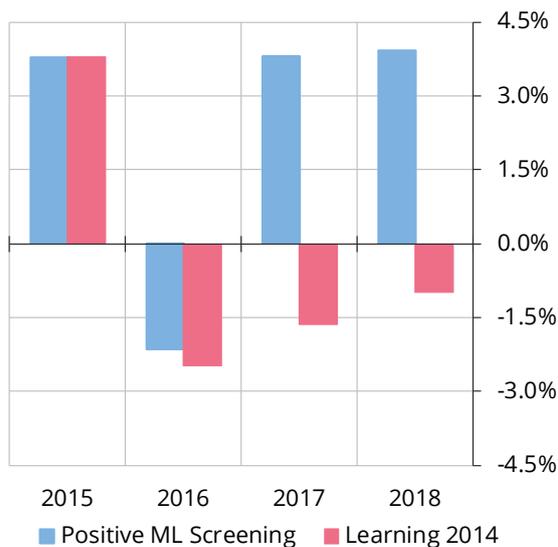
(a) Learning 2012



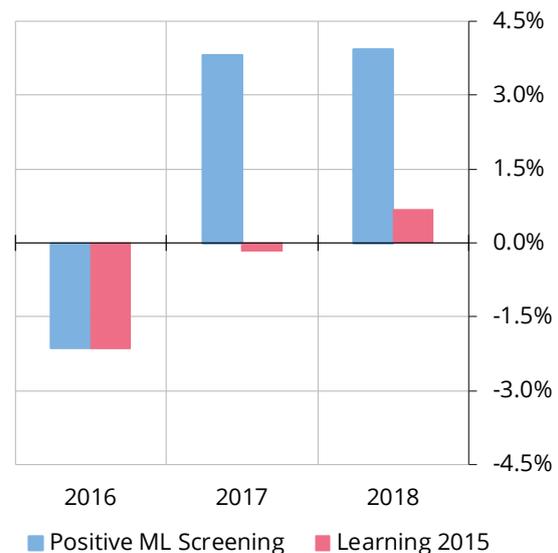
(b) Learning 2013



(c) Learning 2014



(d) Learning 2015



Calendar excess returns of the positive ML screening and the four portfolios LEARNING 2012, LEARNING 2013, LEARNING 2014 and LEARNING 2015 over the MSCI World Index. Data are shown in USD from January 2013 to March 2018.

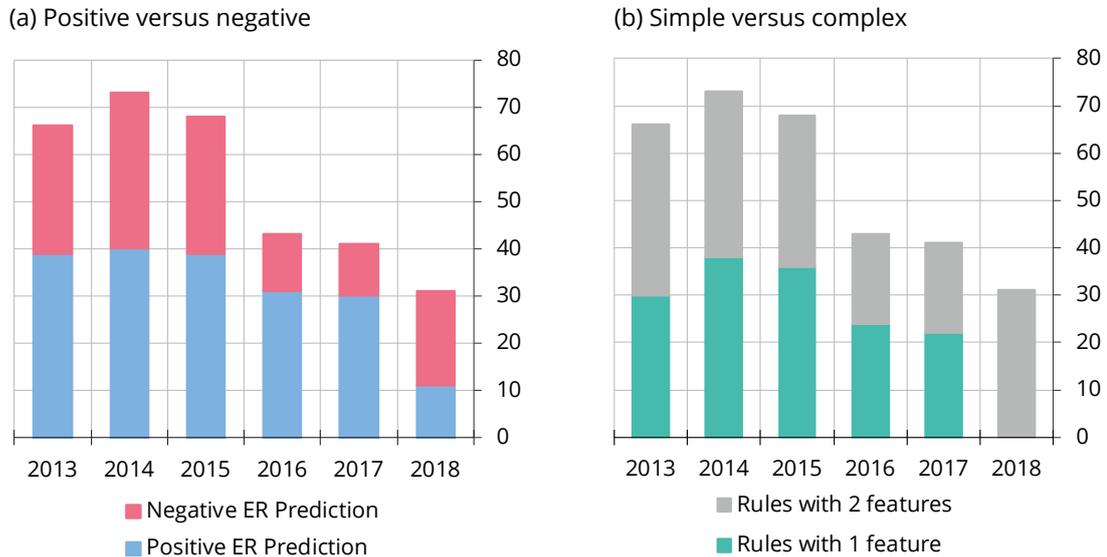
Source: MSCI, Datastream, Sustainalytics.

The number of rules used by the algorithm changes over time: as shown in Panel (a) of Figure 4, this number evolves in the range [31,73] with the split between positive rules (ie rules related to positive predictions of the excess return) and negative ones also changing over time. Interestingly, the number of rules related to negative excess return increased from 12 in 2016 to 20 in the latest 2018 learning. Panel (b) of Figure 4 shows the same number of rules split between simple rules (ie those that only make use of one feature) and complex rules (ie those that use two features, as the examples shown in Table 9). Both Figures 3–4 suggest that, to extract alpha from the ESG

features, one needs to regularly update the algorithm and consider a newly created set of rules to detect patterns between ESG profiles and financial performance.

Number of rules at each update of the algorithm

Figure 4



Number of rules at each update of the algorithm: (a) the split between rules that predict positive or negative excess returns (ER); (b) the split between rules that make use of one feature (simple) or two features (complex).

6. Conclusion

The last few years have seen increasing interest in ESG investing and the integration of socially responsible principles at the portfolio construction level. Managers and investors are asked to complement pure financial objectives with non-traditional financial ones.

Our study brings some new ideas and insights into the way investors could achieve ESG objectives in their investments. The literature on the theme is mixed: initial studies were mostly sceptical of the benefit of ESG integration into the portfolio. Over time the mindset has evolved, and several studies have empirically proved that ESG integration in the portfolio does not lower performance. Most recently, the financial literature has gone one step further and claim that, indeed, ESG integration is a way to extract alpha or, at least, to reduce risks. We recognise the need for serious integration of ESG objectives alongside classic financial ones, and the existence of an economic link between the ESG profile of a company and its financial performance over the long run. Nevertheless, we tend to agree with the pioneers of ESG research, which assert that, at best, ESG integration does not significantly degrade financial performance, especially for large and diversified investment universes.

ESG profiles can impact financial performance in a non-linear way, and the impact can depend on the sector, the country or other specific characteristics of each company. Thus, we designed and implemented a sophisticated machine learning algorithm that identifies patterns between ESG profiles and performance, and is statistically robust across the universe and over time.

The algorithm produces a set of rules, each of which identifies a region in the high-dimensional space of the ESG features, conditionally on which we can make a prediction on the stock's excess return. All the predictions are finally aggregated and transformed into a score taking values in $\{-1, 0, +1\}$, so that in the end we can effectively look at the sign of the excess return rather than its magnitude.

With this algorithm, trained over time to remain updated, we empirically proved that the link between ESG profiles and financial performance exists, but that it can only be accessed with non-linear techniques. Indeed, a simple strategy that selects stocks whose scores are positive significantly outperforms the well known ESG best-in-class approach.

References

- Allouche, J and P Laroche (2005): "A meta-analytic investigation of the relationship between corporate social and financial performance", *Revue de gestion des ressources humaines*, no 57, pp 18.
- Asness, C (2017): "Virtue is its own reward: or, one man's ceiling is another man's floor", *AQR Blog*, <https://www.aqr.com/Insights/Perspectives/Virtue-is-its-Own-Reward-Or-One-Mans-Ceiling-is-Another-Mans-Floor>.
- Aupperle, K, A Carroll and J Hatfield (1985): "An empirical examination of the relationship between corporate social responsibility and profitability", *Academy of Management Journal*, vol 28, no 2, pp 446–63.
- Bragdon, J and J Marlin (1972): "Is pollution profitable?", *Risk Management*, vol 19, no 4, pp 9–18.
- Capelle-Blancard, G and S Monjon (2012): "Trends in the literature on socially responsible investment: looking for the keys under the lamppost", *Business Ethics: A European Review*, vol 21, no 3, pp 239–50.
- Cesa-Bianchi, N and G Lugosi (2006): *Prediction, learning and games*, Cambridge University Press.
- Chong, J and G Phillips (2016): "ESG investing: a simple approach", *The Journal of Wealth Management*, vol 19, no 2, fall, pp 73–88.
- Devaine, M, P Gaillard, Y Goude and G Stoltz (2013): "Forecasting electricity consumption by aggregating specialized experts", *Machine Learning*, vol 90, no 2, pp 231–60.
- Filbeck, G, H Holzhauer and X Zhao (2014): "Using social responsibility ratings to outperform the market: evidence from long-only and active-extension investment strategies", *The Journal of Investing*, vol 23, no 1, spring, pp 79–96.
- Friede, G, T Bush and A Bassen (2015): "ESG and financial performance: aggregated evidence from more than 2000 empirical studies", *Journal of Sustainable Finance & Investment*, vol 5, no 4, pp 210–33.
- Friedman, M (1970): "The social responsibility of business is to increase its profits", *The New York Times Magazine*, 13 September.
- Giese, G, A Ossen and S Bacon (2016): "ESG as a performance factor for smart beta indexes", *The Journal of Index Investing*, vol 7, no 3, winter, pp 7–20.
- Griffin, J and J Mahon (1997): "The corporate social performance and corporate financial performance debate: twenty-five years of incomparable research", *Business & Society*, vol 36, no 1, pp 5–31.
- Humphrey, J and D Tan (2014): "Does it really hurt to be responsible?", *Journal of Business Ethics*, vol 122, no 3, pp 375–86.
- Indrani, D and M Clayman (2015): "The benefits of socially responsible investing: an active manager's perspective", *The Journal of Investing*, vol 24, no 4, winter, pp 49–72.
- Kurtz, L and D Di Bartolomeo (2011): "The long-term performance of a social investment universe", *The Journal of Investing*, vol 20, no 3, fall, pp 95–102.

Margolis, J, H Elfenbein and J Walsh (2009): "Does it pay to be good and does it matter? A meta-analysis of the relationship between corporate social and financial performance", https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1866371.

Peiris, D and J Evans (2010): "The relationship between environmental social governance factors and US stock performance", *The Journal of Investing*, vol 19, no 3, fall, pp 104–12.

Nemirovski, A (2000): "Topics in nonparametric", *Ecole d'Eté de Probabilités de Saint-Flour*, vol 28, pp 85.

Orlitzky, M, F Schmidt, S and Rynes (2003): "Corporate social and financial performance: a meta-analysis", *Organization Studies*, vol 24, no 3, pp 403–41.

Revelli, C and J Viviani (2015): "Financial performance of socially responsible investing (SRI): What have we learned? A meta-analysis", *Business Ethics: A European Review*, vol 24, no 2, pp 158–85.

Shiller, R (2013): "Capitalism and financial innovation", *Financial Analysts Journal*, vol 69, no 1.

Stoltz, G (2010): "Agrégation séquentielle de prédicteurs: méthodologie générale et applications à la prévision de la qualité de l'air et celle de la consommation électrique," *Journal de la Société Française de Statistique*, vol 151, no 2, pp 66–106.

Tsybakov, A (2003): "Optimal rates of aggregation", in B Schölkopf and M Warmuth (eds) *Learning theory and kernel machines*, pp 303–13.

Van Beurden, P and T Gössling (2008): "The worth of values – A literature review on the relation between corporate social and financial performance", *Journal of Business Ethics*, vol 82, no 2, pp 407–24.

Van Duuren, E, A Plantinga and B Scholtens (2016): "ESG integration and the investment management process: fundamental investing reinvented," *Journal of Business Ethics*, vol 138, no 3, pp 525–33.

Wu, M L (2006): "Corporate social performance, corporate financial performance, and firm size: a meta-analysis", *Journal of American Academy of Business*, vol 8, no 1, pp 163–71.

Zoltan, N, A Kassam and L E Lee (2016): "Can ESG add alpha? An analysis of ESG tilt and momentum strategies", *The Journal of Investing*, vol 25, no 2, summer, pp 113-24.

BlackRock vs Norway Fund at shareholder meetings: institutional investors' votes on corporate externalities¹

Marie Brière,^a Sébastien Pouget^b and Loredana Ureche-Rangau^c

Abstract

Do institutional investors engage with companies on corporate externalities such as greenhouse gas emissions? And if so, why? We study voting at shareholder meetings by two emblematic global investors: BlackRock, a major asset manager, and the Norway Government Pension Fund Global (the Norway Fund), a responsible sovereign wealth fund. Our data cover 2014 and include the two institutions' votes on 35,382 resolutions at 2,796 corporations worldwide. Both of these so-called universal owners oppose management significantly more often on externality than on financial issues. The Norway Fund is more active on shareholder resolutions concerning externalities related to environmental and social issues than on governance issues. The difference between the two investors' voting behaviour is larger when we focus on resolutions related to greenhouse gas emissions, a clearly identified externality. Overall, universal ownership (see eg Monks and Minow (1995)) and, more importantly, delegated philanthropy (see eg Benabou and Tirole (2010)) both appear to provide incentives for institutional investors to combat negative externalities generated by firms.

^a Amundi, 90 Bd Pasteur, 75015 Paris, France, Paris Dauphine University, Place du Maréchal de Lattre de Tassigny, 75116 Paris, France; Université Libre de Bruxelles, Solvay Brussels School of Economics and Management, Centre Emile Berheim, Av FD Roosevelt, 50, CP 145/1, 1050 Brussels, Belgium, .

^b Toulouse School of Economics (University of Toulouse 1 Capitole, TSM), 21 allée de Brienne, 31000 Toulouse, spouget@univ-tlse1.fr.

^c CRISEA, University of Picardie Jules Verne, 10 Placette Lafleur, BP 2716, 80027 Amiens Cedex 1, France, loredana.ureche@u-picardie.fr.

¹ The authors thank Andrea Attar, Bruno Biais, Narayan Bulusu, Catherine Casamatta, Fany Declerck, Tom Fearnley, Jonas Jolle, Bert Scholtens, Simone Sepe, Alexis Wegerich, and the participants to the Seventh Public Investors Conference for useful comments and suggestions. Financial support from the research chair on sustainable finance and responsible investing (Chaire FDIR) and from the Amundi research chair on asset management is gratefully acknowledged.

1. Introduction

This paper studies whether and why institutional investors engage companies to reduce the negative externalities they exert on society. As indicated, for example, by Laffont (1987), an externality is the effect produced by an economic activity on parties that are not involved in this activity. Externalities constitute a major source of market failure since market equilibria only reflect private effects that are perceived by the parties involved in the activity, but not overall societal effects. In a report based on research by Trucost, a leading consultancy firm in extra-financial analysis, Mattison et al (2011) estimate that, in 2008, the largest 3,000 publicly listed companies worldwide generated more than US\$ 2.15 trillion or 7% of their combined revenues as environmental externalities such as climate change. This figure, which is already very significant, does not consider companies' social externalities such as consumer safety issues and human rights violations.

To study institutional investors' engagement to reduce companies' negative externalities, we focus on votes at shareholder meetings on resolutions related to both environmental and social (ES) issues. Such a focus is useful because it provides us with a large amount of data on one type of engagement, ie shareholder voting, on societal issues. To be even more precise in terms of identification, we also restrict our attention to greenhouse gas emissions, a clear example of an externality produced by companies.

To understand what motivation may induce investors to care about externalities generated by companies, we focus on the Norway Government Pension Fund Global (the Norway Fund) and BlackRock, two emblematic institutional investors. These two investors had assets under management of more than \$1 trillion and \$5 trillion, respectively, in 2017. They both have a large, global and well diversified equity portfolio. In this sense, both investors are universal owners (see eg Monks and Minow (1995)). The Norway Fund has also a delegated philanthropic mission (see eg Benabou and Tirole (2010)) as it is monitored by the parliament of Norway and a Council on Ethics.² Given their size, the two investors are likely to have a significant influence on corporate behaviour across the world.

Separation between ownership and control is one of the fundamental characteristics of modern companies (Berle and Means (1932)). This separation opens the room for potential conflicts of interests between shareholders and corporate executives (Jensen and Meckling (1976)): managers may not always favour the strategies that are best for shareholders. These potential conflicts call for an active involvement of shareholders in the governance of corporations. This explains why we are interested in institutional investors' engagement.

As described by Bebchuk et al (2017), institutional investors play a central role in today's corporate governance landscape. To mitigate the negative effects of the conflict between shareholders and executives, institutional investors can induce

² "...the Council on Ethics for the Government Pension Fund Global (GPF) is an independent body that makes recommendations to Norway's central bank, Norges bank, to exclude companies from the GPF or place them under observation. The council assesses a company's operations on the basis of guidelines determined by the Norwegian Ministry of Finance. The guidelines contain both product-based exclusion criterion, such as the production of tobacco or certain types of weapons, and conduct-based exclusion criterion, such as gross corruption, human rights violations and environmental damage. The Council has five members and a secretariat with a staff of eight" (Council on Ethics (2016), p7).

executives to follow their guidance by engaging companies, such as (i) holding discussions with executive managers and board members (see eg Dimson et al (2015), Barko et al (2017)), (ii) filing shareholder proposals (see eg Gillian and Starks (2000), Cziraki et al (2010), Renneboog and Szilagyi (2011)) and (iii) voting during shareholder general meetings (see eg Cunat et al (2012), Flammer (2015), Bach and Metzger (2017)).

Two basic arguments justify institutional investors in being actively engaged on externality issues. The first argument rests on the universal owner logic (see eg Monks and Minow (1995), Hawley and Williams (2000), Dimson et al (2013), Azar (2017)). Large institutional investors own shares in virtually all listed companies and have a long horizon. As universal owners, they might engage firms to mitigate the negative externalities imposed on other firms held in their portfolios, to avoid deteriorating their overall value. For example, they may want to consider the negative economic impact that the greenhouse gas (GHG) emissions of a firm might have on other companies' businesses through water, food, health or migration issues.

The universal owner logic is well summarised in Mattison et al (2011): "For a diversified investor, environmental costs are unavoidable as they come back into the portfolio as insurance premiums, taxes, inflated input prices and the physical cost associated with disasters. One company's externalities can damage the profitability of other portfolio companies, adversely affecting other investments, and hence overall market return." Larry Fink, head of BlackRock, indicates that passive investors, as universal owners, have strong incentives to engage companies: "In our index-oriented accounts, we can't sell those stocks even if they are terrible companies. As an indexer, our only action is our voice and so we are taking a more active dialogue with our companies and are imposing more of what we think is correct" (Authers (2015)).

Universal owners can also engage companies to improve the level of coordination among their ES policies, which can be beneficial for all companies' financial value. For example, Benabou and Tirole (2016) show that coordinated policies on managerial compensation issues enable firms to avoid the harmful effects of a bonus culture.

The situation is very different for corporate executives who, in general, own concentrated stakes in their companies, either because most of their capital is in the form of firm-specific human capital or because their incentive plans require them to do so. These different exposure profiles generate conflicts of interests: executives are less likely to be willing to internalise the effects that their companies have on the payoffs and value of other companies.

A second argument that calls for institutional investors to be active in engagement on externality issues is related to delegated philanthropy logic (Benabou and Tirole (2010)). Institutional investors such as pension funds, mutual funds and sovereign funds invest on behalf of clients or citizens who may have preferences regarding externalities that differ from the ones of companies' managers. Institutional investors might thus want to promote these clients' and citizens' values and preferences and induce management to choose the appropriate course of action. One can, for example, think that the level of global risk induced by a firm related to climate change or nuclear energy might not be valued in the same manner by corporate managers and by institutional investors who represent clients or citizens. Investors may thus want to communicate their preferred level of precaution to corporate executives. This can only be achieved via engagement. One important reason why

institutional investors may endorse the delegated philanthropy logic is because they care about their reputation among clients or citizens.

As shown by Morgan and Tumlinson (2012), such engagement by institutional investors on externality issues is socially desirable because (i) companies' actions are less subject to the free-rider problem than that of individual shareholders when deciding to fight against these externalities, and (ii) it makes companies' production decisions more efficient from a social point of view and increases the welfare of shareholders who care about these externalities.

In the delegated philanthropy logic, conflicts of interest may emerge when corporate executives and shareholders have different values and preferences towards corporate externalities. Shareholders will find it important to communicate their values and preferences towards externalities to executives to induce firms to adopt their preferred behaviour.

Universal ownership is the most prevalent reason provided by institutional investors to rationalise their responsible investment and engagement policies. One reason is that the universal ownership logic focuses only on financial returns and is thus consistent with a narrowly defined notion of fiduciary duty. There are, however, several impediments to this logic. On the one hand, the externalities should be correctly evaluated and should be material for companies' profits. On the other hand, the materiality of externalities should not occur too far into the future in order to significantly affect asset valuations. Delegated philanthropy does not suffer from these impediments, but its strength can be attenuated by the difficulty of finding a consensus among clients and citizens regarding the externalities that investors should focus and actively engage companies on.

We aim to put these two basic arguments – ie universal ownership and delegated philanthropy – through an empirical test.

The paper is organised as follows. First, we discuss the related literature, then we present our methodology, data and variables before ending with our empirical analysis and a discussion of our results.

2. Related literature

Several papers have studied how voting at shareholder meetings can alter corporate behaviour. Cunat et al (2012) show that close votes in favour of changes in governance trigger an improvement in the valuation of market capitalisation. Likewise, Flammer (2015) and Flammer and Bansal (2017) show that close votes on ES issues and on long-term executive compensation plans, respectively, are associated with an increase in firms' stock market valuation. Bauer et al (2010) show that firms in less competitive industries are more likely to be targeted by shareholder resolutions. Bach and Metzger (2017) find that shareholder support for a proposal affects firm value because, even if votes are non-binding as is the case in the United States, failure to comply with a majority vote may trigger executive turnover. We complement this literature by analysing in more detail the voting policies of institutional investors and their determinants.

Other papers have studied how behind-the-scenes engagement by investors may affect corporate behaviour and performance, eg Smith (1996) and Becht et al

(2009) tackle governance issues and Dimson et al (2015) and Barko et al (2017) ES issues. These papers focus on the engagement of a given institutional investor. They find that private engagement is effective at triggering changes in targeted companies and that engagement, in general, increases firms' value. We complement this literature by focusing on voting instead of private engagement and by studying the voting strategies of investors with different motivations, standards and responsibilities at a common subset of resolutions.

Azar et al (2017) show that firms in the US airline industry that are held by common institutional investors are less likely to aggressively compete on the same routes. Keswani et al (2016) study the voting behaviour of financial firms at their competitors' general assembly meetings. They find that they are more likely to favour management reducing directors' efficacy and firm valuation. These empirical studies document the hidden cost of universal ownership. The present study aims at documenting a potential positive side, especially the fact that universal owners might have an incentive to internalise part of the corporate externalities, as argued for example by Mattison et al (2011).

Fichtner et al (2017) offer a very interesting description of the voting policies of the three largest passive asset management firms, BlackRock, Vanguard and State Street. They observe that these firms implement a coordinated voting policy across their different funds and that they, in general, vote with management. We complement this analysis by focusing on votes on externality issues, comparing them with votes on other issues, and providing an empirical test of the various reasons why institutional investors may pressure companies to take actions against negative externalities.

3. Methodology

We make an empirical study of institutional investors' votes on externality issues at shareholder meetings. This focus provides us with a relatively large amount of data and allows us to clearly identify conflicts between management and shareholders.³ When management opposes efforts to fight negative externalities, some shareholders may fill in resolutions to be voted upon at shareholder meetings in an attempt to impose a different policy on management.⁴ In this case, it is interesting to investigate what voting stance large institutional investors adopt in order to find out whether they support the idea of companies making such efforts to mitigate negative externalities.

In this paper, we focus on two emblematic global investors: BlackRock and the Norway Fund. BlackRock is an asset management firm with over \$5 trillion dollars under management, of which the total equity portfolio amounts to \$2.6 trillion. According to Fichtner et al (2017), BlackRock is the broadest global blockholder in listed corporations around the world: 3,648 holdings above 3%, 2,632 above 5% and 375 above 10%. In the United States, BlackRock has about 2,000 holdings of 5% among the 3,900 publicly listed US companies. Within its numerous funds, BlackRock

³ One drawback is that we are unable to observe behind-the-doors discussions (see McCahery et al (2016), Dimson et al (2015), Barko et al (2017)).

⁴ If management wanted to promote efforts against negative externalities, it would not wait for a resolution to be filed at the annual shareholder meeting before implementing the appropriate policy.

pursues both passive asset management, through index and exchange-traded funds, and active management. The company's corporate governance team includes 31 people that vote at more than 15,000 shareholder meetings and more than 130,000 proposals every year. BlackRock follows a centralised voting policy.

The Norway Fund is a sovereign wealth fund with over \$1 trillion of assets. It holds equity stakes in about 9,000 companies worldwide, with a total equity portfolio of more than \$500 billion. The average proportion of shares in listed corporations held by the Norway Fund is about 1%. The fund's corporate governance team includes around 12 people who vote on more than 11,000 resolutions at general meetings every year. In 2014, the two investors' holdings seem highly correlated, both in terms of a firm's capitalisation (an 87% correlation coefficient) and the weights of companies in investors' portfolios (a 95% correlation).

Given the amount of managed assets invested in global equity, both BlackRock and the Norway Fund may be characterised as universal owners: they hold a significant equity stake in almost all major publicly listed firms worldwide. However, they differ across several dimensions. On the one hand, BlackRock has been a listed corporation since 2009 and is therefore run by a board that has a fiduciary duty to represent its own shareholders. Among these shareholders, the major ones, with holdings above 3%, are PNC Bank, Norges Bank Investment Management, The Vanguard Group, Wellington Management, Capital Research & Management, State Street Global Advisors Fund Management and BlackRock Fund Advisors. We thus consider BlackRock as the archetype of a standard, well diversified investor. In its Global Corporate Governance and Engagement Principles, BlackRock states that "the trigger for engagement on a particular ES concern is [its] assessment that there is potential for material economic ramifications for shareholders". This is clearly in line with the universal ownership principle described above.⁵

On the other hand, the Norway Fund is a sovereign wealth fund that invests Norway's petroleum revenues to provide steady resources for the country over the long term. As stated by Chambers et al (2012), its goal is "to serve as a long-term savings vehicle which seeks to secure the income from a non-renewable resource by diversifying into a broad portfolio of international securities." The Norway Fund is monitored by the Ministry of Finance, which is itself supervised by the Norwegian parliament. Because of this fiduciary duty to the representatives of the Norwegian people, the Norway Fund is recognised as a leader in the responsible investment community (see Chambers et al (2012)). The fund's commitment to responsible investing is manifested in a Council for Ethics, in charge of evaluating whether the investment policy is consistent with the ethical guidelines adopted by the Ministry of Finance. As indicated in its 2016 annual report, the Council on Ethics' objective when engaging with a company in which the Norway Fund invests is to "gather information to provide a basis for assessing the risk that the company may be contributing to the violation of ethical norms, either now or in the future" (see Council on Ethics (2016)). The Norway Fund is part of "the 25 most responsible asset allocators" list that

⁵ It is possible that BlackRock also pursues a delegated philanthropy approach: "To prosper over time, every company must not only deliver financial performance, but also show how it makes a positive contribution to society. Companies must benefit all of their stakeholders, including shareholders, employees, customers, and the communities in which they operate" (see BlackRock, *A fundamental reshaping of finance*, www.blackrock.com/corporate/en-no/investor-relations/larry-fink-ceo-letter). However, given its size and the diversity of values of its clients, BlackRock might have difficulties in clearly identifying the most important issues.

distinguishes the most responsible sovereign wealth funds and government pension funds across a universe of more than 200 funds worldwide (The Bretton Wood II Leaders List (2017)). Thus, we consider the Norway Fund as the archetypal responsible, well diversified investor.⁶

By comparing the voting behaviours of BlackRock and the Norway Fund at general meetings, we can identify whether universal ownership alone is sufficient to encourage institutional investors to promote corporate action against negative externalities or whether delegated philanthropy is also necessary. We focus on 2014, the first year for which we have detailed information on voting by these two investors. To do our test, we have collected and classified voting data for the two investors on the same resolutions.

Our analysis focuses on understanding investors' opposition to management. At shareholder meetings, management and shareholders may fill in resolutions. Externality resolutions are proposed by shareholders and pertain to ES issues. When interpreting our results, we pay attention to the ultimate meaning of votes: opposing management on a shareholder proposal means voting for the proposition to pass. This is because management almost always opposes shareholder resolutions. Thus, if an investor opposes management on an externality-related shareholder resolution, it means that this investor is encouraging the effort against the negative externality.

We compare investors' votes on externality issues with those on a variety of other issues, notably management proposals on financial and governance matters, and shareholder proposals on governance. This enables us to clearly identify opposition due to externalities rather than to other characteristics of the proposed policies. Moreover, we want to single out the impact of preferences for negative externality mitigation from other effects. For that purpose, our analysis controls for various factors that can explain disagreement with management or among investors. Agency problems (see eg Agrawal and Knoeber (1996), Hong et al (2012), Cheng et al (2013)) can be one reason for investors' disagreement and we thus include a dummy indicating that a resolution has been filed by a shareholder on a governance issue. Differences of opinion can be another reason (see eg Chen et al (2002), Hong and Stein (2003), Boot et al (2006 and 2008)), and we include the dispersion in analysts' forecasts (see eg Diether et al (2002)) as a control.

4. Data and variables

Our data include detailed information for 35,382 resolutions, including 326 on ES practices, voted upon by both BlackRock and the Norway Fund in 2014 on a sample of 2,796 corporations across the world.⁷ We collected this data from BlackRock's SEC filings and the Norway Fund's website.⁸ We obtained firm characteristics from FactSet and firms' environmental, social and governance (ESG) ratings from MSCI. For

⁶ Most sovereign wealth funds have delegated philanthropy objectives that may not, however, always encourage a focus on corporate social responsibility. Norway Fund's focus is clearly stated: "The long-term return depends on sustainable development in economic, environmental and social terms" (see Norges Bank Investment Management, *Strategy 2014–2016*, p12).

⁷ We focus on the equity shares owned and voted by BlackRock (active mutual funds and passive funds such as iShares ETFs).

⁸ www.nbim.no/en/responsibility/our-voting-records/.

additional analyses that require data on the two institutional investors' holdings, we also use a smaller sample based on information retrieved from the Form 13F filings of the SEC's Electronic Data Gathering, Analysis and Retrieval (EDGAR) system, ie a total of 6,037 resolutions, including 110 on ES aspects. Much of the data collection effort revolved around manually classifying data into various categories (financial, governance, environment etc) and subcategories (climate change and GHG emissions, hydraulic fracturing etc).

The period under study was chosen because voting instruction data from Norges Bank Investment Management is available online from 1 July 2013. Using this data set, we were able to collect the management recommendations for each of the resolutions submitted to a vote, as well as the content of the resolution and the vote per se.⁹

4.1. Description of resolutions

Following the proxy voting guidelines issued by Institutional Shareholder Services (ISS),¹⁰ we manually classified the resolutions into five major areas: environmental (E), social (S), governance (G), financial, and others.¹¹ E, S and G resolutions include several themes that include different issues. Table 1 shows summary statistics on the entire data set collected on BlackRock and Norway Fund votes. Out of 35,382 resolutions voted upon by the two investors, 69 were on environmental issues, mainly climate change and reporting of sustainability policies; 257 were on social issues, dealing mainly with firms' charitable contributions, political lobbying and donations, and human right issues; 28,396 were on governance issues, mainly those related to board structure, compensation and audit practices.

4.2. Summary of votes

Table 1 reports the summary statistics on opposition to resolutions by issue of BlackRock and the Norway Fund. It shows that the rate of opposition to management is different for BlackRock and the Norway Fund. BlackRock opposes management on 3% of resolutions, compared with 8% for the Norway Fund. The opposition rates are similar to the general statistics for financial and governance issues. However, opposition rates are different for ES. BlackRock rarely opposes management on these issues, while the Norway Fund opposes management on 101 out of 326 resolutions (31%).

The Norway Fund is particularly active on climate change and GHG emissions, with an opposition rate to management of 83% and sustainability reporting at 50%. On social issues, the Norway Fund's degree of opposition amounts to 75% on board diversity issues, 83% on sexual orientation and 65% on political contributions. All the environmental resolutions and most of the social resolutions are proposed by shareholders. Within the social area, management-sponsored resolutions account for 140 out of a total of 257 (101 on political contributions and 39 on charitable

⁹ The data set collected online was manually cross-checked. We thank Thierry Martial Kengne for excellent research assistance.

¹⁰ www.issgovernance.com/file/policy/2016-us-summary-voting-guidelines-dec-2015.pdf.

¹¹ The category called "others" refers to matters that did not fall into one of the four other areas, eg "open meeting", "close meeting", "amend articles" etc.

contributions). Shareholder resolutions on governance are rare (1% of such resolutions are proposed by shareholders) but they show an interesting divergence between the two investors: the Norway Fund opposes management on these resolutions 36% of the time, BlackRock only 12%.

Voted resolutions and rate of opposition to management

Table 1

	Total number of voted	Rate of opposition to the management		Rate of opposition to the management by sponsor of the resolution			
		BlackRock	Norway Fund	BlackRock		Norway Fund	
				Management	Shareholder	Management	Shareholder
E	69	4%	49%	-	4%	-	49%
Animal welfare	2	0%	50%	-	0%	-	50%
Animal testing	1	0%	0%	-	0%	-	0%
Animal welfare policies	1	0%	100%	-	0%	-	100%
Climate	24	4%	83%	-	4%	-	83%
Climate change and GHG emissions	24	4%	83%	-	4%	-	83%
Environment and sustainability	34	0%	23%	-	0%	-	23%
Hydraulic fracturing	3	0%	67%	-	0%	-	67%
Nuclear safety	15	0%	0%	-	0%	-	0%
Sustainability reporting	16	0%	50%	-	0%	-	50%
Others	9	22%	33%	-	22%	-	33%
S	257	8%	26%	6%	9%	7%	49%
Consumer issues	10	10%	10%	-	10%	-	10%
Genetically modified ingredients	8	13%	13%	-	13%	-	13%
Other consumer responsibility	2	0%	0%	-	0%	-	0%
Diversity	11	9%	73%	-	9%	-	73%
Board diversity	4	25%	75%	-	25%	-	75%
Discrimination	1	0%	0%	-	0%	-	0%
Sexual orientation	6	0%	83%	-	0%	-	83%
General corporate issues	40	22%	22%	23%	0%	23%	0%
Charitable contributions	40	22%	22%	23%	0%	23%	0%
Human rights	20	10%	35%	-	10%	-	35%
Human rights proposals	20	10%	35%	-	10%	-	35%
Political activities	176	4%	24%	0%	9%	1%	55%
Lobbying	29	10%	38%	-	10%	-	38%
Political contributions	147	3%	21%	0%	9%	1%	65%
G	28,396	3%	8%	3%	12%	8%	36%
Audit practices and risk management	3,113	2%	7%	2%	0%	7%	8%
Audit practices	3,111	2%	7%	2%	0%	7%	9%
Risk management	2	0%	0%	-	0%	-	0%
Board accountability and responsive	18	0%	11%	0%	0%	0%	20%
Ability to remove directors	13	0%	0%	0%	0%	0%	0%
Tax transparency	5	0%	40%	0%	0%	0%	100%
Board independence	51	18%	88%	100%	16%	100%	88%
Competitive activities of directors	1	100%	100%	100%	-	100%	-
Independent chairman and directors	50	16%	88%	-	16%	-	88%
Board structure	20,557	2%	7%	2%	5%	7%	7%
Appointment	20,143	2%	7%	2%	0%	7%	1%
Board composition	167	1%	2%	1%	0%	1%	38%
Others board related proposals	86	9%	10%	1%	47%	3%	47%
Related-party transaction	161	1%	6%	1%	-	6%	-
Compensation/Remuneration	4,462	5%	11%	5%	2%	11%	47%
Employee compensation	1,606	6%	11%	6%	0%	11%	15%
Executive compensation	2,856	4%	12%	4%	3%	10%	53%
Shareholder rights	195	27%	33%	21%	35%	20%	49%
Call special meeting	20	20%	40%	0%	33%	0%	67%
Proxy access right	22	36%	41%	0%	47%	0%	53%
Takeover defenses	87	33%	25%	29%	53%	26%	20%
Voting formalities	66	18%	38%	8%	24%	12%	54%
Financial	5,716	3%	5%	3%	6%	5%	26%
Others	944	4%	6%	4%	1%	6%	7%
Total	35,382	3%	8%	3%	9%	7%	34%

This table summarises the number of voted resolutions by the two investors in 2014, within each area, theme and issue (Column 1). Columns 2 and 3 provide the percentage of opposition to management recommendation within each area, theme and issue (independent of the sponsor of the resolution). Columns 4, 5, 6 and 7 report the rate of opposition to management within each area, theme and issue, depending on the sponsor of each resolution (management and shareholder) of BlackRock and the Norway Fund respectively.

Source: Authors' calculations.

4.3. Variables

Dependent variables

The variables we seek to explain are the two investors' opposition to management recommendations regarding the resolutions submitted to a vote. We thus define the following six dummy variables:

- *BR or NF oppose* equals one if either or both investors oppose management recommendation, and zero elsewhere;
- *BR opposes* equals one if BlackRock opposes management recommendation, and zero elsewhere;
- *NF opposes* equals one if the Norway Fund opposes management recommendation, and zero elsewhere.

The main statistics for these dummies are presented in Table 2, Panel A. Opposition to management concerns 9% of resolutions on average, due mainly to the Norway Fund's voting policy.

Summary statistics					Table 2
Variable	N	Mean	Std Dev	Min	Max
Panel A: Measures of agreement					
BR or NF oppose	35,382	0.087	0.282		
BR opposes	35,382	0.027	0.163		
NF opposes	35,382	0.078	0.268		
Panel B: Resolution characteristics					
Shareholder proposal	35,382	0.023	0.150		
Resolution ES	35,382	0.009	0.096		
Resolution G	35,382	0.803	0.398		
Resolution climate	35,382	0.001	0.026		
Resolution ES non climate	35,382	0.009	0.092		
Panel C: Country and firm ESG ratings					
Country Rating ES	35,382	0.619	0.058	0.360	0.690
Country Rating G	35,382	0.783	0.122	0.140	0.950
Company Rating ES	35,382	4.913	1.486	0.500	9.950
Company Rating G	35,382	6.567	2.683	0.000	10.000
Panel D: Firm characteristics					
Market cap	35,382	14923.470	32209.350	47.700	439000.000
ROA	35,382	4.575	8.343	-99.5	189.000
Price-to-book	35,382	3.837	32.693	0.192	1540.000
Sales growth	35,382	0.218	4.702	-1.04	177.000
Asset turnover	35,382	0.790	0.698	0.000	9.390
Volatility	35,382	38.551	13.022	14.038	202.924
Analyst dispersion	35,382	0.130	1.016	-54.22	27.678

This table provides summary statistics for the 35,382 common resolutions the Norway Fund and BlackRock voted upon in 2014. Panel A refers to the disagreement measures, Panel B to the characteristics of the resolutions, Panel C to the extra-financial performance measures for firms and countries, Panel D to firms' financial characteristics. All variables are defined in the Data and Variables section.

Source: Authors' calculations.

Explanatory and control variables

Resolution characteristics

We define several dummy variables to set forth specific dimensions of voted resolutions, namely:

- *Shareholder proposal equals one if the resolution is sponsored by shareholders and zero elsewhere;*
- *Resolution ES equals one if the resolution is either related to E or S issues and zero elsewhere;*
- *Resolution G equals one if the resolution is related to G issues and zero elsewhere;*
- *Resolution climate equals one if the resolution is related to climate issues and zero elsewhere;*
- *Resolution ES non-climate equals one if the resolution is related to all the environmental and social issues except for climate and zero elsewhere.*

Table 2, Panel B summarises the main statistics for these dummies. Two main observations emerge. First, on average, only 2% of the resolutions submitted to a vote are sponsored by shareholders. Second, most resolutions are related to governance (80% on average).

ESG characteristics

Different variables are used to capture the ESG performance of firms and their home countries. To assess firms' ESG performance, two variables are constructed/collected:

- *Company Rating ES equals the average between the E and S scores provided by the MSCI ESG STATS database for corporate social responsibility.¹² We aggregated the E and S scores to obtain a single measure of the firms' performance on societal externalities.*
- *Company Rating G is collected from MSCI ESG STATS.*

The summary statistics in Table 2, Panel C, show that the firms under study perform better on governance issues (with an average score higher than 6) than on ES topics (average score below 5).

We use different proxies to measure the ESG performance of countries where the firms in our sample are domiciled and obtain this data from several sources. We thus construct the following variables:

- *Country Rating ES equals the average between the E rating and S rating for each country, where:*
 - the E rating is the average of five variables that proxy key environmental issues: GHG per unit of GDP; air quality and health; environmental policy stringency index (all from OECD statistics¹³); global per capita CO2 emissions from fossil fuel use and cement production (from the Emissions Database for Global Atmospheric Research¹⁴) and the Environmental Performance Index (EPI) produced by Yale Center for Environmental Law and Policy (YCELP) and the Center for International Earth Science Information Network

¹² See MSCI ESG Research, IVA Methodology, 2014.

¹³ stats.oecd.org.

¹⁴ EDGARv4.3, European Commission, Joint Research Centre (JRC)/PBL Netherlands Environmental Assessment Agency.

- (CIESIN) at Columbia University.¹⁵ Each variable is normalised¹⁶ into an index between zero and one for aggregation purposes;
- the S rating is the average of two variables that proxy key social issues: the Human Development Index and the Gender Inequality Index, both produced by the annual Human Development Reports Office of the United Nations Development Programme.
 - *Country Rating G* equals the average of six index-transformed variables: voice and accountability; government effectiveness; regulatory quality; rule of law; control of corruption; and political stability and absence of violence/terrorism, all collected from the World Bank's Worldwide Governance Indicators database.¹⁷

The statistics reported in Table 2, Panel C, indicate that, on average, the countries in which our sample companies are domiciled are relatively well rated regarding ESG criteria.

Firm financial characteristics

Data for firm characteristics are obtained from FactSet. As illustrated in Table 2, Panel D, these characteristics include:

- *market capitalisation on 31 December 2013, in thousands of dollars;*
- *return on assets on 31 December 2013;*
- *price-to-book ratio on 31 December 2013;*
- *annual sales growth rate on 31 December 2013;*
- *asset turnover ratio on 31 December 2013;*
- *volatility, proxied by the annualised standard deviation of daily stock returns between 2009 and 2013; and*
- *analyst dispersion, measured as the standard deviation of earnings-per-share forecasts scaled by absolute mean earnings forecasts, following Diether et al (2002) and Johnson (2004); we consider, for each firm, analysts' forecasts 6 months before the general meeting date.*

Each company is also associated with its industry in 10 commonly defined sectors:

1. Financials
2. Materials
3. Industrials
4. Consumer discretionary
5. Health care
6. Technology
7. Energy
8. Communications
9. Consumer staples
10. Utilities

All continuous control variables (market cap, ROA, price-to-book, sales growth, asset turnover and analyst dispersion) are normalised.

¹⁵ <http://epi.yale.edu> and <http://sedac.ciesin.columbia.edu/data/set/epi-environmental-performance-index-2014>.

¹⁶ Index = (variable – min)/(max-min). An index closer to 1 indicates a better performance in the area under study.

¹⁷ Details can be found in Kaufmann et al (2010).

Finally, on the reduced sample comprising firms with SEC 13F filings, we also include different measures of holdings to proxy for the financial stake of each of the two investors in each firm. As depicted in Table 6, Panel E, we construct:

- *Weight in BR portfolio (respectively, NF) as the weight that the investment in a given company represents within the entire portfolio of BlackRock, respectively the Norway Fund;*
- *Holding BR (% of capitalisation), respectively NF, as the amount invested in the firm by BlackRock, respectively the Norway Fund, divided by the market cap of the firm, as reported by the 13F filings on 31 December 2013;*
- *Weight in portfolio (average BR NF) as the average between the weights of the two investors; and*
- *Holding (average BR NF) as the average between the two investors' holdings.*

5. Empirical results

We present our main results regarding investors' opposition to management on resolutions related to externalities¹⁸ for the entire sample with country and sector fixed effects, then we provide robustness results from additional regressions without country fixed effects and with bivariate regressions, and regressions that control for the holdings of BlackRock and the Norway Fund and, finally, results on climate change issues that clearly involve externalities.

5.1. Main analyses

Our basic specification studies the two investors' opposition to managers on externality issues, compared with their opposition on other issues related to finance and governance.

The results from our basic specification are reported in Table 3. We regress the likelihood of opposition onto the fact that the resolution relates to ES issues and onto various control variables. Column (1) shows that at least one of the two investors is more likely to oppose corporate management on ES resolutions submitted by shareholders. The coefficient on these issues, 1.867, is significantly different from 0 and from the coefficient on governance issues raised by shareholders, 1.594 (p-value=0.08).

The analysis of marginal effects shows that a resolution on an ES topic increases the likelihood that at least one of the two shareholders will oppose management by almost 60%. This compares with a less than 50% increase in likelihood for shareholder resolutions on governance issues.

¹⁸ Given that management almost always opposes such resolutions, this corresponds to investors' support for combating negative externalities.

Opposition to management with country and industry FE

Table 3

	Probit coefficients			Marginal effects		
	(1)	(2)	(3)	(1)	(2)	(3)
	BR or NF oppose	BR opposes	NF opposes	BR or NF oppose	BR opposes	NF opposes
Shareholder proposal*Resolution ES	1.867***	1.030***	1.818***	0.591***	0.132***	0.559***
Management proposal*Resolution ES	0.089	0.433**	0.146	0.014	0.033*	0.022
Shareholder proposal*Resolution G	1.594***	1.220***	1.507***	0.485***	0.179***	0.435***
Management proposal*Resolution G	0.327***	0.238***	0.296***	0.041***	0.010***	0.035***
Country rating ES	-2.768***	-5.333***	-2.195***	-0.398***	-0.259***	-0.291***
Country rating G	1.108***	1.751***	0.900***	0.159***	0.085***	0.119***
Company rating ES	-0.018	-0.026	-0.016	-0.003	-0.001	-0.002
Company rating G	-0.038***	-0.040***	-0.036***	-0.005***	-0.002***	-0.005***
Market cap	-0.048***	-0.056**	-0.048***	-0.007***	-0.003**	-0.006***
ROA	-0.011	-0.085*	0.009	-0.002	-0.004*	0.001
Price-to-book	0.004	-0.0104	0.003	0.001	-0.001	0.000
Sales growth	0.015	-0.0121119	0.017	0.002	-0.001	0.002
Asset turnover	-0.026	-0.011	-0.028	-0.004	-0.001	-0.004
Volatility	-0.048**	-0.091***	-0.037*	-0.007**	-0.004***	-0.005*
Analyst dispersion	0.001	-0.013	0.003	0.000	-0.001	0.000
Industry fixed effect	yes	yes	yes	yes	yes	yes
Country fixed effect	yes	yes	yes	yes	yes	yes
Observations	35,382	35,382	35,382	35,382	35,382	35,382
Pseudo R2	0.062	0.085	0.058	0.062	0.085	0.058

This table reports the probit coefficients and the marginal effects of variables that may explain disagreement with management. The dependent variable is a dummy variable equal to one if at least one of the two investors oppose management recommendation (1), if BlackRock opposes management recommendation (2), if the Norway fund opposes management recommendation (3), and zero elsewhere. Continuous control variables (market cap, ROA, price-to-book ratio, sales growth rate, asset turnover ratio and analyst dispersion) are normalised. Country and industry fixed effects are included in all regressions. Standard errors are clustered at the firm level. All variables are defined in the Data and Variables section. MacFadden's pseudo-R2 measure the model fit. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

Source: Authors' calculations.

Columns (2) and (3) of Table 3 are at the heart of our investigation: they display the results for opposition to management by BlackRock and the Norway Fund, respectively, in particular on ES issues involving externalities. The coefficients of the variables indicating that a resolution is sponsored by a shareholder, whether on ES or on governance, are significantly positive. This indicates that both investors tend to oppose management more for shareholder-submitted resolutions than for those on financial issues. According to a Wald test, the coefficients for governance resolutions submitted by shareholders, 1.220 for BlackRock and 1.507 for the Norway Fund respectively, are significantly larger than for those submitted by management, at 0.238 and 0.296 respectively (p-value=0.00).

Column (2) of Table 3 shows that BlackRock opposes management more on externality issues than on financial issues, but not more than on shareholder resolutions on governance (1.030 versus 1.220 respectively). Marginal effects suggest that, for BlackRock, the rate of opposition to management increases by 13% for resolutions compared with financial resolutions.

Column (3) of Table 3 shows that the Norway Fund opposes management on externality issues more than on financial issues and on governance resolutions submitted by shareholders. For the Norway Fund, the coefficient on ES issues, 1.818, is significantly different from 0 and from the coefficient on governance shareholder

resolutions, 1.507 (p-value<0.04). Marginal effects suggest that, for the Norway Fund, the rate of opposition to management increases by 56% for externality-related resolutions compared with financial resolutions (versus a 13% increase for BlackRock). For shareholder resolutions on governance, the rate of opposition of the Norway Fund increases by only 44%.

5.2. Robustness checks

To check the robustness of our findings, we run the same regressions as before, but we omit country fixed effects. The results are displayed in Table 4, Columns (1) through (3). They are very similar to those in Table 3. The regressions displayed in Table 4, Column (4) and (5), are estimated jointly. They suggest that our findings are valid when running a bivariate probit regression instead of univariate regressions. Moreover, the joint estimation of BlackRock's and the Norway Fund's voting policies enables us to compare the propensity of each investor to oppose management on externality issues and thus support efforts to improve ES behaviour. For the Norway Fund, the coefficient on ES shareholder resolutions is significantly larger, 1.816, than for BlackRock, 1.131 (p-value=0.00).

Opposition to management and bivariate probit estimations without country FE

Table 4

	Probit coefficients			Bivariate probit Coefficients	
	(1) BR or NF oppose	(2) BR opposes	(3) NF opposes	(4) BR opposes	(5) NF opposes
Shareholder proposal*Resolution ES	1.879***	1.036***	1.828***	1.131***	1.816***
Management proposal*Resolution ES	-0.022	0.267	0.039	0.168	-0.001
Shareholder proposal*Resolution G	1.532***	1.191***	1.447***	1.262***	1.455***
Management proposal*Resolution G	0.313***	0.213***	0.283***	0.210***	0.304***
Country rating ES	-3.429***	-6.609***	-2.852***	-6.143***	-2.837***
Country rating G	0.868***	1.784***	0.692***	1.625***	0.706***
Company rating ES	-0.003	-0.015	-0.002	-0.011	-0.002
Company rating G	-0.036***	-0.038***	-0.035***	-0.038***	-0.034***
Market cap	-0.045***	-0.053**	-0.045***	-0.049**	-0.044***
ROA	-0.004	-0.084*	0.015	-0.088**	0.012
Price-to-book	0.005	-0.009	0.005*	-0.010	0.005*
Sales growth	0.016	-0.012	0.018	-0.019	0.018
Asset turnover	-0.029	-0.024	-0.031	-0.023	-0.030
Volatility	-0.043**	-0.089**	-0.031	-0.086***	-0.033
Analyst dispersion	-0.005	-0.013	-0.003	-0.019	-0.003
Industry fixed effect	yes	yes	yes	yes	yes
Country fixed effect	no	no	no	no	no
Observations	35,382	35,382	35,382	35,382	35,382
Pseudo R2/Prob Wald Chi2	0.047	0.068	0.044	0.000	0.000

This table reports the probit coefficients of variables that may explain disagreement with management ((1), (2), (3)) without country fixed effects. The dependent variable is a dummy variable equal to one if at least one of the two investors oppose management recommendation (1), if BlackRock opposes management recommendation (2), if the Norway fund opposes management recommendation (3), and zero elsewhere. Columns (4) and (5) report the coefficients of variables that may explain disagreement with management from a bivariate probit estimation without country fixed effects. Specifications (4) and (5) are estimated simultaneously to capture the joint effect of BlackRock opposing to management when the Norway Fund agrees with management recommendation (4), and the Norway Fund opposing to management when BlackRock agrees (5). Continuous control variables (market cap, ROA, price-to-book ratio, sales growth rate, asset turnover ratio and analyst dispersion) are normalised. Industry fixed effects are included in all regressions. Standard errors are clustered at the firm level. All variables are defined in the Data and Variables section. MacFadden's pseudo-R² measure the model fit. For the bivariate probit, the last row reports the probability of the Wald Chi2 test that measures the model fit. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

Source: Authors' calculations.

To check that our results hold when controlling for investors' holdings in firms, we restrict our attention to the firms listed in the SEC 13F filings that record the holdings of institutional investors. Tables 5 and 6 display the same type of information as Tables 1 and 2, but for the sample the information is restricted to firms in the 13F filings. In this sample, we find larger firms but the overall image in terms of the type of resolution voted upon is qualitatively similar.

Voted resolutions and rate of opposition to management: reduced sample

Table 5

Sponsor of the resolution	BR disagrees with management		NF disagrees with management		Total number of voted resolutions
	<i>Management</i>	<i>Shareholder</i>	<i>Management</i>	<i>Shareholder</i>	
E					32
Animal welfare					2
Animal testing	-	0%	-	0%	1
Animal welfare policies	-	0%	-	100%	1
Climate					18
Climate change and GHG emissions	-	6%	-	78%	18
Environment and sustainability					8
Hydraulic fracturing	-	0%	-	67%	3
Nuclear safety	-	-	-	-	-
Sustainability reporting	-	0%	-	40%	5
Others	-	25%	-	50%	4
S					78
Consumer issues					6
Genetically modified ingredients	-	0%	-	20%	5
Other consumer responsibility	-	0%	-	0%	1
Diversity					4
Board diversity	-	50%	-	100%	2
Discrimination	-	-	-	-	-
Sexual orientation	-	0%	-	50%	2
General corporate issues					1
Charitable contributions	-	0%	-	0%	1
Human rights					13
Human rights proposals	-	15%	-	38%	13
Political activities					54
Lobbying	-	13%	-	38%	24
Political contributions	-	10%	-	57%	30
G					5,748
Audit practices and risk management					565
Audit practices	0%	-	2%	-	564
Risk management	-	0%	-	0%	1
Board accountability and responsiveness					2
Ability to remove directors	-	0%	-	0%	1
Tax transparency	-	0%	-	100%	1
Board independence					26
Competitive activities of directors	-	-	-	-	-
Independent chairman and directors	-	8%	-	85%	26

Board structure					4,251
Appointment	1%	0%	7%	0%	4,223
Board composition	0%	-	0%	-	4
Others board related proposals	0%	33%	0%	67%	24
Related-party transaction	-	-	-	-	-
Compensation/Remuneration					827
Employee compensation	5%	0%	8%	50%	182
Executive compensation	2%	3%	7%	61%	645
Shareholder rights					77
Call special meeting	0%	33%	0%	67%	19
Proxy access right	0%	43%	0%	29%	8
Takeover defenses	0%	80%	0%	20%	26
Voting formalities	14%	24%	0%	41%	24
Financial	2%	18%	20%	68%	156
Others	9%	-	22%	-	23
Total					6,037

This table summarises the percentage of opposition to management recommendation and the number of voted resolutions by the two investors in 2014, within each area, theme and issue. The sample includes only the firms for which we managed to collect data on their characteristics and holdings from 13F filings. Columns 2, 3, 4 and 5 report the rate of opposition to management within each issue, depending on the sponsor of each resolution (management and shareholder). Column 6 provides the number of voted resolutions by the two investors in 2014, within each area, theme and issue, on this reduced sample.

Source: Authors' calculations.

Summary statistics: reduced sample

Table 6

Variable	N	Mean	Std Dev	Min	Max
Panel A: Measures of disagreement					
Oppos_atleystone	6,037	0.09	0.29		
Oppos_blackrock	6,037	0.02	0.12		
Oppos_norway	6,037	0.09	0.28		
Panel B: Resolution characteristics					
Shareholder proposal	6,037	0.04	0.20		
Resolution ES	6,037	0.02	0.13		
Resolution G	6,037	0.95	0.21		
Panel C: Country and firm ESG ratings					
Country Rating ES	6,037	0.7	0.00	0.65	1
Country Rating G	6,037	0.75	0.02	0.75	0.89
Company Rating ES	6,037	4.62	1.48	0.55	9.95
Company Rating G	6,037	6.73	2.79	0	10.00
Panel D: Firm characteristics					
Market cap	6,037	29,375	54,882	249	439,000
ROA	6,037	4.88	7.61	-44.80	64.20
Price-to-book	6,037	7.56	76.66	0.53	1,540.00
Sales growth	6,037	0.08	0.27	-1.00	3.79
Asset turnover	6,037	0.72	0.70	0.00	4.45
Volatility	6,037	38.58	14.83	14.08	150.77
Analyst dispersion	6,037	0.10	0.76	-8.14	20.10
Panel E: Company holdings					
Weight in BR portfolio	6,037	0.08%	0.32%	0.00%	3.00%
Holding BR (% of capitalisation)	6,037	0.76%	2.17%	0.00%	28.29%
Weight in NF portfolio	6,037	0.07%	0.30%	0.00%	3.00%
Holding NF (% of capitalisation)	6,037	2.81%	7.37%	0.00%	101.57%
Weight in portfolio (average BR NF)	6,037	0.13%	0.28%	0.00%	2.92%
Holding (average BR NF)	6,037	1.79%	4.60%	0.00%	51.14%

This table provides summary statistics for the 6,037 common resolutions the Norway Fund and BlackRock voted upon in 2014 for which we managed to collect, besides firms' characteristics, the holdings of the two investors as provided by the 13F filings. Panel A refers to the disagreement measures, Panel B to the characteristics of the resolutions, Panel C to the extra-financial performance measures for firms and countries, Panel D to firms' financial characteristics and Panel E to the Norway Fund and BlackRock's holdings in these firms. All variables are defined in the Data and Variables section.

Source: Authors' calculation

The results are in Table 7. The second line of the table includes NA values because there are no management proposals on ES issues in this reduced sample. We control for two types of holding measure: the weight of the firm in the investor's portfolio (company weight in portfolio) and the proportion of the firm's stock held by the investor (company holding). Columns (1) through (3) display the results of the same regression as before, ie without including holdings as a control, but on the sample the results are restricted to firms in the 13F filings. Columns (4) through (6) display the results for the regressions that include holdings as a control. Holdings appear not to affect the voting policy of the two investors.

Opposition to management: reduced sample

Table 7

	Probit coefficients					
	(1) BR or NF oppose	(2) BR opposes	(3) NF opposes	(4) BR or NF oppose	(5) BR opposes	(6) NF opposes
Shareholder proposal*Resolution ES	0.728***	0.471*	0.693***	0.722***	0.459*	0.689***
Management proposal*Resolution ES	NA	NA	NA	NA	NA	NA
Shareholder proposal*Resolution G	1.015***	0.703***	0.834***	1.020***	0.709***	0.837***
Management proposal*Resolution G	-0.985***	-0.816***	-0.961***	-0.984***	-0.813***	-0.960***
Country rating ES	-56.961***	10.192	-54.133***	-56.969***	10.283	-54.076***
Country rating G	NA	NA	NA	NA	NA	NA
Company rating ES	-0.005	0.041	-0.005	-0.006	0.038	-0.006
Company rating G	-0.015	-0.032	-0.006	-0.015	-0.032	-0.005
Company weight in portfolio	NA	NA	NA	10.242	15.563	9.467
Company holding (% of capitalisation)	NA	NA	NA	-0.756	-0.988	-0.432
Market cap	-0.077***	-0.054**	-0.066***	-0.079***	-0.057**	-0.067***
ROA	-0.004	-0.243*	0.010	-0.007	-0.257*	0.007
Price-to-book	-0.003	-0.299	-0.002	-0.003	-0.321	-0.002
Sales growth	-0.481	1.275*	-0.639	-0.476	1.282*	-0.630
Asset turnover	-0.214***	-0.194*	-0.193***	-0.209***	-0.186*	-0.191***
Volatility	-0.088*	-0.221**	-0.0763742	-0.086*	-0.221**	-0.076
Analyst dispersion	-0.012	0.008	-0.014	-0.012	0.008	-0.013
Industry fixed effect	yes	yes	yes	yes	yes	yes
Observations	6,037	5,648	6,037	6,037	5,648	6,037
Pseudo R2	0.143	0.181	0.130	0.143	0.182	0.131

This table reports the probit coefficients of variables that may explain disagreement with management. The sample includes only the firms for which we managed to collect data on their characteristics and the holdings of the two investors. The dependent variable is a dummy variable equal to one if at least one of the two investors oppose management recommendation (1) and (4), if BlackRock opposes management recommendation (2) and (5), if the Norway Fund opposes management recommendation (3) and (6), and zero elsewhere. Columns (1), (2) and (3) report the results without holdings, while columns (4), (5) and (6) summarise the results with holdings. Company holding/company weight in portfolio refer to holding (average BR NF)/weight in portfolio (average BR NF) in specification (4), holding BR (% of capitalisation)/weight in BR portfolio in specification (5), holding NF (% of capitalisation)/Weight in NF portfolio in specification (6). Continuous control variables (market cap, ROA, price-to-book ratio, sales growth rate, asset turnover ratio and analyst dispersion) are normalised. Industry fixed effects are included in all regressions. Standard errors are clustered at the firm level. All variables are defined in the Data and Variables section. MacFadden's pseudo-R2 measure the model fit. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

Source: Authors' calculations.

In the two specifications included in Table 7, the results are as follows: compared with proposals on financial issues, both investors appear to oppose management (i) more often for shareholder proposals, whether on ES or on governance issues, and (ii) less often for management proposals on governance. Regarding shareholder proposals, the result that opposition to management on ES issues is larger than on governance issues is no longer present. There is a clear sample effect. For example, in this sample, management resolutions on governance arouse significantly less opposition from institutional investors than financial resolutions do, a result that is reversed compared with our full sample.

However, it is still the case that the Norway Fund opposes management more often than BlackRock on shareholder-sponsored resolutions on externality issues. When we include holdings as a control variable, the coefficient for the Norway Fund

is 0.689, larger than the 0.459 coefficient for BlackRock. Finally, the coefficients on the holdings' variables are not significant.

5.3. Climate change resolutions

We now study shareholder resolutions that request firms to adopt policies to combat climate change. This type of resolution calls on management to, for example, "Report on financial and physical risks of climate change", "Indicate quantitative goals for GHG and other air emissions", and "Review public policy advocacy on climate change."

To study how BlackRock and the Norway Fund vote on resolutions that are clearly related to an externality, we include a dummy variable for a resolution asking the firm to adopt a climate change mitigation policy. At shareholder meetings, these resolutions are always submitted by shareholders, and management always opposes them.¹⁹

Table 8 displays our results. We find that the Norway Fund opposes management more often on climate-related resolutions than on other externality resolutions and on shareholder resolutions on governance (p-value=0.00). This indicates that the Norway Fund has a strong tendency to vote in favour of climate change mitigation policies, despite management's negative recommendations. The results for BlackRock are very different: it does not oppose management on climate-related resolutions more than on financial issues (the coefficient on the climate-related resolution dummy is insignificant). Moreover, BlackRock opposes management on climate resolutions less than on other ES resolutions and even less than on governance resolutions.

¹⁹ This is not to say that firms' management never implement climate change mitigation policies on their own initiative. Our sample focuses only on the firms in which resolutions were filed by shareholders to impose externality-related policies on corporate management. Hence, behind-the-doors engagement has failed (McCahery et al (2016)). Firms in which management has voluntarily implemented policies to mitigate negative externalities are thus excluded.

	Probit coefficients		
	(1) BR or NF oppose	(2) BR opposes	(3) NF opposes
Shareholder proposal*Resolution climate	2.785***	0.720	2.806***
Shareholder proposal*Resolution ES non climate	1.754***	1.063***	1.694***
Management proposal*Resolution ES non climate	0.088	0.435**	0.145
Shareholder proposal*Resolution G	1.594***	1.219***	1.507***
Management proposal*Resolution G	0.327***	0.238***	0.297***
Country rating ES	-2.777***	-5.331***	-2.205***
Country rating G	1.107***	1.752***	0.899***
Company rating ES	-0.018	-0.026	-0.016
Company rating G	-0.038***	-0.040***	-0.036***
Market cap	-0.048***	-0.056**	-0.048***
ROA	-0.011	-0.085*	0.009
Price-to-book	0.004	-0.010	0.003
Sales growth	0.015	-0.012	0.017
Asset turnover	-0.025	-0.011	-0.028
Volatility	-0.048**	-0.091***	-0.037*
Analyst dispersion	0.001	-0.013	0.003
Industry fixed effect	yes	yes	yes
Country fixed effect	yes	yes	yes
Observations	35,382	35,367	35,382
Pseudo R2	0.062	0.085	0.058

This table reports the probit coefficients of variables that may explain disagreement with management with focus on the impact of climate resolutions. The dependent variable is a dummy variable equal to one if at least one of the two investors oppose management recommendation (1), if BlackRock opposes management recommendation (2), if the Norway Fund opposes management recommendation (3), and zero elsewhere. Continuous control variables (market cap, ROA, price-to-book ratio, sales growth rate, asset turnover ratio and analyst dispersion) are normalised. Industry and country fixed effects are included in all regressions. Standard errors are clustered at the firm level. All variables are defined in the Data and Variables section. MacFadden's pseudo-R2 measure the model fit. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

Source: Authors' calculations.

6. Discussion

Our basic specification makes it possible to characterise the voting policy on externality issues of BlackRock and of the Norway Fund, separately, and thus to test the relative influence of universal ownership and delegated philanthropy.

The results of our basic specification indicate that externality issues generate more opposition from the two investors than financial and governance resolutions, including those governance resolutions submitted to a vote by shareholders. It thus suggests that there may be something peculiar about ES issues, which we interpret

as being related to externalities. Moreover, the fact that opposition of both investors is larger for governance proposals submitted by shareholders than for governance proposals sponsored by the management is in line with the existence of agency conflicts (see eg Jensen and Meckling (1976)).

When we disentangle the behaviour of each investor regarding its opposition to management on the different resolutions, we set forth some additional insights. More specifically, the results for BlackRock indicate that universal ownership is a significant motivation for engagement of this institutional investor, although not as strong as agency conflicts, as manifested in opposition to shareholder resolutions on governance. Indeed, BlackRock does not oppose management more on externality-related than on governance-related shareholder resolutions. This is exactly contrary to what we observe for the Norway Fund, indicating that delegated philanthropy may also be present in the data and constitutes a driver of engagement on externality issues for some institutional investors.²⁰

Overall, our results suggest that BlackRock and the Norway Fund actively oppose managers at shareholder meetings: both tend to oppose management more frequently on shareholder-sponsored proposals than on management-sponsored ones. However, only the Norway Fund opposes management more for shareholder resolutions on externality issues than for governance issues. This suggests that delegated philanthropy provides additional incentives beyond universal ownership for institutional investors to curb negative externalities generated by firms.

These results seem robust when we test them using several different specifications. The joint estimation of BlackRock's and the Norway Fund's voting policies in a bivariate setting confirms that the Norway fund's propensity to oppose management and support efforts to improve ES behaviour is significantly higher than for BlackRock. Furthermore, the coefficients on industry dummies²¹ show evidence that the propensity of BlackRock to vote against management on ES proposals is higher in some industries where negative ES externalities are lower, eg the financial and technology sector, compared with industries like energy or materials in which negative externalities may be more present. This is not the case for the Norway Fund, whose voting policy on ES proposals is not driven by a particular industry.

Moreover, institutional investors' opposition to management depends neither on the proportion of a firm's equity they hold nor on the proportion of a firm in an investor's portfolio and our main conclusion is not reversed in the 13F sample: More often than BlackRock, the Norway Fund favours firms' policies oriented towards mitigating negative externalities. Investors' holdings seem not to affect their voting policy.

Finally, we focus on climate resolutions because (i) climate change poses a major economic challenge, with potentially disastrous consequences, and (ii) GHG emissions

²⁰ Column (1) of Table 3 shows that investors do not oppose corporate management more for management-sponsored proposals on ES issues than for financial issues. In the view of investors, some ES management proposals are as beneficial as traditional financial proposals. This might be because some corporate policies related to ES might be good for the firms. This would explain why they are proposed by management and not refused by shareholders. Moreover, this result reinforces our interpretation that ES proposals made by shareholders are related to policies to curb externalities. These policies may be viewed as being detrimental to firm value, explaining why management opposes them, while beneficial to society, which is why (some) investors favour them.

²¹ Available upon request from the authors.

represent a clear externality that firms impose on society. Indeed, firms emit GHG into the atmosphere but avoid the cost of this negative externality because there is no appropriate global regulation in place, whether based on GHG taxes (Pigou (1920)) or on emission allowance markets (Coase (1960)).²² As indicated by Gollier and Tirole (2015): “Most benefits of mitigation are global and distant, while costs are local and immediate.” Firms are thus likely to emit more GHG than a benevolent global social planner would require.

The fact that the Norway Fund opposes management more often on climate-related resolutions than on other externality resolutions and on shareholder resolutions on governance, while it is exactly the contrary for BlackRock, is consistent with universal ownership not being a sufficiently strong motivation to get institutional investors to engage with corporations to fight negative externalities.

Two arguments may nevertheless nuance this main result. First, it is now well known that major money managers, such as BlackRock, engage constantly with companies they hold in their portfolios through different types of direct and private communication with corporate managers. Thus, voting in line with management might also express the fact that potential divergences were discussed and solved ahead of the general meeting, behind closed doors, or negotiated against other management concessions. Unfortunately, as these exchanges are private, it is very difficult to track them and have a clear picture of the topics discussed. However, the same institutional investors are not reluctant to vote against management if they appreciate that firms did not show significant evidence of progress on the issues discussed or if the management is not responsive enough, as stated by Larry Fink in 2018 in a letter to the CEOs of BlackRock portfolio companies.²³ Second, not all ES resolutions submitted by shareholders may be clearly stated, reasonable, or have the same impact in terms of the ES outcome. As an example, on the 13 shareholder resolutions on nuclear safety, neither BlackRock nor the Norway Fund opposed management who were not in favour of the resolutions. A close look at these resolutions shows that they are all proposed for Japanese companies, with a very vague and unclear content. Unfortunately, the limited number of ES resolutions we have in this sample makes it complex to create various clusters on the different issues and get statistically meaningful results. Also, we do not have access to the final outcome of the votes, which could have provided a clearer picture of the resolutions that managed to gather general support beyond the two institutional investors under study and those which were indeed considered as having negligible ES positive outcomes by the same majority of shareholders.

²² Emissions markets exist across the globe, but they are sometimes partial (as in the case of the European Union Emissions Trading System), and do not ensure a single global price for GHG emissions.

²³ “BlackRock can choose to sell the securities of a company if we are doubtful about its strategic direction or long-term growth. In managing our index funds, however, BlackRock cannot express its disapproval by selling the company’s securities as long as that company remains in the relevant index. As a result, our responsibility to engage and vote is more important than ever.” (www.blackrock.com/corporate/en-no/investor-relations/larry-fink-ceo-letter).

7. Conclusion

This paper studies voting at shareholder meetings by two emblematic investors: BlackRock, a large, well diversified, financially oriented investor, and the Norway Fund, a large, well diversified, responsible sovereign wealth fund. Both are universal owners (see eg Monks and Minow (1995)), and the Norway Fund also has a delegated philanthropic mission (see eg Benabou and Tirole (2010)), monitored by the Norwegian parliament and a Council on Ethics.

Our data cover 2014 and include the two institutions' votes on 35,382 resolutions at 2,796 corporations worldwide, as well as managers' recommendations. We find that both investors oppose management more significantly for shareholder resolutions on ES issues than on financial resolutions. The data seem to thus suggest that a universal owner philosophy is at work. Moreover, we find that support for resolutions on reducing negative externalities is stronger at the Norway Fund than at BlackRock. This suggests that the delegated philanthropy logic is also at work. Our findings hold with and without country fixed effects and also if we restrict our analysis to meetings of firms for which investors' holdings are available. Our results are even more significant when we restrict our analysis to climate change issues. For these issues that clearly constitute externalities, we find that only the Norway Fund and its delegated philanthropy logic oppose management in an attempt to fight climate change.

Overall, our findings suggest that delegated philanthropy is stronger than universal ownership at providing incentives for institutional investors to combat negative externalities generated by firms. These findings have important policy implications. They suggest that corporations – in particular large multinationals – that have significant influence on the future of the planet, are unlikely to be disciplined by institutional investors simply because these investors hold well diversified portfolios. Instead, we find that institutional investors' corporate engagement policies ought to reflect the values of their clients or beneficiaries. This could be achieved by setting up pass-through voting or, more generally, by basing engagement policies on mechanisms designed to measure clients' values on the main externality issues.

References

- Agrawal, A and C Knoeber (1996): "Firm performance and mechanisms to control agency problems between managers and shareholders", *Journal of Financial and Quantitative Analysis*, vol 31, no 3, pp 377–97.
- Authers, J (2015): "Responsible investment: vice versus nice", *Financial Times*, 25 June.
- Azar, J (2017): "Portfolio diversification, market power, and the theory of the firm", *IESE Research Papers*, D/1170.
- Azar, J, M Schmaltz and I Tecu (2017): "Anti-competitive effects of common ownership", *Ross School of Business Paper*, no 1235.
- Bach, L and D Metzger (2017): *How do shareholder proposals create value?*, <http://dx.doi.org/10.2139/ssrn.2247084>.
- Barko, T, M Cremers and L Renneboog (2017): "Activism on corporate social responsibility", *ECGI Finance Working Paper*, no 509.
- Bauer, R, R Braun and M Viehs (2010): "Industry competition, ownership structure and shareholder activism", *ECCE Working Paper*, Maastricht University.
- Bebchuk, L A, A Cohen and S Hirst (2017): "The agency problems of institutional investors", *Journal of Economic Perspectives*, vol 31, no 3, pp 89-102.
- Becht, M, J Franks, C Mayer and S Rossi (2009): "Returns to shareholder activism: evidence from a clinical study of the Hermes UK Focus Fund", *Review of Financial Studies*, vol 22, no 8, pp 3093–129.
- Benabou, R and J Tirole (2010): "Individual and corporate social responsibility", *Economica*, vol 77, pp 1–19.
- (2016): "Bonus culture: competitive pay, screening, and multitasking", *Journal of Political Economy*, vol 124, no 2, pp 305–70.
- Berle, A and G Means (1932): *The Modern Corporation and Private Property*, Macmillan.
- Boot, A, R Gopalan and A Thakor (2006): "The entrepreneur's choice between private and public ownership", *Journal of Finance*, vol 61, no 2, pp 803–36.
- (2008): "Market liquidity, investor participation, and managerial autonomy: why do firms go private?", *Journal of Finance*, vol 63, no 4, pp 2013–59.
- Chambers, D, E Dimson and A Ilmanen (2012): "The Norway model", *Journal of Portfolio Management*, vol 38, no 2, pp 67–81.
- Chen, J, H Hong and J Stein (2002): "Breadth of ownership and stock returns", *Journal of Financial Economics*, vol 66, no 2–3, pp 171–205.
- Cheng, I H, H Hong and K Shue (2013): "Do managers do good with other peoples' money?", *Chicago Booth Research Paper*, no 12–47.
- Coase, R (1960): "The problem of social cost", *Journal of Law and Economics*, vol 3, October, pp 1–44.
- Council on Ethics (2016): *Council on Ethics for the Norwegian Government Pension Fund Global: Annual Report 2016*, https://nettsteder.regjeringen.no/etikkradet3/files/2017/03/Etikkradet_annual_report_2016_web.pdf.

- Cunat, V, M Gine and M Guadalupe (2012): "The vote is cast: The effect of corporate governance on shareholder value", *Journal of Finance*, vol 67, no 5, pp 1943–77.
- Cziraki, P, L Renneboog and P Szilagyi (2010): "Shareholder activism through proxy proposals: the European perspective", *European Financial Management*, vol 16, no 5, pp 738–77.
- Diether, K, C Malloy and A Scherbina (2002): "Differences of opinion and the cross section of stock returns", *Journal of Finance*, vol 57, no 5, pp 2113–41.
- Dimson, E, I Kreutzer, R Lake, H Sjo, and L Starks (2013): *Responsible Investment and the Norwegian Government Pension Fund Global: Main Report*, Norwegian Ministry of Finance Strategy Council.
- Dimson, E, O Karakas and X Li (2015): "Active ownership", *Review of Financial Studies*, vol 28, no 12, pp 3225–68.
- Fichtner, J, E Heemskerk and J Garcia-Bernardo (2017): "Hidden power of the Big Three? Passive index funds, re-concentration of corporate ownership, and new financial risk", *Business and Politics*, vol 19, no 3, pp 298–326.
- Flammer, C (2015): "Does corporate social responsibility lead to superior financial performance? A regression discontinuity approach", *Management Science*, vol 61, no 11, pp 2549–68.
- Flammer, C and P Bansal (2017): "Does a long-term orientation create value? Evidence from a regression discontinuity", *Strategic Management Journal*, vol 38, no 9, pp 1827–47.
- Gillian, S and L Starks (2000): "Corporate governance proposals and shareholder activism: the role of institutional investors", *Journal of Financial Economics*, vol 57, no 2, pp 275–305.
- Gollier, C and J Tirole (2015): "Negotiating effective institutions against climate change", *Economics of Energy and Environmental Policy*, vol 4, no 2, pp 5–27.
- Hawley, J and A Williams (2000): "The emergence of universal owners: some implications of institutional equity ownership", *Challenge*, vol 43, no 4, pp 43–61.
- Hong, H, J Kubik and J Scheinkman (2012): "Financial constraints on corporate goodness", *NBER Working Paper*, no 18476.
- Hong, H and J Stein (2003): "Differences of opinion, short-sales constraints, and market crashes", *Review of Financial Studies*, vol 16, no 2, pp 487–525.
- Jensen, M and W Meckling (1976): "Theory of the firm: managerial behavior, agency costs and ownership structure", *Journal of Financial Economics*, vol 3, no 4, pp 305–60.
- Johnson, T C (2004): "Forecast dispersion and the cross section of expected returns", *Journal of Finance*, vol 59, no 5, pp 1957–1978.
- Keswani, A, D Stolin and A Tran (2016): "Frenemies: how do financial firms vote on their own kind?", *Management Science*, vol 63, no 3, pp 631–54.
- Kaufmann, D, A Kraay and M Mastruzzi (2010): "The worldwide governance indicators: a summary of methodology, data and analytical issues", *World Bank Policy Research Working Paper*, no 5430.
- Laffont, J (1987): "Externalities", in J Eatwell, M Milgate and P Newman (eds), *The New Palgrave: A Dictionary of Economics*, Palgrave Macmillan.

Mattison, R, M Trevitt and L Van Ast (2011): "*Universal ownership: why environmental externalities matter to institutional investors*", Principles for Responsible Investment and UNEP Finance Initiative.

McCahery, J, Z Sautner and L Starks (2016): "Behind the scenes: the corporate governance preferences of institutional investors", *Journal of Finance*, vol 71, no 6, pp 2905–32.

Monks, R and N Minnow (1995): *Corporate Governance*, Cambridge, Basil Blackwell.

Morgan, J and J Tumlinson (2012): "Corporate provision of public goods", *Working Paper*, University of California Berkeley.

Pigou, A (1920): *The Economics of Welfare*, Macmillan.

Renneboog, L and P Szilagyi (2011): "The role of shareholder proposals in corporate governance", *Journal of Corporate Finance*, vol 17, no 1, pp 167–88.

Smith, M (1996): "Shareholder activism by institutional investors: evidence from CalPERS", *Journal of Finance*, vol 51, no 1, pp 227–52.

The Bretton Wood II Leaders List (2017): *BWII Leaders List: The 25 Most Responsible Asset Allocators*,

www.dalberg.com/system/files/2017-10/BW2_event_document_FINAL11.pdf.

Part 3

Portfolio Construction

Robust optimisation by constructing near-optimal portfolios

Martin van der Schans,¹ Tanita de Graaf,² and Wessel van Eeghen¹

Abstract

Many investors use optimisation to determine their optimal investment portfolio. Unfortunately, optimal portfolios are sensitive to the optimisation's required input, ie they are not robust. Traditional robust optimisation approaches seek to provide an optimal and robust portfolio and, in doing so, replace the investor's investment decision process. In practice, however, portfolio optimisation supports but seldomly replaces the investment decision process. In this paper, we present an approach that both solves the robustness problem and aims to support rather than replace the investment decision process. We determine a region with near-optimal portfolios that, especially in light of the robustness problem, are all satisfactory allocation decisions. Then, as is already common practice, the investor can bring in expert opinions and additional information to select a preferred near-optimal portfolio. We will show that the region of near-optimal portfolios is more robust than the optimal portfolio itself.

JEL classification: C61, G11.

¹ Ortec Finance, PO Box 4074, 3006 AB Rotterdam, The Netherlands, www.ortec-finance.com. This version: 5 June 2019.

² ORTEC BV.

1. Robust optimisation

Many investors use portfolio optimisation to determine their optimal investment portfolio. Unfortunately, optimal portfolios are sensitive to the optimisation's input parameters. Although this sensitivity is often studied in the context of mean-variance optimisation, where optimal portfolios are sensitive to the estimated mean and covariance matrix (Frankfurter et al (1971), Michaud (1989), Chopra and Ziemba (1993)), sensitivity is a generic problem in portfolio optimisation (Kondor et al (2007), Ciliberti et al (2007)). As discussed in Hurley and Brimberg (2015), the sensitivity is caused by an interaction of an estimation error in the input and the optimisation objective.

The literature proposes several robust optimisation approaches to deal with the sensitivity problem. Although they differ methodologically, approaches such as shrinkage (Ledoit and Wolf 2004), robust statistics (Reyna et al (2005)), Black-Litterman inverse optimisation (Bertsimas et al (2012)) and Bayesian optimisation (Schöttle et al (2010)) reduce the sensitivity by diminishing the role of the data on which the optimisation parameters are estimated, ie they reduce estimation errors in the optimisation input or their effect. Other methods, such as regularisation, change the optimisation objective to make it less sensitive to estimation errors in the optimisation parameters (DeMiguel et al (2009), Brodie et al (2009)). Also, there are hybrid methods that do both. For example, the optimisation community proposes a general robust optimisation framework. Given a convex optimisation problem, the framework proposes a robust counterpart that lets the input vary within a specified range and selects the worst case outcome (Ben-Tal and Nemirovski (1998)). Finally, there is the resampled frontier (Michaud (1998)), which is constructed by resampling the input from a distribution and averaging over the resamplings' optimisation results.

Robust optimisation approaches work well when all market information is quantified and incorporated in the optimisation problem. However, despite efforts to incorporate information such as transaction costs, expert opinion and liquidity into portfolio optimisation problems, they remain a simplification of reality. In practice, investors often combine the optimal portfolio with additional information that was not or could not be incorporated. Thus, for investors, portfolio optimisation is a tool that supports but does not replace their decision process. In this paper, we take this as the starting point for developing a robust optimisation approach.

2. Near-optimal portfolios

Generally, the result of a portfolio optimisation problem is an efficient frontier with optimal portfolios. Now, given an optimal portfolio w_0 on the efficient frontier, we construct a near-optimal region just below the efficient frontier as indicated by the shaded region in Figure 1. The portfolios in this region are referred to as near-optimal portfolios. As shown in Chopra (1993) and Section 4, the near-optimal portfolios' weights can differ completely from those of the optimal portfolio. The idea is that all near-optimal portfolios still have a satisfactory risk-return trade-off so that investors can, as is already common practice, bring in additional arguments and select their preferred near-optimal portfolio.

To find the near-optimal region represented by the shaded region in Figure 1, we construct near-optimal portfolios w_1, \dots, w_n far away from each other and show that any weighted average of these portfolios, ie any portfolio in their convex hull,

$$\text{Conv}(w_1, \dots, w_n) := \{ \sum_{i=1}^n \theta_i w_i \mid \theta_i \geq 0, \sum_{i=1}^n \theta_i = 1 \}, \quad (1)$$

Figure 1: Efficient frontier and near-optimal region

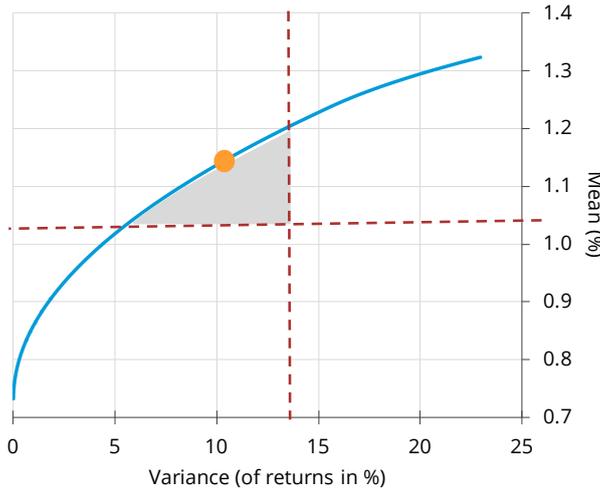


Table 1: Statistics and optimal allocation

	Stocks	Bonds	T-bills
Mean	1.32%	1.03%	0.73%
Std. dev.	4.79%	3.98%	0.22%
Correlations:			
Stocks	1.000		
Bonds	0.341	1.000	
T-bills	-0.081	0.050	1.000
Optimal allocation	58.10%	22.80%	19.10%

Figure 1 shows a mean-variance efficient frontier based on the statistics in Table 1, an optimal allocation (orange dot) and a shaded region with near-optimal portfolios. Table 1 shows statistics of monthly returns from January 1980 to December 1990 as reported by Chopra (1993) and the portfolio weights of the orange dot in Figure 1.

is near-optimal. We continue constructing near-optimal portfolios until their convex hull sufficiently covers the near-optimal region. We will show that the near-optimal region is more robust than the optimal portfolio w_0 itself. The intuitive understanding is that, with slightly different input parameters, the near-optimal region slightly changes in shape. For example, the old optimum becomes near-optimal and one of the near-optimal portfolios becomes optimal. In particular, we will show that most near-optimal portfolios remain near-optimal. And, because near-optimal portfolios have, by construction, satisfactory risk-return trade-offs, the investor's allocation does not need to be revised and becomes more robust.

3. Methodology

3.1 Constructing near-optimal portfolios

The construction of near-optimal portfolios consists of the following steps:

1. start with an efficient frontier and an optimal portfolio w_0 on the frontier as in Figure 1;
2. specify the near-optimal region as indicated by the shaded region in Figure 1;
3. find the portfolio w_1 in the near-optimal region with the highest risk and return;
4. find the portfolio w_2 in the near-optimal region that differs most in allocation from w_1 ;
5. find the portfolio w_3 in the near-optimal region that differs most in allocation from the weighted averages of w_1 and w_2 , ie from all portfolios in $\text{Conv}(w_1, w_2)$; and
6. continue until $\text{Conv}(w_1, \dots, w_n)$ covers the near-optimal region up to a required precision ϵ .

In this section, we specify these steps in more detail. Although near-optimal portfolios can be found below any efficient frontier, we assume for simplicity that the investor is interested in near-optimal mean-variance portfolios. Thus, the efficient frontier is constructed by solving

$$\min_w \lambda w^T \Sigma w - (1 - \lambda) w^T \mu, \quad (2a)$$

$$Aw = b, \quad (2b)$$

$$Gw \leq h, \quad (2c)$$

where the vector μ contains the asset's mean returns, Σ is their covariance matrix, A is a matrix representing, together with the vector b , the equality constraints, G is a matrix representing, together with the vector h , the inequality constraints and $0 \leq \lambda \leq 1$ represents the investor's risk aversion. Also, equality constraints (2b) should at least enforce that the sum of all portfolio weights w equals one. Apart from this restriction, constraints (2b) and (2c) can be chosen freely and, for example, be used to prevent short selling or fix the allocation to certain asset classes. In Step 1, the investor solves mean-variance optimisation problem (2) for a number of risk aversion parameters λ and obtains the efficient frontier as denoted in Figure 1 by the solid blue line. From the frontier, the investor selects an optimal portfolio w_0 with an appropriate risk-return trade-off. In Step 2, the investor specifies a region R of near-optimal portfolios around the optimal portfolio w_0 by specifying a minimum return μ_{\min} so that

$$w^T \mu \geq \mu_{\min}, \quad (3)$$

and a maximum variance σ_{\max}^2 so that

$$w^T \Sigma w \leq \sigma_{\max}^2. \quad (4)$$

The near-optimal region $R(\mu_{\min}, \sigma_{\max}^2)$ thus consists of all portfolios that satisfy (2b), (2c), (3) and (4), and, it is represented by the shaded region in Figure 1. It can be shown that the near-optimal region is convex, which means that any weighted average of the near-optimal portfolios is near-optimal (see Appendix A).

Next, in Step 3, we find a portfolio w_1 with maximum risk and return which is located in the region's upper right corner. Because, apart from degenerate cases, the efficient frontier is concave, increasing and consists of unique portfolios, this portfolio cannot be written as a weighted average of other near-optimal portfolios. Therefore, it is a natural starting point for the construction of the near-optimal region. In Step 4, we find a portfolio w_2 in the region R that differs most from w_1 , ie w_2 satisfies:

$$w_2 = \arg \max_{w \in R(\mu_{\min}, \sigma_{\max}^2)} \|w_1 - w\|, \quad (5)$$

where $\|\cdot\|$ indicates the Euclidean norm, ie the root of the sum of the components squared. Note that maximising the Euclidean norm favours large deviations in one component over small deviations in several components. More generally, as follows from Lemma A1, once we find $i-1$ near-optimal portfolios w_1, \dots, w_{i-1} , all portfolios in the convex hull $\text{Conv}(w_1, \dots, w_{i-1})$, ie all weighted averages, are near-optimal. Therefore, in Step 5, we find w_i by finding the portfolio that differs most from all the portfolios in the convex hull of w_1, \dots, w_{i-1} :

$$w_i = \arg \max_{w \in R(\mu_{\min}, \sigma_{\max}^2)} d(w, \text{Conv}(w_1, \dots, w_{i-1})), \quad (6)$$

where the function d indicates the distance to the convex hull:

$$d(w_i, \text{Conv}(w_1, \dots, w_{i-1})) = \min_{w \in \text{Conv}(w_1, \dots, w_{i-1})} \|w_i - w\|. \quad (7)$$

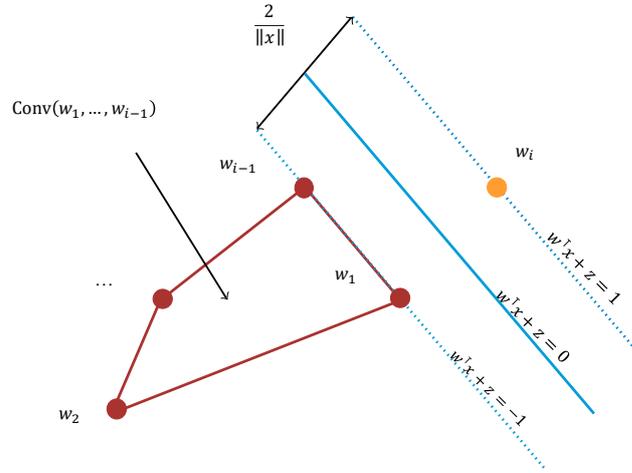
As indicated in Step 6, we continue constructing near-optimal portfolios until the constructed convex hull $\text{Conv}(w_1, \dots, w_n)$ covers the near-optimal region $R(\mu_{\min}, \sigma_{\max}^2)$ up to a required precision $\varepsilon > 0$, ie until:

$$d(w_n, \text{Conv}(w_1, \dots, w_{n-1})) < \varepsilon. \quad (8)$$

Although we cannot say a priori to what extent a certain choice of ε will capture the entire near-optimal region, criterion (8) enforces that there are no near-optimal portfolios that have allocation weights that differ more than ε with the nearest portfolio in the convex hull of w_1, \dots, w_n . Therefore, in practice, we recommend choosing ε equal to the absolute difference in allocation that is considered insignificant.

3.2 Support vector machines

Although Section 3.1 specifies how to construct the near-optimal region, optimisation problem (6) is difficult to solve since evaluating its objective requires calculating the distance of a portfolio to a convex hull which in turn requires solving optimisation problem (7). Optimisation problem (6) can be simplified by applying theory on support vector machine (SVM) classification developed since Vapnik (1963).



The blue line separates the green portfolios from the orange portfolio and has maximal separation margin $2/\|x\|$.

In its easiest form, SVM classification is a machine learning method that tries to separate two classes of points in space by a plane (Boser et al (1992)). Figure 2 shows a two-dimensional example where the green points, representing portfolios w_1 to w_{i-1} , are separated in space from the orange point, representing the portfolio w_i , by the line $w^T x + z = 0$; here x is a vector and z is a scalar.

In SVM classification, the separating line $w^T x + z = 0$ maximises its distance to the two classes, which, as indicated in Figure 2, can be shown to equal $1/\|x\|$. As shown in Boser et al (1992), the separating line of an SVM classification problem can, when it exists, be found by solving a quadratic programming problem:

$$\min_{x,z} \|x\|^2, \quad (9a)$$

$$x^T w_j + z \leq -1 \text{ for } j = 1, \dots, i-1, \quad (9b)$$

$$x^T w_i + z \geq 1. \quad (9c)$$

As shown in Bennett and Bredensteiner (2000), Bennett and Campbell (2000), and Mavroforakis and Theodoridis (2006), the separation margin $2/\|x\|$ equals the distance between the convex hull $\text{Conv}(w_1, \dots, w_{i-1})$ and the point w_i as defined by (7). Optimisation problem (6) can therefore be written as:

$$\min_{x,z,w_i} \|x\|^2, \quad (10a)$$

$$x^T w_j + z \leq -1 \text{ for } j = 1, \dots, i-1 \quad (10b)$$

$$x^T w_i + z \geq 1. \quad (10c)$$

$$A w_i = b, \quad (10d)$$

$$G w_i \leq h, \quad (10e)$$

$$w_i^T \mu \geq \mu_{\min}, \quad (10f)$$

$$w_i^T \Sigma w_i \leq \sigma_{\max}^2, \quad (10g)$$

where constraints (10d) to (10g) enforce that w_i is near-optimal, ie w_i is in the region $R(\mu_{\min}, \sigma_{\max}^2)$, and constraints (10b) and (10c) together with objective (10a) enforce that w_i is at maximum distance of the convex hull $\text{Conv}(w_1, \dots, w_{i-1})$.

Figure 3: Mean-variance efficient frontier with a near-optimal region

Table 2: Near-optimal portfolios and near-optimal region's centre of mass

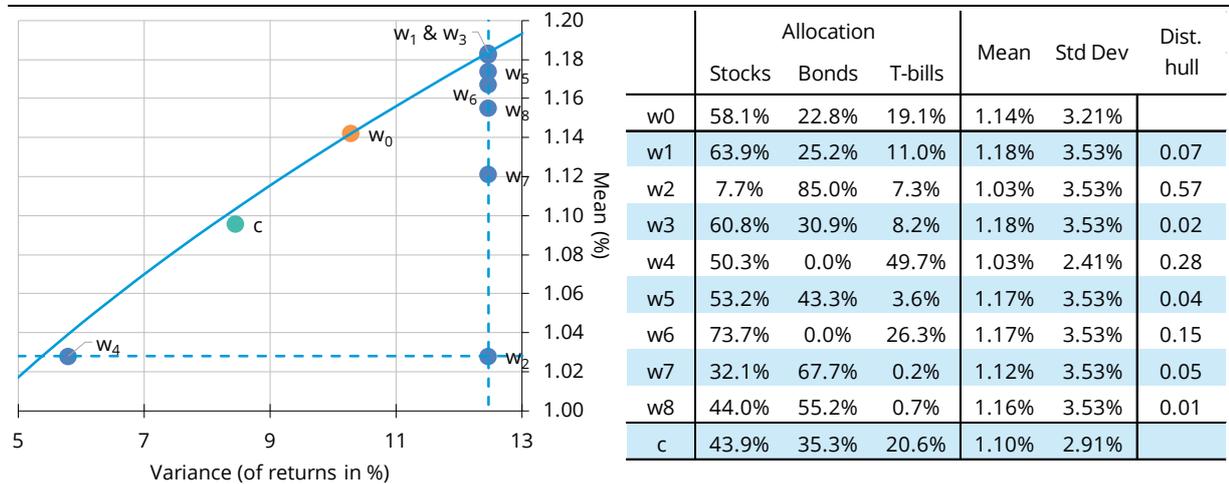


Figure 3 shows, as in Figure 1, the mean-variance efficient frontier and optimal portfolio (orange dot) based on the statistics in Table 1. In addition, the figure shows a near-optimal region (shaded) covered by the convex hull of several near-optimal portfolios (blue dots). The green dot represents the convex hull's centre of mass. Table 2 lists the allocation weights and statistics of the optimal portfolio w_0 , the near-optimal portfolios and centre of mass portfolio c . The last column indicates the distance to the convex hull of the preceding near-optimal portfolios.

Although optimisation problem (10) is non-linear and not even convex due to inequality constraint (10c), it is solvable with standard optimisation software such as the NLOPT's SLSQP solver (Johnson 2010) once a good starting point for the algorithm is found. Finding a good starting point is not straightforward. Imagine, for example, the near-optimal region as a circle and the convex hull approximating it as a triangle connecting three points on the boundary of the circle. The main difficulty is that the complement of the triangle consists of three disjoint parts of the circle. To find a good starting point, we use the so-called hit-and-run (Smith 1984) and shake-and-bake (Boender et al (1991)) algorithms to sample points uniformly from the boundary of the near-optimal region. In practice, a large enough sample will generate a starting point for the optimisation that is far enough away from the convex hull if it exists.

4. Example

Discussion

In this section, we continue with the Chopra (1993) example as shown schematically in Figure 1. Under a no short selling constraint, we perform a mean-variance optimisation using the statistics in Table 1. This results in the efficient frontier (blue line) and the optimal portfolio w_0 (orange dot) indicated in Figure 3. The near-optimal region, indicated by the shaded region in Figures 1 and 3, consists of portfolios that,

compared to the optimal portfolio (orange dot), have an expected return that is relatively at most 10% lower and a standard deviation is relatively at most 10% higher. To find the near-optimal region, we apply the optimisation procedure presented in Section 3 and obtain portfolios w_1 to w_8 . The procedure continues until no portfolio can be found with an allocation difference larger than 1% to the convex hull found so far (see also the last column in Table 2). So, the convex hull of the portfolios w_1, \dots, w_8 covers the near-optimal region up to a precision $\varepsilon = 0.01$, see equation (8).

Chopra (1993) studies (almost) the same near-optimal region with the purpose to show that near-optimal portfolios can have completely different weights. There are two important differences to note. First, for the purpose in Chopra (1993), a different definition of the near-optimal region is used: portfolios that, compared to the optimal portfolio, provide at least 90% of the expected return for less than 90% of the standard deviation are left out, ie portfolios in the lower-left corner of the near-optimal region indicated in Figure 3 are left out. For our purpose, however, we do consider these portfolios near-optimal because there are portfolios considered near-optimal with the same average return and a higher standard deviation. Regardless of this difference in definition, we verified that, as is implied by the difference, all near-optimal portfolios constructed in Chopra (1993) are contained in the convex hull of w_1, \dots, w_8 .

Second, Chopra (1993) constructs near-optimal portfolios through a grid search, ie tries all possible portfolios, and searches for near-optimal portfolios with the highest upward and downward deviation in one asset class. Because there are three assets, this results in six near-optimal portfolios that, as intended, differ completely in weights. It can, however, be verified that the convex hull of the near-optimal portfolios found in Chopra (1993) does not at all cover the near-optimal region. For example, except for w_4 , all portfolios are near-optimal in the definition of Chopra (1993), but w_1 , w_3 , w_5 and w_8 cannot be written, also not approximately, as a weighted average of the near-optimal portfolios reported in Chopra (1993). This shows that merely constructing portfolios with the highest upward and downward deviation is not sufficient for finding the complete near-optimal region. Additionally, a grid search algorithm is, contrary to the methods presented here, only feasible in low dimensions.

Since optimisation problem (10) is not convex, we performed two consistency checks to ensure that the convex hull of the near-optimal portfolios w_1, \dots, w_8 indeed covers the near-optimal region. First, as noted, we verified that all near-optimal portfolios constructed in Chopra (1993) are contained in the convex hull of w_1, \dots, w_8 . Second, we also verified that all portfolios both on the efficient frontier and in the near-optimal region are contained in the convex hull of w_1, \dots, w_8 . Together, this gives sufficient confidence in the convergence and accuracy of the numerical solvers used.

Finally, note that Figure 3 might wrongly give the impression that the near-optimal portfolios of which the convex hull covers the near-optimal region should lie on the boundary of the region indicated by the shaded region in Figure 3. It can easily be shown that this is not the case. For example, when the near-optimal region is increased to portfolios with a return of at least 1% and a variance of at most 4%, the first near-optimal portfolio found has a 100% allocation to treasury bills and does not lie on the boundary of the region indicated by the shaded region in Figure 3.

Selecting a preferred near-optimal portfolio

Once the near-optimal region is covered by the convex hull of near-optimal portfolio w_1, \dots, w_8 , the investor is free to select a preferred portfolio from this region. Without further information, the convex hull's centre of mass c , indicated by the green dot in Figure 3, can be a good default choice. Geometrically, it can be interpreted as the average portfolio when we would sample portfolios (uniformly) from the convex hull of w_1, \dots, w_8 . To construct the convex hull's centre of mass, we used the Quickhull algorithm (Barber et al (1996)).

The centre of mass's geometric interpretation indicates why it makes a good default choice. Now, there are many sets of portfolios whose convex hull approximates the near-optimal region, and we have merely chosen w_1, \dots, w_8 because they conveniently differ from each other as much as possible. Unless the investor has reasons not to, a preferred portfolio should have the property that it depends on the near-optimal region, but not on the particular choice of portfolios whose convex hull approximates it. Because, loosely speaking, the centre of mass is the average portfolio of a region, similar regions have approximately the same centre of mass. Therefore, the centre of mass has this property. Also, conceptually, it is one of the most simple portfolios with this property. Note that, for example, a simple average of portfolios whose convex hull approximates the near-optimal region does not have this property and does depend on the portfolios chosen.

In Section 4.3, we will show that the centre of mass portfolio is robust in two ways. First, the centre of mass portfolio is less sensitive to changing input parameters than portfolios on the frontier. Second, because the centre of mass portfolio is not on the boundary of the convex hull, it is expected to remain near-optimal with slightly different input parameters.

Although the centre of mass portfolio can be a good default choice, preferably, the investor brings in additional arguments to select a preferred near-optimal portfolio. In light of the sensitivity problem, these arguments should be additional to the risk-return statistics in Table 1, eg selecting the portfolio with the highest sharp ratio would not suffice. As an example, suppose the investor prefers, for whatever reason, not to invest in bonds. It follows from Table 2 that, in that case, any weighted average of portfolio w_4 and w_6 would suffice: such a portfolio is near-optimal and has no allocation to bonds. When the investor currently owns 60% equity and 40% cash, which is to be invested in either stocks or treasury bills, the investor could, for example, choose to leave his current exposure to equity intact and to buy treasury bills with his cash.

In practice, the additional arguments brought in for a selecting a preferred near-optimal portfolio can take many forms. Although some arguments to select a preferred near-optimal portfolio can be quantitative, eg low transaction costs, good performance on another investment horizon or good performance in another risk measure, the methodology especially lends itself to be combined with qualitative arguments such as sustainable investing preferences. The methodology gives the range of possibilities through the near-optimal region, after that, it is up to the investor. Also, the methodology can serve as a decision support tool for investment boards. An advisor can present the range of possibilities, ie the near-optimal portfolios. Then, the investment board can choose from the range of possibilities following a line of reasoning they prefer.

Robustness of the preferred near-optimal portfolio

To investigate the robustness of the near-optimal portfolio method, we view the mean and covariance matrix presented in Table 1, ie the optimisation’s input parameters, as estimated on a sample of size N . To perturb the optimisation’s input, we draw a sample of size N from a normal distribution with mean and covariance as in Table 1 and estimate a perturbed mean and covariance matrix.¹ So, the larger the sample size the closer the perturbed means and covariance matrices are to original ones in Table 1. For each perturbed mean and covariance matrix, we determined the perturbed optimal mean-variance portfolio, near-optimal portfolios, convex hull and centre of mass portfolio.

Figures 4 and 5 show the average robustness using 100 samples for each sample size. In Figure 4, the yellow dots and their trend line show for each sample size the average percentage overlap using 100 samples of size N between the original and perturbed convex hulls. The increase of the yellow trend shows that, with perturbed input parameters, roughly 60% to 90% of the near-optimal portfolios remains near-optimal and, consequently, no new investment advice is required.

Figure 4: Near-optimal region vs mean variance

Figure 5: Near-optimal portfolios

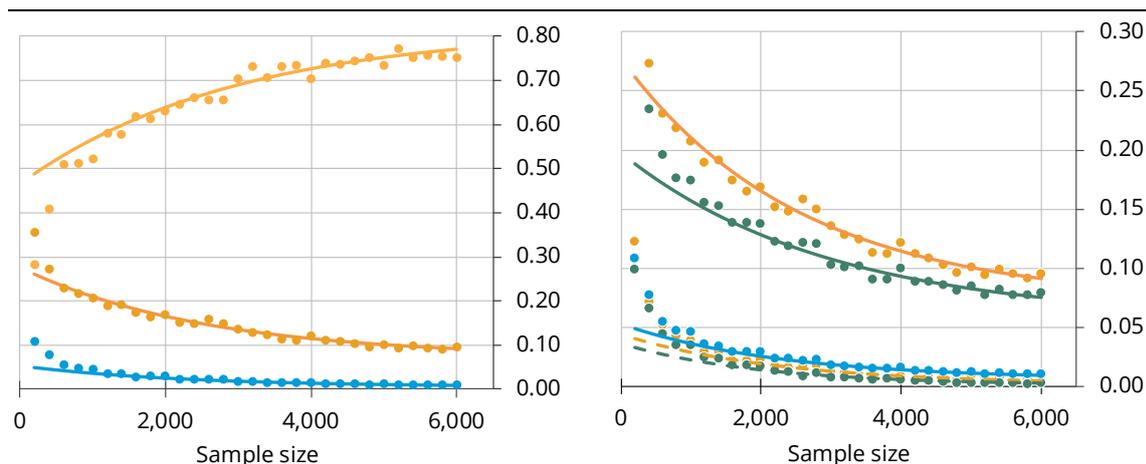


Figure 4 and 5 show, as a function of the sample size N , the average percentage of overlap between the original and perturbed convex hull (yellow), the average turnover between the original and the perturbed mean-variance optimal portfolio (orange) and the average turnover between portfolios in the original convex hull and the closest portfolio in the perturbed convex hull (blue). In addition, Figure 5 shows the original and the perturbed centre of mass (green). And, the average turnovers between the closest portfolio in the perturbed convex hull and: the original convex hull (blue), the original mean-variance optimal portfolio (orange dotted), and the original centre of mass portfolio (green dotted).

The comparison with mean-variance optimisation can best be made using the turnover measure:

$$T(u, v) = \frac{1}{2} \sum_i |u_i - v_i|,$$

¹ Equivalently, we could have drawn perturbed means and covariance matrices from a normal inverse-Wishart distribution, which is the conjugate prior to the multivariate normal distribution.

where the sum is taken over all the components of the vectors. The turnover measure can be interpreted as the fraction of the portfolio u that has to be sold and reinvested to obtain portfolio v . In Figures 4 and 5, the orange dots and their trend line show that the average turnover between the mean-variance optimal and perturbed mean-variance optimal portfolio ranges from roughly 10% to 25%. The average turnover is significantly decreased with the near-optimal portfolio method: the blue dots and their trend line in Figures 4 and 5 indicate that, on average, less than 5% of the portfolio has to be sold and reinvested to obtain a near-optimal portfolio when input parameters are re-estimated.

Figure 5 shows for two typical near-optimal portfolios, the mean-variance optimal portfolio and centre of mass portfolio, that their robustness significantly increases when their robustness is measured with respect to the near-optimal region. At least for these two near-optimal portfolios, the near-optimal method achieves its increase in robustness, because it measures robustness with respect to a region of near-optimal portfolios instead of with respect to a single optimal portfolio. In other words, the increase in robustness is a free advantage when all near-optimal portfolios are considered valid investment decisions.

5. Conclusion

We have shown how to construct a region of near-optimal portfolios below the efficient frontier. Also, we discussed how this methodology is both robust and is designed to support rather than replace the investor's investment decision process.

There are several directions that can be explored in future research. First, the methodology can be applied to more computationally intensive optimisation objectives such as mean-CVaR optimisation. Second, the robustness with respect to estimates for the mean and covariance matrices can be explored separately. Also, the robustness can be compared with other robust optimisation approaches such as resampling. And finally, the method's integration with the investment decision process or use as a decision support tool can be further explored.

6. Acknowledgments

We would like to thank Dr O.W. van Gaans of Leiden University for contributing to this research and supervising two master thesis projects on this topic. Additionally, we would like to thank our colleagues Marnix Engels, John Kuijt and Loranne van Lieshout for their valuable feedback.

A Convexity of the near-optimal region

Lemma A1. The near-optimal region $R(\mu_{\min}, \sigma_{\max}^2)$ consisting of all portfolios that satisfy (2b), (2c), (3) and (4) is convex.

Proof. When given two near-optimal portfolios $u, v \in R(\mu_{\min}, \sigma_{\max}^2)$, we have to show that any weighted average $w = tu + (1-t)v$ is also near-optimal, ie w satisfies (2b), (2c), (3) and (4) for all $0 \leq t \leq 1$. First, since u and v satisfy (2b), it directly follows that w satisfies (2b). Also, inequality constraints (2c) and (3) follow directly. That w satisfies inequality constraints (4) follows from convexity of the left-hand side of (4) and applying Jensen's inequality:

$$w^T \Sigma w \leq t u^T \Sigma u + (1-t) v^T \Sigma v \leq t(w_0^T \Sigma w_0 + \delta_\Sigma) + (1-t)(w_0^T \Sigma w_0 + \delta_\Sigma) = w_0^T \Sigma w_0 + \delta_\Sigma.$$

References

- Barber, C, D Dobkin and H Huhdanpaa (1996): "The quickhull algorithm for convex hulls", *ACM Transactions on Mathematical Software*, vol 22, no 4, December, pp 469-483.
- Ben-Tal, A and A Nemirovski (1998): "Robust convex optimization", *Mathematics of Operations Research*, vol 23, no, 4, pp 769–805.
- Bennett, K and E Bredensteiner (2000): "Duality and geometry in SVM classifiers", *Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000)*, Stanford University, USA, 29 June–2 July, pp 57–64.
- Bennett, K and C Campbell (2000): "Support vector machines: hype or hallelujah?" *SIGKDD Explorations Newsletter*, vol 2, no 2, December, pp 1–13.
- Bertsimas, D, V Gupta and I Paschalidis (2012): "Inverse optimization: a new perspective on the Black-Litterman model", *Operations Research*, vol 60, no 6, pp 1389–403.
- Boender, C, R Caron, J McDonald, A Rinnooy Kan, H Romeijn, R Smith, J Telgen and A Vorst (1991): "Shake-and-bake algorithms for generating uniform points on the boundary of bounded polyhedra", *Operations Research*, vol 39, no 6, November–December, pp 945–54.
- Boser, B, I Guyon and V Vapnik (1992): "A training algorithm for optimal margin classifiers", *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pp 144–152, COLT '92. Pittsburgh, USA.
- Brodie, J, I Daubechies, C De Mol, D Giannone and I Loris (2009): "Sparse and stable Markowitz portfolios", *Proceedings of the National Academy of Sciences*, vol 106, no 30, pp 12267–12272.
- Chopra, V and W Ziemba (1993): "The effect of errors in means, variances, and covariances on optimal portfolio choice", *The Journal of Portfolio Management*, vol 19, no 2, pp 6–11.
- Chopra, V (1993): "Improving optimization", *The Journal of Investing*, vol 2, no 3, pp 51–59.
- Ciliberti, S, I Kondor and M Mézard (2007): "On the feasibility of portfolio optimization under expected shortfall". *Quantitative Finance*, vol 7, no 4, pp 389–396.
- DeMiguel, V, L Garlappi, F Nogales and R Uppal (2009): "A generalized approach to portfolio optimization: improving performance by constraining portfolio norms", *Management Science*, vol 55, no 5, pp 798–812.
- Frankfurter, G, H Phillips and J Seagle (1971): "Portfolio selection: the effects of uncertain means, variances, and covariances", *Journal of Financial and Quantitative Analysis*, vol 6, no 5, pp 1251– 1262. doi: 10.2307/2329859.
- Hurley, W and J Brimberg (2015): "A note on the sensitivity of the strategic asset allocation problem", *Operations Research Perspectives*, vol 2, pp 133–136.
- Johnson, S (2010): *The NLOpt nonlinear-optimization package*.
- Kondor, I, S Pafka and G Nagy (2007): "Noise sensitivity of portfolio selection under various risk measures", *Journal of Banking & Finance*, vol 31, no 5, pp 1545–1573.

Ledoit, O and M Wolf (2004): "Honey, I shrunk the sample covariance matrix". *The Journal of Portfolio Management*, vol 30, no 4, pp 110–119.

Mavroforakis, M and S Theodoridis (2006): "A geometric approach to Support Vector Machine (SVM) classification". *IEEE Transactions on Neural Networks*, vol 17, no 3, pp 671–682.

Michaud, R (1998): *Efficient asset management: a practical guide to stock portfolio management and asset allocation*, Financial Management Association Survey and Synthesis, Boston, HBS Press.

Michaud, R (1989): "The Markowitz optimization enigma: Is 'optimized' optimal?", *Financial Analysts Journal*, vol 45, no 1, pp 31–42.

Reyna, F, A Duarte Jr, B Mendes and O Porto (2005) : "Optimal portfolio structuring in emerging stock markets using robust statistics", *Brazilian Review of Econometrics*, vol 25, no 2, pp 139.

Schöttle, K, R Werner and R Zagst (2010): "Comparison and robustification of Bayes and Black-Litterman models", *Mathematical Methods of Operations Research*, vol 71, no 3, pp 453–75.

Smith, R (1984): "Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions", *Operations Research*, vol 32, no 6, pp 1296–1308.

Vapnik, V (1963): "Pattern recognition using generalized portrait method", *Automation and remote control*, vol 24, pp 774–780.

Copyright © 2019 Ortec Finance bv. All rights reserved. No part of this paper may be reproduced, in any form or by any means, without permission from the authors. Short sections may be quoted without explicit permission provided that full credit is given to the source. The views expressed are those of the individual author(s) and do not necessarily reflect the views of Ortec Finance bv.

Strategic asset allocation from theory to practice: new decision-support tools*

Golan Benita and David Hoffman**

Abstract

The mean-variance (MV) model is widely used in portfolio management, particularly when choosing a strategic asset allocation (SAA) benchmark. But it suffers from several drawbacks, such as sensitivity to small changes in the estimates of mean returns and concentration in a small number of assets. Therefore, when portfolio managers use the MV model as part of their SAA process, they usually do not accept its results “as-is”, but rather make adjustments to them – adjustments which are often based on arbitrary rules or intuition. We describe two decision-support tools that were developed at the Bank of Israel which can help portfolio managers decide whether such adjustments are needed, and examine how they would affect the SAA benchmark’s return in various scenarios. Both tools use a forecast of financial market developments, referred to as the base scenario, to calculate the expected returns provided to the MV model. The first tool identifies a few portfolios which are very similar to the MV optimal portfolio in risk and expected return under the base scenario, but which have quite different exposures to financial market risk factors, including duration, credit spreads and equity returns. The performance of these alternative portfolios can then be compared to that of the MV optimal portfolio under market scenarios different from the base scenario. The second tool allows the portfolio manager to see whether, and by how much, small changes in the base scenario’s financial market forecasts would cause large changes in the makeup of the MV optimal portfolio, and also to see how such differences would affect the portfolio’s return. The information provided by these two decision-support tools gives portfolio managers a more solid quantitative foundation upon which to base their decisions about how to deviate from the MV optimal portfolio in setting their SAA benchmark.

JEL classification: G110, G170.

Keywords: asset allocation, investment decisions, portfolio choice, risk analysis.

* The views expressed in this paper are those of the authors, and do not necessarily reflect the position of the Bank of Israel. This version: October 2019.

** David Hoffman, Markets Department, Bank of Israel, POB 780, Jerusalem 91007, Israel. Phone: 972-2-655-2385. Fax: 972-2-666-9385. Email: david.hoffman@boi.org.il.

1. Introduction

One of the pillars of modern finance is the mean-variance (MV) model developed by Markowitz (1952, 1956). The MV model is widely used in academic research and among practitioners, *inter alia* for selecting an investment portfolio's strategic asset allocation.¹ However, the model has several drawbacks, of which the two most well known are the following: First, the composition of the portfolios located on the efficient frontier is very sensitive to the assumptions made about the distributions of return-on-assets, particularly about their means. Second, imposing a "no short sell" constraint – ie requiring the weight of every asset in the portfolio to be positive – often causes the model to generate an efficient frontier comprised of portfolios that include a very small number of assets.

If the future distribution of return-on-assets were known with certainty, these drawbacks would be less severe. However, in practice, there is usually considerable uncertainty regarding the nature of this distribution – particularly about the means of return-on-assets. The uncertainty about the true mean of the distribution, coupled with the tendency of small changes in means to lead to large changes in portfolio composition, often results in asset allocations concentrated in a small number of assets. This magnifies the potential impact of any error, which means the results of an MV optimisation cannot be implemented as the asset allocation of an investment portfolio without first being subject to critical examination.

There may also be uncertainty about the target level of risk appropriate for a given investor – which often is best expressed as a narrow range, rather than as a precise value – and about the choice of risk measure. Although portfolio risk is often summarised by a single risk measure (such as volatility, the risk measure used by the MV model), its level is usually a function of the portfolio's exposure to a variety of risk factors. For example, central bank foreign exchange reserves portfolios may be exposed to four main risk factors: duration, credit, equity and currency. In such portfolios, very similar combinations of risk and return can sometimes be obtained with different combinations of exposures to these risk factors. For instance, there may be two portfolios with similar levels of expected return and volatility over a one-year horizon, where the volatility of one is mainly due to its allocation to equities, while that of the other is mainly due to duration exposure (ie a higher interest rate risk).

Faced with a choice between the two, an investment manager may prefer the portfolio that is exposed to factors that they believe have *less uncertainty*. In addition, if the manager found these or any other portfolios on the frontier to be concentrated in very few assets, they may go as far as finding an additional *portfolio* that lies *slightly below* the efficient frontier, but has greater diversification benefits. The latter allocation would have a level of risk that is only slightly different from the target, but would place weights on a greater number of assets.

¹ According to the capital asset pricing model (Sharpe (1964), Lintner (1965, 1969)), all investors should select the market portfolio and mix it with a riskless asset according to their risk preferences. However, in practice there is no pure investable riskless asset available to large-scale investors with relatively long investment horizons. For instance, the appropriate riskless asset for an investor who has a one-year investment horizon should be a 12-month Treasury bill. However, large-scale investors cannot practically invest such a high proportion of their portfolio in a single security. Given that there is, in practice, no riskless asset, the optimal portfolio should be selected from the efficient frontier according to the investor's risk preferences.

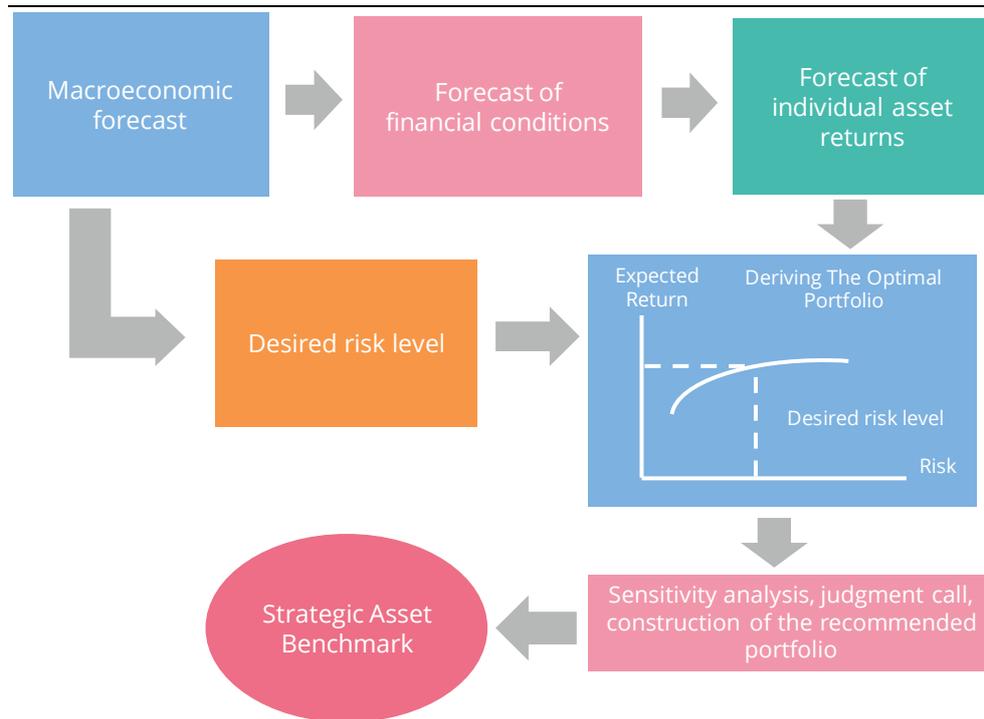
By identifying portfolios using these notions, the investment manager could reduce their exposure to uncertainty about the future distribution of return-on-assets, and may be able to select a strategic asset allocation (SAA) that appears more appropriate than the one obtained by simply and mechanically applying the MV model to solve the problem.

As understood in this paper, the SAA decision is a structured decision-making process, conducted at regular intervals (eg annually or semi-annually). During the process, inputs such as economic and financial forecasts are chosen by the portfolio manager, which are used to select an allocation of funds among asset classes available for investment. This allocation is kept as a *benchmark* portfolio until the next SAA decision is made, and serves as a comparison for the *actual* portfolio of assets held. The portfolio manager may manage against this benchmark by either passively tracking it, or with some latitude allowed to deviate from its composition (in other words, to engage in active management).

A schematic view of one approach to SAA is given in Figure 1. In this example, the process starts with a macroeconomic forecast, which then is mapped both to a forecast of how conditions in the financial markets, eg bond yields, will change over the coming year, and to the choice of a target risk level that the stewards or stakeholders of the portfolio feel is appropriate in the forecasted macroeconomic conditions. The issues involved in formulating such a macroeconomic forecast, and in mapping it into a forecast of financial conditions and an appropriate choice of risk level, are outside the scope of this paper. However, the uncertainty about whether such a forecast will be realised as anticipated, and the difficulty in expressing such uncertainty quantitatively, underlines the importance of taking into account the uncertainty about the means of return-on-assets distributions and about the appropriate level at which to set the target of the risk measure.

Example of a strategic asset allocation process

Figure 1



Source: Bank of Israel.

For a given distribution of asset returns and a target risk level, the MV model can be used to derive the efficient frontier, as shown in the blue rectangle of Figure 1. The composition of the portfolio on the efficient frontier for the target risk level can then be identified as theoretically (*ex-ante*) optimal. However, due to the aforementioned limitations of the MV model, the SAA decision-making process does not normally end there. In other words, the MV model serves as a decision-support tool, not a decision-making tool. In most cases, the portfolio obtained from it will be modified in ways that reduce the impact of its shortcomings on the result – for example, by introducing additional diversification in an ad hoc manner (ie personalising the weight of each asset on the optimal solution). Such modifications may be supported by analyses such as stress testing, but the final call is often made on the basis of the portfolio manager’s intuition or knowledge, which may sometimes be perceived as arbitrary.

To improve the quality of the decisions made in this final step, we have developed two new decision-support tools, which are described in this paper.

1. **The near-optimal analysis (NOA) tool**, which allows the investment manager to compare the MV optimal portfolio with other portfolios that have similar risk and return characteristics but different exposures to three of the four risk factors listed above – ie duration, credit and equity.
2. **The forecast sensitivity analysis (FSA) tool**, which allows the investment manager to analyse the sensitivity of the composition of the *ex-ante* optimal portfolio against changes in the forecasts of risk factors over reasonable ranges. Mapping the optimal allocation over a range of values allows the investment manager to examine how outcomes of the financial variables that are different from those in their forecast would affect the asset allocation chosen for the portfolio.

By allowing the investment manager to explore the trade-off between the various asset classes and different risk-return assumptions, these tools can be of help in selecting an SAA that is close to optimal under given assumptions, while having less exposure than the theoretically optimal portfolio to uncertainty about how accurate those assumptions are.

The approach suggested is not intended to provide a single, unequivocal answer to the question of how to invest, but rather to give the portfolio manager additional information, which will allow the final call to be based less on intuition and more on a solid quantitative foundation.

The remainder of the paper is organised as follows. In Section 2, we describe a simple case study to help illustrate how these two tools are used. In Sections 3 and 4, respectively, we describe how each tool works, and give an example of its results, based on the case study. Section 5 discusses how the results of the two tools together can be used to gain insights that may influence the SAA decision in the case study. Section 6 concludes the paper.

2. The case study

We consider the decision of a portfolio manager who has nine assets available to invest: three US Treasury indices with maturity ranges of 0–1, 1–5 and 1–10 years, respectively; three European domestic government bond indices of the same maturities; an index of investment grade (IG) corporate bonds issued in the United States; an index of IG euro-denominated corporate bonds; and a multicurrency, developed market equity index. The portfolio manager has no desire to expose the portfolio to currencies other than the US dollar and the euro, which entails that a proportion of the equity's currency risk will be hedged as part of the asset allocation decision.

The optimal portfolio must conform to the following constraints: its currency composition must be 65% in US dollars and 35% in euros; the allocation to non-government securities (equity and corporate bonds) cannot exceed 25%; short positions and leverage are not allowed; and finally, the risk of the portfolio, as measured by the volatility (standard deviation) of returns, must fall in the range between 1.95 to 2.05% – this is the portfolio manager's target risk level range. The goal, then, is to find the portfolio with the highest level of return for this level of risk. This is an optimisation problem, which in theory could be solved by applying the traditional Markowitz MV model with restrictions.

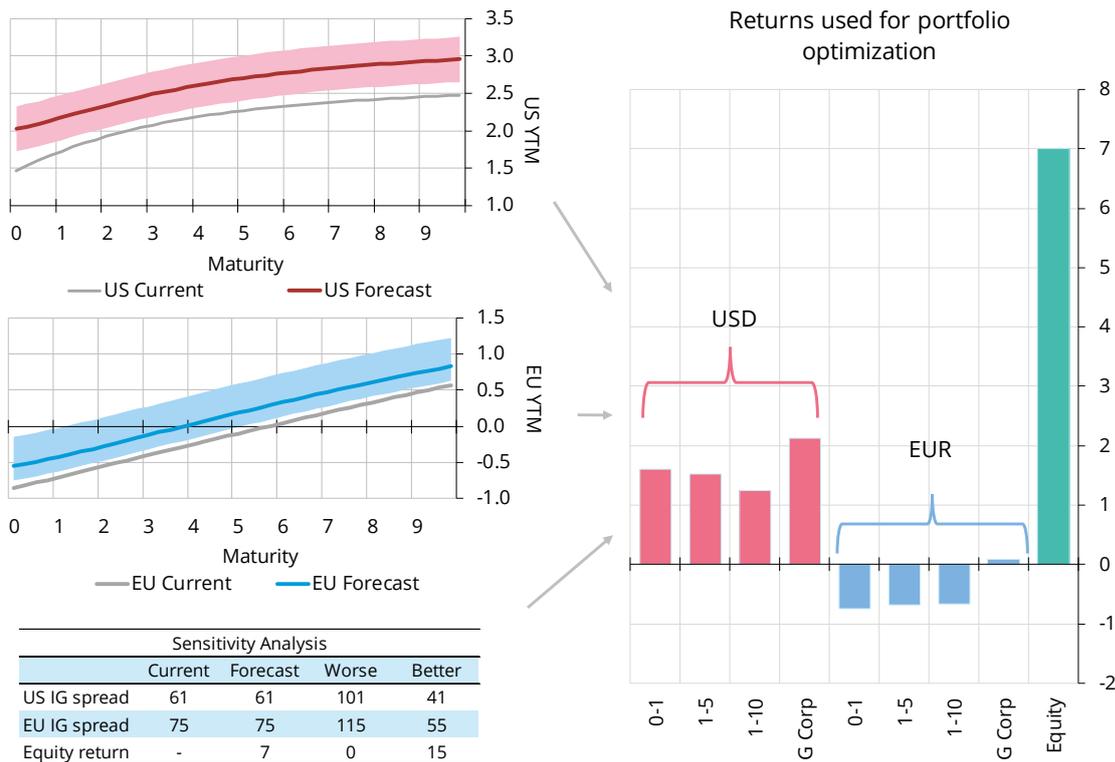
In order to use the Markowitz MV model to select an optimal portfolio, assumptions are required about the parameters of the distribution. For example, to use a multivariate normal distribution to model the returns of the aforementioned assets, assumptions on the vector of expected returns and their variance-covariance matrix are required.²

² The tools we have developed do not address the issues involved in macroeconomic forecasting or the conversion of a macroeconomic forecast into a forecast of financial market outcomes, as represented by the first two boxes in the top row of Figure 1, which, as previously stated, are outside the scope of this paper. Nor do they address the problem of defining the variance-covariance index, which is an exogenous input. However, they do include a process by which a financial forecast can

These elements may be, in practice, derived from other inputs such as (i) forecasts for the government yield curves throughout the projection period specified for the SAA process (in this example, one year ahead), together with their current values; (ii) forecasts for the average credit spreads of the two IG corporate bond indexes that are included in the available assets, together with their current values; and (iii) a forecast for the return of the equity index. Discussing the generation of these inputs, and how these are translated into asset returns is out of the scope of this paper, but the process is illustrated in Figure 2.^{3,4}

Financial market and asset return forecasts

Figure 2



Source: Bank of Israel.

be directly translated into expected returns on specific assets, reflecting the move from the second (purple) box to the third (green) box. For the FSA tool, this ability is not just a convenience, but is integral to its functioning, as will be explained.

- ³ The FSA tool also requires a range of uncertainty around each of the forecasts as part of its input. This is shown by the shaded area around the yield curves in Figure 2, and by the two rightmost columns in the table at the bottom left corner of the figure. As with the forecasts themselves, setting these ranges belongs to the stage of formulating a macroeconomic forecast and translating it into a forecast of financial conditions, which has already been done before the tools we are presenting come into play.
- ⁴ The process of translating financial forecasts into asset return forecasts is computationally intensive for the bond indexes, but simple in its conception. The transition from the initial to the final values of the yield curves and credit spreads is assumed to occur at a constant rate over the course of the year. We also assume that the makeup of the index is the same at the beginning of each month throughout the year. We can then use the forecast yield-to-maturity of each bond, its weight in the index and its duration to calculate monthly holding-period returns of the index, the geometric sum of which gives its annual holding-period return.

By using the returns shown on the right-hand panel of Figure 2 as the expected-returns vector, and supplying a variance-covariance matrix for returns on these assets, it is straightforward to find the efficient frontier of portfolios and identify the MV optimal portfolio at our target level of risk. In what follows, we describe how the NOA and the FSA may aid the decision-maker by improving this process.

3. The near-optimal analysis (NOA) tool

To address the issues that arise when using the traditional MV model, we will identify other portfolios that have similar risk and return to the “efficient” or “optimal” one, but, insofar as possible, different exposures to duration risk, credit risk and equity risk. This is done in four steps:

- (1) Identify the MV optimal portfolio at our target level of risk, and take note of its expected return.
- (2) Identify a large number of portfolios that are below the efficient frontier, but close to it.
- (3) From among the portfolios found in Step 2, select those that have risk and return approximately equal to that of the MV optimal portfolio identified in Step 1.
- (4) From among the subset of nearly optimal portfolios found in Step 3, identify a few that are as different as possible, both from the MV optimal portfolio and from one another, in terms of the three risk factors that are of interest.

Step 1 requires the application of the Markowitz MV model to derive the efficient frontier using the expected-returns vector, the variance-covariance matrix and subject to the constraints imposed. The portfolio that has the required standard deviation – the target risk level – is then identified. In the case study, this was defined as a range between 1.95 and 2.05%. For this first step, we choose a standard deviation at the middle of this range, that is, 2.0%. The expected return of this portfolio is found to be 1.72%.

To carry out Step 2, we repeat Step 1 many times, introducing random fluctuations into the expected return vector each time. To generate such random fluctuations, we define a range where the return of each asset may fall, modelled through a uniform distribution. The range where this uniform distribution is defined (upper and lower bound) can be derived in many ways.⁵ For this case study, we calculate them as 75% and 125% of the base expected return for the lower and upper bound, respectively.⁶ Using a different expected return vector every time allows us to derive a quasi-efficient frontier in every iteration. However, these new portfolios’ expected returns are then calculated using the base expected return for every asset, which, by construction, places them on or below the efficient frontier. This is repeated

⁵ Some other possibilities to consider include (i) ranges based on each asset’s standard deviation of return, of some fixed number of standard deviations above and below the base mean; and (ii) ranges above and below the base mean set in absolute terms by the analyst as a judgement call, either per asset or by groups of assets.

⁶ These base expected returns were those derived from our financial forecasts, which were used to find the original efficient frontier itself.

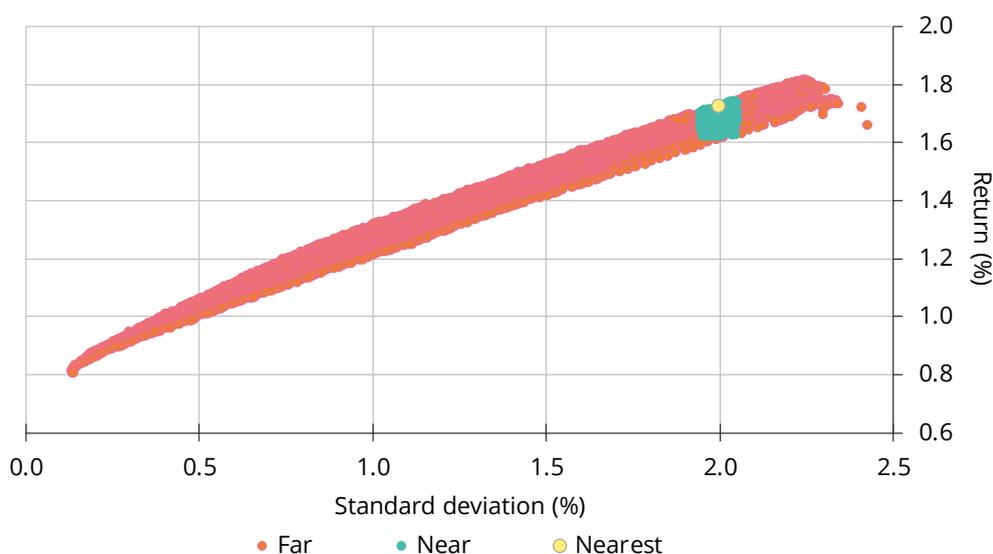
until we have accumulated a number of simulations sufficient for empirical analysis. In this case study, 25,000 sub-optimal portfolios were used.⁷

In Step 3, we define the maximum allowed differences between a portfolio's expected return and standard deviation and those of the MV optimal portfolio. We have earlier specified a standard deviation target in the range of 1.95 to 2.05%. So we look for portfolios with plus/minus five basis points of volatility around the standard deviation of 2.0. For the return, in this case, we choose a lower bound of eight basis points below the return of the MV optimal portfolio found in Step 1, which is assumed to be acceptable by the decision-makers. The 25,000 portfolios found in Step 2 are then examined, and only those meeting the bounds defined around the return and the volatility of the MV optimal portfolio are retained. In our case, this set includes about 1,000 portfolios.

Figure 3 illustrates Steps 1 to 3 graphically. The pink region represents the outcome of Step 2, and the green one, the outcome of Step 3. Note that the actual number of portfolios identified in Step 2 is larger than those plotted in the graph; for greater readability, and due to the limits of the software graphics package that we used, those beyond a certain distance from the efficient frontier are omitted.

Portfolios near the efficient frontier

Figure 3



Source: Bank of Israel.

Step 4 calls for us to find a *small* number of portfolios from among the roughly 1,000 kept in Step 3 that are quite different from one another in terms of the three

⁷ This procedure is similar in spirit to that employed by Michaud and Michaud (2008), but applied somewhat differently. In their model, the goal is to find a near-efficient frontier, in which each portfolio consists of a larger number of assets than there are in the portfolios on the actual efficient frontier. Ours is to identify portfolios that do not necessarily contain more assets than those on the efficient frontier, but do have different risk exposures. For this reason, we introduce a greater degree of randomness, so that a part of the region in the risk-return space bounded above by the efficient frontier is filled with sub-optimal portfolios. Also, we do not average asset allocations across portfolios along the whole length of the efficient frontier, as they do, but rather consider only the final few portfolios found in Step 4 as explicit alternatives, both to the MV optimal portfolio and to each other.

key risk factors that interest us: duration, equity and credit. To do this, we first calculate the exposure of each portfolio to these factors. Each exposure is expressed as follows: portfolio duration in years, exposure to equity in percentage, and exposure to credit in percentage allocated to corporates.⁸

Having thus associated each portfolio with a set of three numbers – which can be thought of as a location vector in three-dimensional space – we next need to calculate the difference, or “distance”, of any one portfolio from another. A natural choice would be the familiar Euclidean distance, found by subtracting one vector from the other, and then squaring the numbers obtained in this manner and adding the squared values together. However, two other distance measures provide results, which we have found more satisfactory.⁹

In the first of these, distance is defined by adding the absolute values of the three values obtained by subtracting one vector from the other. In other words, the distance between points **x** and **y** is calculated as:

$$d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n |x_i - y_i|$$

This measure is sometimes referred to as the “Manhattan distance”, since it represents the distance travelled between two points when one is constrained to follow a grid of equidistant, perpendicular paths, as in a stylised version of the street map of Manhattan.

In the second measure, distance is defined by the formula:

$$d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n \frac{|x_i - y_i|}{|x_i + y_i|}$$

Values of *i* for which the denominator is zero are ignored. This measure is referred to as the “Canberra distance”, for reasons unknown to the authors. Note that the three measures of distance listed above always have positive values.¹⁰

The distances between all possible pairs of portfolios are calculated using whichever one of the distance measures discussed above is selected. The analysis then proceeds as follows:

- a) First, designate the MV optimal portfolio as portfolio 1.
- b) Then, find the portfolio that is farthest from the MV optimal portfolio; call this portfolio 2.
- c) For every other portfolio, add together its distance to portfolio 1 and its distance to portfolio 2. Call the portfolio with the greatest combined distance portfolio 3.
- d) Continuing in the same manner, portfolio 4 will be that for which the sum of the distances to portfolios 1 to 3 is greatest; 5 will be the one for which the sum of

⁸ In principle, these should then be multiplied by factors that equate the profit or loss from a one standard deviation price change for each type of risk. In practice, we have found that using the unmodified values gives satisfactory results.

⁹ This is, of course, a subjective impression with which other practitioners are free to differ.

¹⁰ See Bodenhofer et al (2014) pp 37–42 for further discussion of these and other distance measures.

the distances to portfolios 1–4 is greatest, and so on, for as many alternative portfolios as we wish to consider. We choose four portfolios in our case.

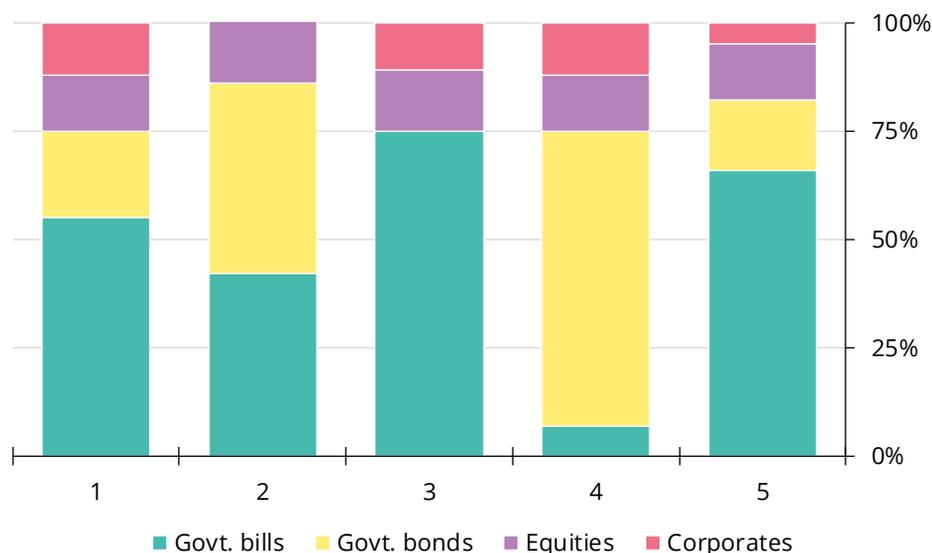
Examining the results of our analysis for the data in our case study (shown in Figure 4) we see that the means and standard deviations of the four alternative portfolios identified are very close to those of the MV optimal portfolio. However, their durations span a broad range, from 0.6 to 2.5 years, and so do their allocations to corporate bonds, which are between 0 and 12%. By contrast, the range of allocations to equity is narrow, from 13 to 15%. From this we can infer that, given the forecasts of asset returns input, there is not any alternative portfolio, having risk and expected return close to those of the MV optimal one, which does not include an allocation to equity within or very near this range.

As an example, it is instructive to compare the two leftmost portfolios (in Figure 4, labeled 1 and 2), which are the MV optimal portfolio and the first alternative identified. If the managers of a portfolio are reasonably comfortable with the quality of their forecasts for equity returns and yield-curve development, but have less confidence in their forecast for corporate credit spreads, they may want to consider a lower allocation to corporates, or perhaps not investing in them at all, while slightly increasing duration and equity exposure to make up the difference in total expected return. The NOA tool's results show that they can do this without deviating greatly from the risk-return profile of the MV optimal portfolio.

It is also possible to examine the performance of the five portfolios that were chosen under alternative scenarios, or to perform stress testing on them. For example, if it turns out that in a reasonably plausible alternative scenario, one of the four alternative portfolios has a slightly positive return, while that of the MV optimal portfolio is significantly negative, an investor who is more sensitive to losses than gains may prefer the alternative portfolio to the original MV optimal portfolio.

Characteristics of case study portfolios identified by the NOA tool

Figure 4



Summary Statistics					
Mean	1.72	1.68	1.71	1.65	1.66
Std Dev	2	2.04	2.04	2.03	1.95
Duration	1.2	1.5	0.6	2.5	1.2
Govt bills	55	42	75	7	66
Govt bonds	20	44	0	68	16
Equities	13	15	14	13	13
Corporates	12	0	11	12	5

Source: Bank of Israel.

4. The forecast sensitivity analysis (FSA) tool

Analysis done in the second step of the NOA tool can provide some insight into how sensitive portfolio allocations using the MV algorithm are to small changes in the vector of expected returns. Nonetheless, the main purpose of the NOA is to offer portfolio alternatives for the decision-maker under the baseline scenario. The FSA tool was developed to more formally explore the sensitivity of the MV algorithm's portfolio allocation to such deviations, within limited ranges. The concern is that small differences in expected returns – which can be interpreted as small deviations of realised financial conditions from the forecast – can lead to a large difference in the composition of the MV optimal portfolio.

Performing the FSA requires that we define bounds of uncertainty for the forecasts of the risk factors. For example, our uncertainty about the direction of yield-curve factors, corporate credit spreads and the return on equity. This is exemplified by the shaded areas above and below each yield curve in Figure 2, and the values labeled "Sensitivity Analysis" in the table at the lower left of the figure.

These bounds reflect estimates about the degree of uncertainty that exists regarding financial market outcomes as implied by the base macroeconomic scenario, rather than the degree of uncertainty on the unconditional realisation of financial market outcomes. In other words, the lack of confidence of the portfolio manager is not on their mapping between macroeconomic and financial variables (which is out of the scope of this paper), but on the realisation of its mean expected path for macroeconomic conditions. As such, the range of possible financial variable outcomes may be narrower than what might be indicated by simply observing the historical (unconditional) distribution of data. It is important to note that it is not necessary for the upper and lower bounds of the outcomes to be symmetric around the mean forecast, nor do they need to be the same range (in percentage points, for example) for similar instruments in different markets.

Intuitively, it is possible to classify the uncertainty about each of the financial market variables that are used to forecast asset returns of this case study as primarily due to one of two risk factors: yield curve risk or business cycle risk. Thus, uncertainties about whether government-bond yield curves will be higher or lower than forecast are classified as yield curve risk. Uncertainties about equity returns and credit spreads are classified as business cycle risk, since business conditions that are better than forecast could lead to a higher equity return and narrower credit spread than expected, while business conditions worse than forecast could lead to the opposite.

The FSA process involves defining a large number of scenarios in a bounded two-dimensional space, in which financial market variables classified as yield curve risk vary in the vertical direction (with higher yields being higher along the vertical axis), while financial market variables classified as business cycle risk vary in the horizontal direction (with higher business conditions being farther to the right along the horizontal axis). Each financial market variable is then mapped to its respective axis, scaled according to the range of its bounds of uncertainty.

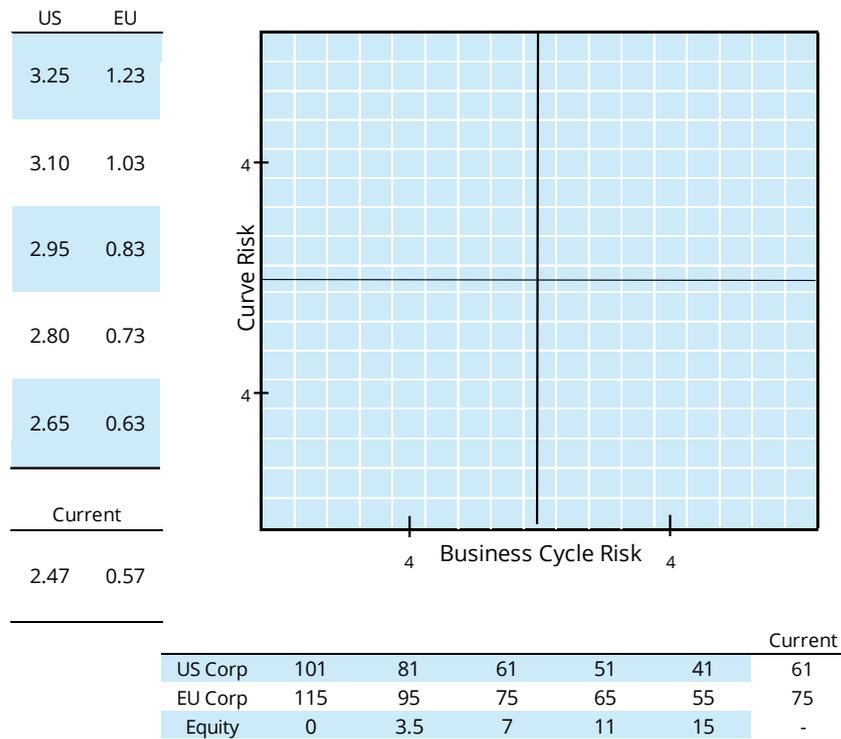
For example, if the range of uncertainty regarding the credit spreads of US IG corporate bonds is defined as 20 basis points (bp) below and 40 bp above the base scenario's forecast, it will be mapped along the horizontal axis from the centre to the left bound as going from 0 to +40 bp, and from the centre to the right bound as going from 0 to -20 bp. In the uppermost of the three horizontal scales in Figure 5, below, this corresponds to the transition from 61 bp (the base forecast) to 101 bp (base + 40) when moving leftwards from the centre line, and from 61 bp to 41 bp (base - 20) when moving rightwards. Implicit in this procedure is the assumption that changes in variables on the same axis (eg US and European credit spreads), move in the same direction and also by the same degree, *relative to the sizes of their bounds of uncertainty in the given direction*.¹¹ It is important to emphasise that, in each alternative scenario, the stepwise forecasts of financial variables over the uncertainty range refer to their ex ante average ("expected") values, *not* their end-of-period realised values.

To allow the creation of a finite number of alternative scenarios, each axis is partitioned into a fixed number of steps above and below the centre, which represents the base case value. The number of steps used can be chosen to partition the space into as many scenarios as the manager feels are useful for the analysis. In our work as practitioners, we have found eight steps above and below the base case (for a total of 17, including the base case) to be most convenient, and this is the

¹¹ This simplifying assumption is necessary for the results of the analysis to be shown in two dimensions.

number used in the case study. Division of each axis into eight steps above and below the base forecast in this manner defines a grid of 288 scenarios, represented by the small squares in Figure 5. The square in the exact centre of Figure 5 represents no deviation from the base forecast for any financial variable, and all others are scenarios in which one or more do deviate from their forecast values.

Grid of scenarios created by the forecast sensitivity analysis tool Figure 5



Source: Bank of Israel.

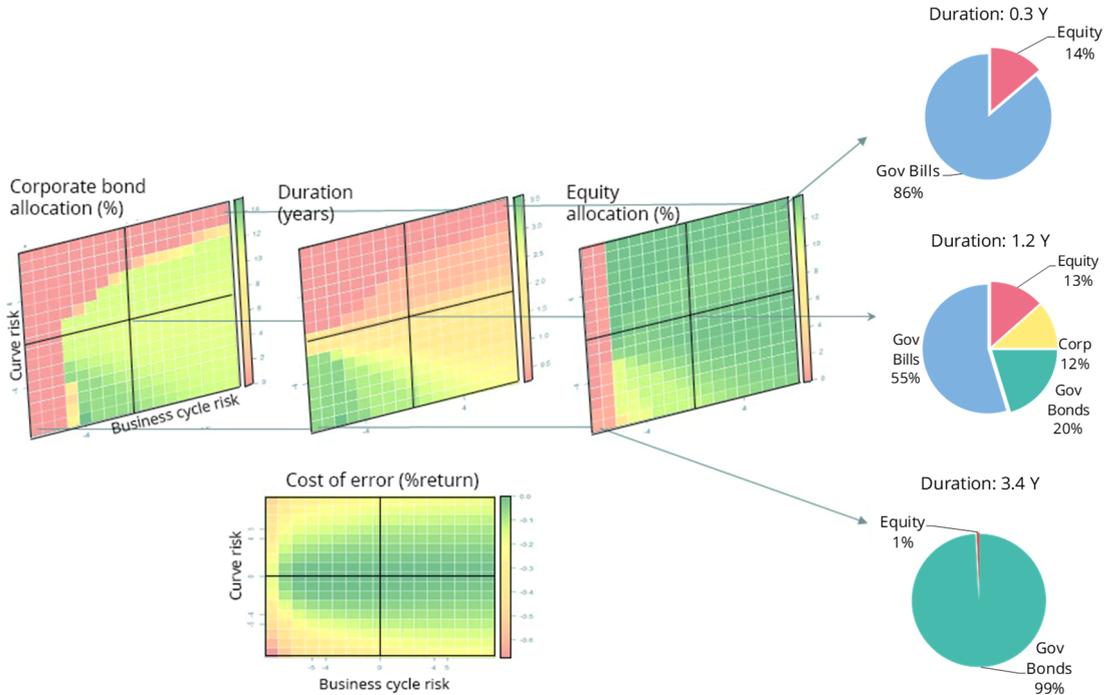
In the scales on the horizontal and vertical axes of the grid in Figure 5, only positions plus 4 and minus 4, out of the range $[-8, +8]$, are marked. However, the tables in the horizontal and vertical margins of the grid give the values of selected financial variables at positions $-8, -4, 0, +4$ and $+8$, going from left to right on the horizontal scale and from bottom to top on the vertical scale, as well as their values at the beginning of the projection period (labeled "Current"). For the business-cycle risk factor on the horizontal axis, all three financial variables are given. For the curve risk factor on the vertical axis, the levels of 10-year yields-to-maturity are given as a proxy for the level of the yield curve as a whole.

For each of the 288 scenarios represented by the small squares in the grid, the assumed values of financial market variables are translated into a vector of asset returns, using the method outlined in Section 2, above. The Markowitz MV model is then used to derive the efficient frontier, given the available assets' characteristics, and subject to the constraints imposed, and the MV optimal portfolio at the target risk level is identified. This MV optimal portfolio, representing a possible asset allocation, is associated with the respective grid square. The process is run 288 times. Finally, we summarise each portfolio by the numerical values of the three risk factors we are interested in – its duration, its allocation to corporate bonds and its allocation to equity – just as was done when using the NOA tool. Rather than using these values

as a position vector with which to define the distance between one portfolio and another, we use them to create heat maps, as shown in Figure 6. A brighter shade of red denotes a lower allocation, while a shade of green denotes a higher allocation to a given risk factor.

Forecast sensitivity analysis results for the case study

Figure 6



Source: Bank of Israel.

Looking at the rightmost of the three heat maps in the upper row of Figure 6, we see that the MV optimal portfolio’s allocation to equity is very similar over a broad range of possible scenarios. Only if the expected return on equity approaches zero does it fall off, and then it does so dramatically. This is consistent with the results from the NOA tool, where we also saw that the allocation to equity is very robust.

On the other hand, even a relatively small rise in yields beyond what was forecast will significantly reduce the duration of the MV optimal portfolio. Its allocation to corporate bonds is influenced by changes in both the curves and the credit spreads. Moving vertically from the centre, we see that up to a certain point, a larger-than-forecast rise in yields does not lead to a reduction in the allocation to corporates because total portfolio duration is reduced by lowering the duration of the government bond portfolio. But, beyond the point at which government bond portfolio duration can no longer be reduced, the allocation to corporates falls off sharply. Moving horizontally from the centre, we see that, in a manner similar to the allocation to equity, so long as spread widening does not pass a certain threshold, the allocation to corporates is not reduced, but once spreads widen beyond that point, there is a sharp transition in optimality.

Finally, the bottom heat map in Figure 6 addresses the following question: If our forecast is incorrect, how much of a loss will that cause, relative to how we would have invested given a correct forecast? In other words, how much worse would the

optimal portfolio implied by the base scenario perform in an alternative scenario, as compared with the portfolio that would be, ex ante, optimal for that scenario? Thus, the middle square of the heat map always has a value of zero, since the base portfolio and the alternative portfolio are the same, while other squares have negative values, as by definition the optimal portfolio from the base scenario will be sub-optimal in an alternative scenario, and so will do less well when compared with the optimal portfolio associated with that scenario. For example, we see that, although even a small addition to our forecast for the change in the level of the yield curve causes the MV model to significantly reduce duration, the effect of this reduction on the difference in returns between the two portfolios is not very large. Having this information might allow portfolio managers to accept a higher duration than they would otherwise have felt comfortable with.

5. Conclusion

The output of the MV model is typically not used “as-is” to make a final decision about the SAA of a portfolio. One reason is because the model’s results can be very sensitive to the assumptions made about the distribution of the return-on-assets and, particularly, about their means. Furthermore, there may also be some uncertainty about the target level of risk, which often can best be expressed as a narrow range rather than as a precise value. To cope with these problems, adjustments to the MV optimal portfolio found by the model are usually made when choosing a strategic asset allocation, but these adjustments have often been based on arbitrary rules or intuition.

In this paper, we have described two decision-support tools that were developed at the Bank of Israel to address these issues, and that are applied in the consultative process by which SAA decisions are made. The two are based on a common software foundation, which allows forecasts over a planning horizon for certain financial market variables – future government yield curves, future corporate credit spreads and the holding period return on an equity portfolio – to be transformed into values of return-on-assets. These values can then be treated as expected returns, and used to identify the MV optimal portfolio at a given risk level by performing Markowitz MV optimisation to derive an efficient frontier.

The tools discussed can be used by portfolio managers to more clearly see whether adjustments of this type are desirable, and what their consequences would be in various scenarios. The first – the NOA – does so by providing the portfolio manager with a number of alternative portfolios, which are very similar to the MV optimal portfolio in terms of risk and expected return, but significantly different in terms of their exposure to various financial market risk factors, including the general level of yields, credit spreads and equity market returns. These alternative portfolios can then be considered, in light of their performance in alternative scenarios and in light of the relative degree of confidence reposed by the portfolio manager in various aspects of the financial markets forecast.

The second tool – the FSA – allows the portfolio manager to see whether, and to what extent, small changes in the financial markets forecast lead to large changes in the makeup of the MV optimal portfolio, and to what extent such differences in the portfolio composition would impact the portfolio’s expected return. With this information in hand, any decisions made by the portfolio manager about adjustments

to the MV optimal portfolio can be made based on a more solid quantitative foundation.

Acknowledgements

The development of the decision-support tools we have described was done in the R statistical computing language (R Development Core Team (2008)). A number of R packages are also incorporated into the software implementing the decision-support tools. Of these, the most important for mathematical computations are **apcluster** (Bodenhofer et al (2011)), which is based on work by Frey and Dueck (2007), and **quadprog** (Weingessel (2013)). Two graphics packages that have roles of major importance in the system are **ggplot2** (Wickham (2016)) and **plot3D** (Soetaert (2017)), which were used to produce most of the figures appearing in this paper. Several other R packages play supporting roles. Each of these packages can be accessed through an URL which follows the pattern <https://CRAN.R-project.org/package=quadprog>.

We wish to express our thanks to colleagues in the Markets Department of the Bank of Israel for useful feedback during the development of the tools we have described, and to the editors for useful comments and suggestions on the paper. The usual disclaimer applies.

References

- Bodenhofer, U, A Kothmeier, and S Hochreiter (2011): "APCluster: an R package for affinity propagation clustering", *Bioinformatics*, vol 27, pp 2463–64.
- Bodenhofer, U, J Palme, C Melkonian, and A Kothmeier (2014): *APCluster: an R package for affinity propagation clustering*, Institute of Bioinformatics, Johannes Kepler University Linz.
- Frey, B and D Dueck (2007): "Clustering by passing messages between data points", *Science*, vol 315, pp 972–7.
- Lintner, J (1965): "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets", *Review of Economics and Statistics*, vol 47, pp 13–37.
- (1969): "The aggregation of investors' diverse judgements and preferences in purely competitive markets", *Journal of Financial and Quantitative Analysis*, vol 4, pp 346–82.
- Markowitz, H (1952): "Portfolio selection", *Journal of Finance*, vol 7, pp 77–91.
- (1956): "The optimization of a quadratic function subject to linear constraints", *Naval Research Logistics Quarterly*, vol 3, pp 111–33.
- Michaud, R and R Michaud (2008): "Estimation error and portfolio optimization: a resampling solution", *Journal of Investment Management*, vol 6, no 1, pp 8–28.
- R Development Core Team (2008): *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- Sharpe, W (1964): "Capital asset prices: a theory of market equilibrium under conditions of risk", *Journal of Finance*, vol 19, pp 425–42.
- Soetaert, K (2017): *plot3D: plotting multi-dimensional data*, R package version 1.1.1.
- Weingessel, A (2013): *quadprog: functions to solve quadratic programming problems*, R port of original S function by B Turlach, R package version 1.5-5.
- Wickham, H (2016): *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag, New York.

Portfolio optimisation problems with hard-to-optimise objective functions

Martin Vesely*

Abstract

The main aim of this paper is to introduce a new algorithm for portfolio optimisation tasks with hard-to-optimise objective functions. The algorithm is based on a heuristic approach called simulated annealing and is suitable for problems with objective functions that are non-differentiable, discontinuous, or have many local extremes (or a combination of all of these features). The performance of the new method is demonstrated on two common problems in quantitative finance: currency portfolio optimisation with a VaR objective function and replication of a benchmark index by its significantly smaller subset of securities. To compare the performance of the new algorithm, several other heuristic methods are introduced. The paper can also be considered as a brief introduction to heuristic optimisation methods for users in economics and finance.

*"Simplicity is power."
Traditional proverb and motto of this paper.*

* Risk Management Department, Czech National Bank; email: martin.vesely@cnb.cz.

1. Introduction and motivation

It is well known that practical problems differ from academic ones, as the real world is complex and contains links that are not always apparent. This is especially true for economic and social issues such as portfolio optimisation.

Even conservative investors such as central banks and international development organisations sometimes have to reconsider their investment strategy as a result of changing financial markets (eg a prolonged period of negative interest rates). A quantitative approach can help decision-makers to better understand markets and project different outcomes of their investment decisions under different scenarios and circumstances. Portfolio optimisation is an inseparable part of this decision-making support, as strategic asset allocation is crucial for long-term capital preservation and income generation. But achieving a realistic optimisation result can be complex, as many variables have to be incorporated into a model.

Mathematical functions describing financial market behaviour without much simplification are not mathematically elegant, especially when statistically and empirically derived relations are incorporated. These functions can be non-linear, non-differentiable or discontinuous. Moreover, functions in optimisation problems can have many local minima, maxima or both. The number of extremes can be so high that a function graph might suggest an “egg holder” rather than a mathematical function¹. If a function has some of these features, it can be impossible to employ some optimisation algorithms for extreme searching, because they usually rely on a good deal of assumptions about the function’s shape and features. Functions describing real world behaviour are hardly ever linear (although sometimes they can be good approximations). Thus, linear programming is often avoided. One can say that this is not a problem since many algorithms for non-linear optimization exist, such as quadratic programming and gradient methods (eg Newton or conjugate gradient methods). However, to assume that the function is quadratic could sometimes be too restrictive as well². Gradient methods can cope with other kinds of non-linearity but they work properly only with functions having at least a first derivative with respect to all variables. This means that in case the function does not have them, one needs to deploy non-derivative methods, eg the simplex method³. While the non-derivate optimisation methods provide means to skirt most of the potential problems so far described, even these methods can have problems with discontinuous functions. Additionally, each method discussed above can be trapped in local extremes of the function. In other words, the method could find some extremes, however, not the global one that had been sought.

The difficulties brought about by the effort to describe market behaviour as precisely as possible can be addressed by the so-called heuristic methods. These are

¹ The so-called “egg holder” function is used for testing an optimisation algorithm’s ability to find extremes. See for example: en.wikipedia.org/wiki/Test_functions_for_optimisation.

² Moreover, quadratic programming guarantees finding of the global extreme only for a subset of quadratic functions (in particular those with a positive definite matrix). Otherwise, the global extreme does not have to be found.

³ This method should not be confused with linear programming simplex. The name for both is the same but the principle and group of problems they can solve are different. For more information on the simplex method see Nelder and Mead (1965).

based on generating random solutions to an optimisation problem. As simplistic as it sounds, random algorithms can be highly sophisticated. Indeed, heuristic algorithms are usually based on physical and biological processes described by mathematical rules but rooted in the probability realm instead of a deterministic one. For example, heuristic algorithms are inspired by processes like a cooling of hot materials (simulated annealing⁴), cosmology (so-called multiverse optimisation⁵ based on the theory that many parallel universes exist simultaneously), the evolution of life-forms (family of so-called genetic algorithms⁶) or the behaviour of predators (eg algorithm mimicking hunting habits of grey wolf packs⁷).

The most important feature of heuristic algorithms is an absence of any special assumptions about the optimised function properties, such as continuity, convexity etc. While the function itself can be mathematically complex, a heuristic algorithm is, in contrast to deterministic methods, still able to find an optimum. On the other hand, it has some disadvantages. Due to its stochastic nature, there is no guarantee that the heuristic method can find a global or even a local optimum. This disadvantage can be addressed by running optimisation many times from the same or different initial points, and then simply picking the best solution. Since generating a huge amount of random numbers can be time-consuming, heuristic methods are generally slower compared with deterministic ones. As powerful computers are more affordable nowadays, one can cope with this drawback easily. A bigger challenge is probably having to choose the right heuristic method for a particular problem because there is no straightforward guideline for this. The best practice is to employ several heuristic methods, such as those inspired by different natural phenomenon, and select the best solution. This paper will introduce some of these methods and present a new one designed purely for portfolio optimisation.

The rest of the paper is organised as follows. First, some optimisation methods are briefly introduced. The new method is then described in detail, followed by a demonstration of the new method on two portfolio optimisation tasks: (i) a minimisation of empirical VaR (ie problem with non-linear objective function) and (ii) an index replication (ie problem with non-linear constraints). Finally, the conclusion is presented.

2. Examples of heuristic optimisation methods

In this section, some examples of optimisation algorithms are presented, beginning first with the simplex method (the only representative of deterministic approach), followed by the simulated annealing technique (which is introduced as a basis for the new method) and, finally, multiverse and grey wolf optimisers.

⁴ See Metropolis et al (1953), Izrailev and Scharf (1990) and Corana et al (1987).

⁵ See Mirjalili et al (2016).

⁶ See Rasheed et al (1997).

⁷ See Mirjalili et al (2014).

Before introducing the algorithms, it is important to understand the term “stop condition”, which is, in short, a criterion that the optimisation algorithm has been completed. There are two types of stop conditions:

1. the maximal number of iterations, ie the number of repeating computational steps of the algorithm, is reached; and
2. the desired precision of the solution is reached

These two types of stop conditions are usually combined. The first one ensures that the algorithm will stop within a finite time to prevent the process from being stuck in an endless loop. The second is used for saving computational time when the required precision is reached.

2.1 Simplex method (Nelder-Mead method)

The simplex method is the only deterministic method presented in this paper, as it does not use derivatives and thus does not rely on assumptions about the shape of the problem objective function. This means that one does not have to sacrifice the accuracy of a model for restrictions imposed by a used method.

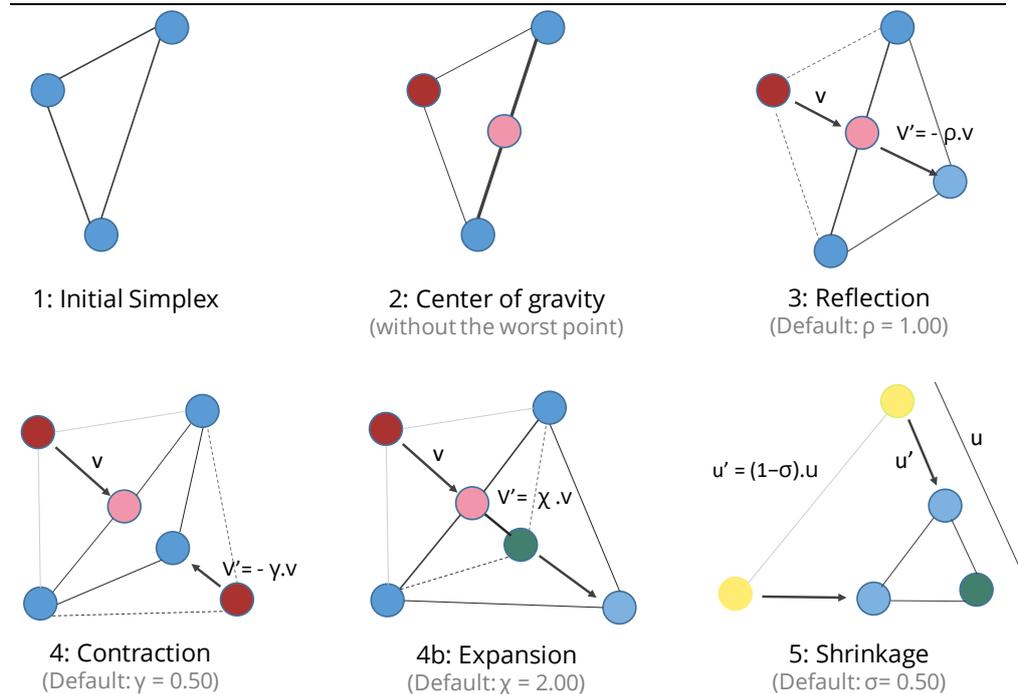
The simplex method uses a geometrical shape called a simplex. In the n dimensional space, the simplex is a regular body with $n+1$ vertices (eg in 1D space, it is a line, in 2D it is an equilateral triangle and in 3D a regular tetrahedron).

An initial solution is specified by the user in the first step of the algorithm. Then a simplex with the initial solution in its geometrical locus is created. In the next step, the objective function values in the simplex vertices are determined. The next step involves choosing a point with the lowest value⁸ of the objective function from the vertices set and the simplex centre. Then, the simplex is modified according to the location of the point with the lowest value in order to surround the point as closely as possible. The simplex can be changed by three operations – reflection, contraction (shrinking) or expansion. These steps are repeated until the stop condition is met. During the process the simplex goes to the minimum of an objective function and reduces itself as illustrated in Figure 1.

⁸ This is assumed that minimising of a function is desired. The highest value is chosen in case of maximisation.

Illustration of the simplex algorithm

Figure 1



Source: <http://capsis.cirad.fr/capsis/documentation/optimisation>, licensed under Creative Commons BY-NC-SA 3.0.

A detailed description of the simplex method (including the formulas for operations shown in Figure 1) is provided in Nelder and Mead (1965).

2.2 Simulated annealing (Metropolis algorithm)

Simulated annealing belongs among the oldest heuristics⁹, with a long record of successful use in sectors including basic research, defense, chemistry, machinery etc.

Simulated annealing is based on simulation of material cooling. In technical terms, it searches for the configuration of particles set with the lowest possible energy at a given temperature. The algorithm was initially used in areas of physics and chemistry before being extended to other fields including business optimisation tasks. However, for portfolio optimisation, this application requires modifications that will be outlined in the third section.

Let us start with the general simulated annealing algorithm. Note that minimising the objective function is assumed. The algorithm has the following steps:

1. An initial solution is established.
2. The objective function value is calculated for the initial solution.
3. The following steps (4 to 7) repeat until the stop condition is met.
4. A "noise operator" is applied to the solution, ie some member(s) of the solution is (are) changed randomly.

⁹ The algorithm was invented by physicist Nicholas Metropolis during his collaboration with Edward Teller, the father of the hydrogen bomb. Therefore, the alternative name "Metropolis algorithm" is often used.

5. The objective function is calculated for the new solution obtained in step 4.
6. In case the objective function value is better than the previous one, the new solution is preserved and the algorithm returns to step 4.
7. In case the objective function is worse than the previous one, the following sub-algorithm is used:
 - 7.1. A uniformly distributed number between zero and one is generated (denoted by x_i)¹⁰.
 - 7.2. Probability p_i of whether the worse solution can be preserved is determined.
 - 7.3. In case $p_i > x_i$ the solution is preserved; otherwise, the solution is abandoned.
 - 7.4. The algorithm returns to step 4.
8. At the stop condition, the best found solution is returned as the final result.

In the first step, the initial solution of the problem is generated. It can be established in many ways, eg it can be a purely random, an expert estimation or a result evaluated by another optimisation algorithm. However, for simplicity, it is sufficient to consider the initial solution as an external input (a method on how to prepare random initial solutions will be shown later). The second step is self explanatory. The third step deals with the stop condition as previously defined. The fourth step (the "noise operator") is aimed towards the random change in the solution of the problem. There is no general approach to noise operator generation. This depends on the nature of the solved problem, eg a random vector with uniformly distributed members is added to the current solution as proposed by Metropolis (1953). The noise operator application merely leads to the solution change; hence it performs the actual process of an optimisation. The fifth and sixth steps are self explanatory.

What can be unusual for readers familiar with deterministic methods is step 7. Its purpose is to avoid getting trapped in a local optimum (so-called "nesting"). The sub-algorithm can be likened to a process "to shake" the solution out of a potential nest. Let us have a look at some subtleties of this sub-algorithm.

The calculation of the probability in step 7.2 can be done in many ways. The one proposed by Metropolis (the so-called Metropolis criterion (Corana et al (1987)) is derived from the laws of thermodynamics. The probability is determined by the formula

$$p_i = \exp\left(-\frac{f_{new} - f_{old}}{kT_i}\right), \quad (1)$$

where f_{new} is the current value of the objective function (ie worse one), f_{old} is the former one, k is the so-called Boltzmann constant and T_i is temperature in current iteration. The Boltzmann constant is a very important number in thermodynamics¹¹, however, for application in finance, the constant k can be an arbitrary positive real number, which can be used for fine-tuning the algorithm performance. Temperature

¹⁰ Note that the new random number is generated in each iteration.

¹¹ The Boltzmann constant plays a role in the relationship between entropy and number of quantum states of particle sets. As a result, it can be found in many formulas in thermodynamics, eg for pressure of gas, energy distribution of particles (so-called Boltzmann distribution) etc. Its value is approximately $1.381 \times 10^{-23} \text{JK}^{-1}$.

T does not have any interpretation in the finance world either, but for historical reasons this parameter will be called so in this paper.

The temperature should decrease during the optimisation process. This dynamic ensures that an increasingly smaller number of solutions with worse objective function values are preserved in the process because decreasing temperature leads to decreasing probability p (as T in formula (1) is approaching zero, p is approaching zero as well). This approach links the method to the material cooling simulation. With decreasing temperature, particles of the material are more stable and do not jump to higher energy levels as often (ie the shaking the solution is not as intensive as at the beginning and it gradually diminishes).

The temperature decreasing process (the so-called "cooling scheme") is described by the following formula (the so-called exponential rule¹²):

$$T_i = T_0 q^{-i}, \quad (2)$$

where q is a fixed number in an interval between zero and one (excluding interval boundaries), T_0 is initial temperature (any positive real number) and i is a natural number increasing during the run of the algorithm. In practical implementation, i is the number of current iteration.

The approach shown above is the original one used for optimisation based on thermodynamics laws. In case of portfolio optimisation, one can adapt formulas (1) and (2) more freely. The best practice is to try to do optimisation with different values of parameters k in (1) and q in (2) or even replace the exponential function in (2) with a linear one (this will be shown in Section 3) and in the end to choose the best algorithm setting.

Because of step 7 and a possible worsening of objective function value, the best solution found so far and its respective objective function value have to be saved in some auxiliary variables and both have to be updated if a better solution is found. This is done in step 5. In the last step (8), the saved value of the objective function is compared with the actual one and a better solution is returned as the final result. Apparently, three variables for solutions and the respective objective function values are used – actual solution, previous solution (these two are updated in each algorithm iteration) and the best solution (this is updated only when better solution is found).

2.3 Multiverse optimisation

Section 2.3 and 2.4 describe two examples of optimisation algorithms developed in this decade. The first method¹³ is inspired by recent discoveries in the field of cosmology, namely multiverse theories¹⁴. These assume that many universes exist at one point in time. The multiverse optimisation considers a set of a problem's possible solutions to be a set of universes. Hence the method operates with more than one solution simultaneously in contrast to simulated annealing.

According to the theory, the universes can be interconnected with white holes (ie an entrance-only gate to a universe) and black holes (ie an exit-only gate from a

¹² See Corana et al (1987).

¹³ See Mirjalili et al (2016).

¹⁴ Plural is used since there are more theories (eg quantum and superstring theories) predicting the presence of more than one universe.

universe). Particular parts of one universe are connected by wormholes (ie shortcuts among different parts of a universe). Solutions are modified by exchanging their specified parts (ie some vector members) through white and black holes. This means that part of a particular solution is replaced by a generally different part of another solution. In other words, one solution accepts the part via a white hole and the other solution donates it via a black hole. Which solution will be the acceptor and which one will be the donor depends on the circumstances in each universe (details are described in the original paper). Wormholes are used for interchanging parts of one particular solution. It is noteworthy (especially for readers already familiar with optimisation heuristics) that the method is similar to the genetic optimisation¹⁵ – the former way of changing a solution (ie white-black holes connection) is akin to a crossing of genes, while the latter one (ie wormhole connection) is similar to a mutation. These two ways of changing a solution can occur with some probability. During the run of the algorithm (ie multiverse aging), the probability that a change occurs through a wormhole is increasing; conversely the probability of a change through black-white holes connection is decreasing. It is an analogy to temperature decrease in simulated annealing. The objective function value for each universe is reevaluated in each iteration and the best one is preserved until a better one is found. The algorithm stops when the stop condition is fulfilled.

2.4 Grey wolf optimisation

The grey wolf optimiser is based on the hunting behaviour of a grey wolf pack¹⁶. It assumes that the pack (ie set of possible solutions) comprises four types of wolves – alpha, beta, delta and omega individuals. While the number of omega wolves is not limited, the other types of wolves have only one representative each. The alpha individual is the head of the pack and thus is considered to be the strongest and smartest wolf. In the optimisation realm, the alpha is identified with the best solution of a problem found so far. Analogically, beta is the second best solution and delta is the third best one. Other solutions are considered to be omegas. During a hunt, the alpha, beta and delta are guides for omega wolves. These wolves adjust their position according to their guides in order to encircle prey. The very same approach is used in optimisation. Omega solutions are slightly changed randomly, but with respect to the three best solutions in all iterations. In other words, a change in each element of a particular omega solution follows both random numbers and the respective elements of alpha, beta and delta solutions. Mathematically speaking, each new omega is a random linear combination of alpha, beta and delta solutions. When omega solutions are updated, the objective function is calculated for them and new alpha, beta and delta and omegas are chosen. The alpha is compared with the best solution obtained so far in previous iterations. In case the new alpha is better, it is preserved as the best solution onward. Steps of algorithm are repeated until the stop condition is fulfilled.

2.5 A few words on algorithm selection

Several heuristics were introduced, however it is not always obvious which method is more suitable for a specific task. One method can easily solve some tasks while

¹⁵ See for example Rasheed et al (1997) for more information on genetic optimisation.

¹⁶ See Mirjalili et al (2014).

completely fail in others – there is no panacea for optimisation tasks. A common practice is to horse-race multiple methods, fine tune their parameters and choose the winners for a hard-to-optimise problem.

3. The new method

In this section, a modified version of the simulated annealing (hereafter “the new method”) for portfolio optimisation is introduced¹⁷.

Methods for initial solution creation and imposing constraints are discussed as well. These approaches are not limited only to portfolio optimisation and may be used for any optimisation.

3.1 Modification of simulated annealing for portfolio optimisation

To solve a particular problem (the portfolio optimisation in this case), one could adapt three parts of the simulated annealing algorithm:

1. Noise operator
2. Temperature decreasing process (cooling scheme)
3. Nesting in local extremes avoidance (shaking the solution)

The modifications of these algorithm’s parts proposed below are designed to enable the incorporation of the following constraints for asset weights into the algorithm directly (ie a user can be sure that these conditions are fulfilled automatically):

1. $w_i \in < 0; 1 >$
2. $\sum_{assets} w_i = 1$

These constraints are commonly used in portfolio optimisation without short positions¹⁸.

To present the design of the algorithm, let us start with the description of the noise operator. Assume that an initial solution fulfilling the above-listed constraints was prepared (a method facilitating this will be described in Section 3.2). Then a random number uniformly distributed between -0.5 and 0.5 is generated¹⁹ – denote it as ϵ . Additionally, assume that w is vector containing n asset weights.

¹⁷ This is the author’s actual contribution.

¹⁸ They are used in tasks dealing with probabilities as well. This fact expands the set of tasks where the proposed method can be used.

¹⁹ There is no particular reason for choosing an interval between -0.5 and 0.5 . Any other interval can be used. The behaviour of the algorithm can be fine-tuned by changing interval boundaries. The proposed value worked well for tasks presented in Sections 4 and 5.

Two weights of assets are chosen randomly (denote them w_i and w_j) and their values are changed according to the following formulas:

$$w_i := w_i + \varepsilon \quad (3)$$

$$w_j := w_j - \varepsilon \quad (4)$$

This modification of weight ensures that constraint no 2 is preserved because one weight is increased by ε , while another one is decreased by the same number. However, the first constraint is not necessarily preserved at the same time (eg if $\varepsilon = 0.4$ and $w_j = 0.1$, then $w_j := 0.1 - 0.4 = -0.3 < 0$). To avoid this constraint breaching, formula (4) has to be rewritten as

$$w_j := \min[\max(w_j - \varepsilon, 0), 1]. \quad (5)$$

A similar modification has to be done for formula (3) too. Thanks to formula (5), the first constraint is fulfilled; however, the second one fulfilled before can be broken now because of weight cutting in formula (5).

Let us therefore introduce a new variable:

$$\Delta = \varepsilon - (w_i^{new} - w_i^{old}), \quad (6)$$

where w_i^{new} is the value of i^{th} asset weight after the modification by formula (5) and w_i^{old} is the same weight before the modification. Hence the difference between these weights expresses how much of ε was "consumed" during i^{th} weight modification. Clearly, Δ stands for the remainder (ie the "unconsumed part") of ε . In the next step, a new weight is selected randomly and modified according to formula (5) but ε is set to be equal to Δ (ie $\varepsilon := \Delta$). Again, a new Δ is determined by formula (6). This cycle runs until Δ reaches zero, ie original ε is fully consumed. Once Δ is equal to zero, the same approach is used for the second weight w_j with ε set back to the original value (ie random number generated before first change of the w_i). While ε was added to the weight w_i , it is subtracted from weight w_j . Note that for weight w_j , the cycle can run differently because generally $w_i \neq w_j$ and there does not have to be any necessity to use the formula (6). For example, for $\varepsilon = 0.4$ and $w_j = 0.5$, $w_j := 0.5 - 0.4 = 0.1$, which is not a forbidden weight value, thus its modification is done in one step. As a result of these modifications in asset weights, both the first and second constraints are fulfilled and the solution is changed as desired.

The second and third modifications of the original algorithm, ie temperature changing and nesting avoidance, are interconnected, therefore they are described together. The temperature decreases in each iteration according to the formula:

$$T_i = 1 - \frac{i}{I_{max}}, \quad (7)$$

where i is the number of current iteration and I_{max} is the maximal number of iterations required. To incorporate the cooling scheme into the algorithm, temperature calculated by formula (7) is used in the noise operator for random number ε modification as follows:

$$\varepsilon := \varepsilon T_i. \quad (8)$$

Clearly, formula (8) ensures that the solution change is getting smaller during the run of the algorithm.

To avoid local minima, temperature decreasing is omitted (ie T_i is set to one) in some iterations. For example, cooling is stopped each tenth iteration²⁰.

The number of iterations (I_{max}) can be changed according to user requirements and/or the nature of the problem solved.

During practical tests, it was observed that exponential cooling (2) and linear cooling (7) led to the same ability of the method to find the optimal solution. Moreover, it was discovered that using the Metropolis criterion (step 7 in the original simulated annealing algorithm) did not have any positive impact on results and can even have a negative one. Besides, removing the criterion leads to a more comprehensible and simple algorithm.

To finalise this section, a pseudo-code of the method is provided below. Numbers at the end of some rows represent links to respective formulas provided in the algorithm description above. Note that step numbering differs from the one used in Section 2.2 as the method is only inspired by simulated annealing.

Pseudocode for the new method:

```

1. n := number of assets in portfolio;
2. Sol_New := create initial solution with n assets;
3. FOR iters = 1 TO maxIter REPEAT
  3.1. Sol_Old := Sol_New;
  3.2. T := 1 - iters/maxIter; (7)
  3.3. epsilon := RAND() - 0.5;
  3.4. IF MOD(iters,10) <> 0 THEN epsilon := epsilon*T; (8)
  3.5. delta := epsilon;
  3.6. FOR j = 1 TO 2 REPEAT
    3.6.1. WHILE delta <> 0 DO
      3.6.1.1. p := RAND_INTEGER(FROM 1 TO n);
      3.6.1.2. x := Sol_New (p);
      3.6.1.3 Sol_New (p) := MAX(MIN(Sol_New(p) + delta,1),0); (5)
      3.6.1.3. delta := delta - (Sol_New(p) - x); (6)
    3.6.2. END WHILE
    3.6.3. delta := -1*epsilon;
  3.7. END FOR
  3.8. IF objectiveFun(Sol_New) > objectiveFun (Sol_Old) THEN
    Sol_New := Sol_Old;
4. END FOR
5. RETURN Sol_New;

```

3.2 Initial solution

The initial solution, maximal number of iteration and objective function formula are inputs provided by the user. While determining the two latter inputs is quite easy, defining the initial solution could be harder. To ensure correct running of the method, the initial solution has to fulfill the asset weight constraints discussed at the beginning of Section 3.1.

One possible way is to set one of the initial vector members equal to one and others to zero. However, this can lead to strange behaviour of the algorithm (such as nesting) or the algorithm can stop far from an optimal solution. Practical tests show that a performance of the method improves with a high degree of randomness

²⁰ There is no exact explanation for chosen number. It is purely a practical fine-tuning of the method and a matter of empirical experience. However, it was observed that this setting provides good results for the optimisation of testing objective functions (eg the Rosenbrock banana function or egg holder function).

among initial solutions. The reason for this is simple; in order to get as many as possible different solutions to pick up the best one, it seems logical to start searching for it from as many as possible different initial points.

The method of producing random initial solutions fulfilling the above-discussed constraints follows. Firstly, n random numbers on an interval between zero and some positive number m are generated. Their distribution is arbitrary, however, it should be bounded on a finite interval (uniform and beta distributions are good examples). These numbers are then normalised by their sum. A sum of rescaled weights is naturally one. Moreover, since generated random numbers are positive and divided by their sum, which is apparently higher than any of them, rescaled numbers lie in the interval between zero and one. Clearly, both constraints imposed on portfolio weights in Section 3.1 are fulfilled.

3.3 Constrained optimisation and penalisation function

Except for the previously discussed constraints imposed on asset weights, the method has been hitherto considered as an unconstrained optimisation tool. However, in practice other constraints are often required (eg a maximum share of a particular issuer in portfolio market value or a target portfolio duration). To take them into account, the objective function has to be modified by adding a penalty function.

Firstly, let us assume that a problem contains a constraint of "less than" type, ie $l(w) \leq L$, where $l(w)$ is a general non-linear real function and L is a real number. Note that the variable w is a vector. Let us consider that the original objective function $f(w)$ is minimised. The modified objective function has the following form:

$$F(w) = f(w) + \lambda \max(l(w) - L; 0). \quad (9)$$

This new function is minimised as well. In case the constraint is fulfilled, the difference $l(w) - L$ is negative, thus the maximum of the difference and zero remains zero and the objective function $f(w)$ is unaffected. Conversely, when the constraint is not fulfilled, a positive number is added to the minimised objective function, which leads to its value increase (the penalty). Parameter λ is set by the user; it expresses the strength of the penalty or is used for scaling.

Analogically, it is possible to deal with a "greater than" constraint $g(w) \geq G$, where $g(w)$ is a non-linear real function and G is a real number. The term added to the objective function has following form:

$$\max(G - g(w); 0). \quad (10)$$

Equal type constraints, ie $e(w) = E$, where $e(w)$ is a non-linear real function and E is a real number, can be described by the term:

$$\text{abs}(E - e(w)). \quad (11)$$

Hence the objective function enriched with the penalty function has the following shape (combination of (9), (10) and (11)):

$$F(w) = f(w) + \sum_{i=1}^{n_1} \lambda_i \max(l_i(w) - L_i; 0) + \sum_{i=1}^{n_2} \tilde{\lambda}_i \max(G_i - g_i(w); 0) + \sum_{i=1}^{n_3} \ddot{\lambda}_i \text{abs}(e_i(w) - E_i; 0). \quad (12)$$

4. Practical task no 1: portfolio optimisation with VaR objective function

In this section, the method introduced in Section 3.1 is tested on a practical portfolio optimisation task – optimal portfolio currency composition.

4.1 The task

Fifteen currencies²¹ and gold were included in the optimisation and their percentage weights were searched for in order to minimise VaR of the portfolio. Historical data and the Monte Carlo simulation were used for the preparation of 1,000 scenarios with different annual returns of each currency and gold against the Czech Koruna (CZK). It is worth noting that these data are an external input and their preparation is not connected with the method for portfolio optimization, thus details on the simulation and capital market assumptions are not provided.

Based on these scenarios and on weights for all currencies in the portfolio, empirical VaR can be calculated. Let us denote \mathbf{A} a matrix containing annual returns in all scenarios (rows) and for all currencies (columns). In this setup, the matrix element a_{ij} represents the return of j^{th} currency against CZK in i^{th} scenario. Let us denote w a column vector of weights, ie j^{th} vector member is weight of j^{th} currency. It is clear that the result of matrix multiplication $\mathbf{A}w$ will be column vector of portfolio expected returns for particular weights in all scenarios. When α percentile of vector $\mathbf{A}w$ members is calculated, one gets $(1 - \alpha)100\%$ empirical VaR of the portfolio for particular weights w .

The goal of the task is to find such weights w that would result in the lowest possible loss, ie to minimise objective function

$$f(w) = \text{VaR}_{(1-\alpha)100\%}(\mathbf{A}w) = -\text{percentile}_{\alpha 100\%}(\mathbf{A}w). \quad (13)$$

4.2 Issues with current optimisation methods

Multiple attempts to run this optimisation using the MS Excel Solver failed to deliver usable results. Apparently this tool is not able to minimise function (13) with the constraints discussed in Section 3.1. Any possible change in Solver settings did not improve its performance. Irrespective of the initial solution, the solver simply “sat in place” and was not able to move anywhere. Results were better in the case of using a MATLAB function implementing a non-derivative simplex method, however, this approach led to very different results for different starting solutions. The failure to solve the task using deterministic methods was the principal motivation to employ a heuristic approach instead; eventually it led to the developing of the new method described in Section 3.1.

Let us look at the function (13) properties. Because of using empirical VaR and 16 variables, it is difficult to imagine the shape of the function and decide about the presence of local extremes and/or other problems that would prevent algorithms implemented in MS Excel and MATLAB from working. For the case with three variables, a 3D graph can be drawn. It is true that three variables should yield a 4D

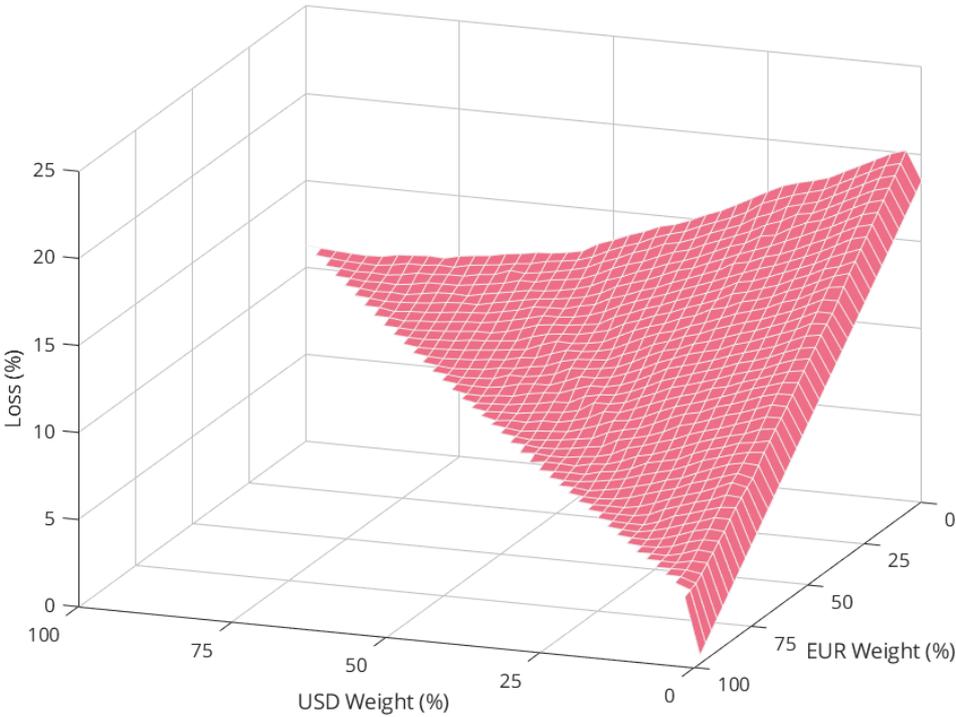
²¹ In particular USD, EUR, GBP, JPY, AUD, CAD, SEK, CHF, NOK, DKK, PLN, KRW, CNY, NZD and SDR.

graph but only two variables can be changed independently, the third one is determined by the fact that the sum of weights has to equal one. A graph of expected loss in portfolio consisting of the euro, US dollar and gold is shown in Figure 2. As can be seen from the graph, the function is continuous, however, one can see a sharp edge on the right-hand side. Moreover, close examination of the graph reveals that the function is similar to a rippled water surface. That means that the function has many local extremes.

To better illustrate this property, a section of the graph with marked local extremes for a portfolio with zero euro weight is shown in Figure 3. Since the ripples are relatively small (ie extremes are shallow), many optimisation algorithms would have a problem with them²². What is more, the function is neither convex nor concave globally. With global convexity being often assumed in deterministic optimisation algorithms, this ambiguity in the function shape is another reason to avoid a deterministic approach and use a heuristic one instead.

Loss function for a portfolio consisting of the euro, US dollar and gold

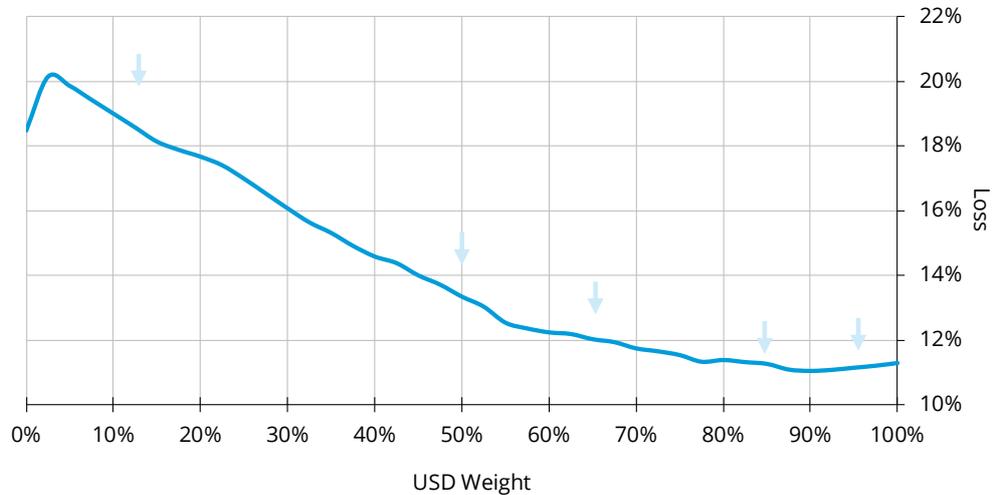
Figure 2



Source: Author's calculations.

²² This issue is similar to the long and slowly decreasing valley in the Rosenbrock banana function (see for example en.wikipedia.org/wiki/Test_functions_for_optimisation).

Percentage



Note: The weight of the euro is set to zero.

Source: Author's calculations.

To conclude this discussion on the function properties; despite the function is seeming relatively uncomplicated at first glance, the issues discussed above disqualify deterministic optimisation algorithms from working even for a relatively simple case of three assets (EUR, USD and gold).

4.3 New method results

The new method introduced in Section 3.1 was compared with several other algorithms. The first one was general simulated annealing²³. Number two and three were a grey wolf optimisation and a multiverse optimisation²⁴. The last algorithm was the simplex method with a random initial solution (randomised simplex²⁵).

None of the general methods discussed in this paper are able to fulfill both above described constraints imposed on asset weights automatically. Therefore, the objective function (13) has to be modified by adding a penalty function.

Hence (13) is replaced by

²³ Function anneal created by Joachim Vandekerckhove in 2008 was downloaded from the MATLAB Central Website (www.mathworks.com/matlabcentral/fileexchange/10548-general-simulated-annealing-algorithm?s_tid=prof_contriblnk) for this purpose. The function can be used freely according to license.

²⁴ Grey wolf optimiser and multiverse optimiser were programmed in MATLAB by author of this paper based on Mirjalili et al (2014) and Mirjalili et al (2016), respectively.

²⁵ Although the simplex method is a deterministic one, it can be converted to semi-heuristic by randomising the starting point. The simplex method itself was implemented in MATLAB function `fminsearch`.

$$f(w) = -\text{percentile}_{\alpha 100\%}(Aw) + \text{abs}\left(1 - \sum_{i=1}^n w_i\right) + \sum_{i=1}^n \max\{-w_i; 0\} + \sum_{i=1}^n \max\{w_i - 1; 0\}, \quad (14)$$

where the first additional term ensures that the sum of weights will equal one, and the second and third ones ensure that weights will be between zero and one, respectively²⁶. It is worth emphasising again that the mentioned constraints are incorporated in the new method by design. Thus (13) can be minimised directly by the new method.

Table 1 compares the performance of the new method with that of others. These statistics are based on 1,000 repetitions of the optimisation launched from different random initial points. The maximal number of iterations was set to 10,000 for all employed methods. It is worth noting that for the purposes of this paper, performance means the ability of an algorithm to approach the objective function global optimum as close as possible; consumption of time and computation power is not of interest.

Comparison of newly developed method with other optimisation algorithms Table 1

Method	Objective function value				
	Best value	Worst value	Average value	Standard deviation	Variation coefficient
New method	0.1101	0.1253	0.1116	0.0012	1.08%
Multiverse optimisation	0.1111	0.1218	0.1143	0.0018	1.57%
Grey wolf optimisation	0.1111	0.1434	0.1165	0.0048	4.12%
Randomised simplex	0.12	0.2545	0.1683	0.0174	10.34%
Simulated annealing	0.1303	0.1858	0.1565	0.0109	6.96%

Source: Author's calculations.

As can be seen in Table 1, the new method delivered the lowest value of the optimised function. However, since the method is heuristic, this does not mean that the global minimum was reached. The value of 0.1101 might still be the local minimum even though the lowest one found so far.

Similarly, the average and variation of optimal value are the best among the compared methods. It is interesting that general simulated annealing, the inspiration for the new method, recorded the worst results measured by the best value of the optimised function. Randomised simplex shows better performance than simulated annealing in terms of the best objective function value, however, it has the highest results fluctuation. Moreover, the average of optimised function values is the worst, too. Both multiverse and grey wolf optimisers reached the same best value of the optimised function but with different variations in results.

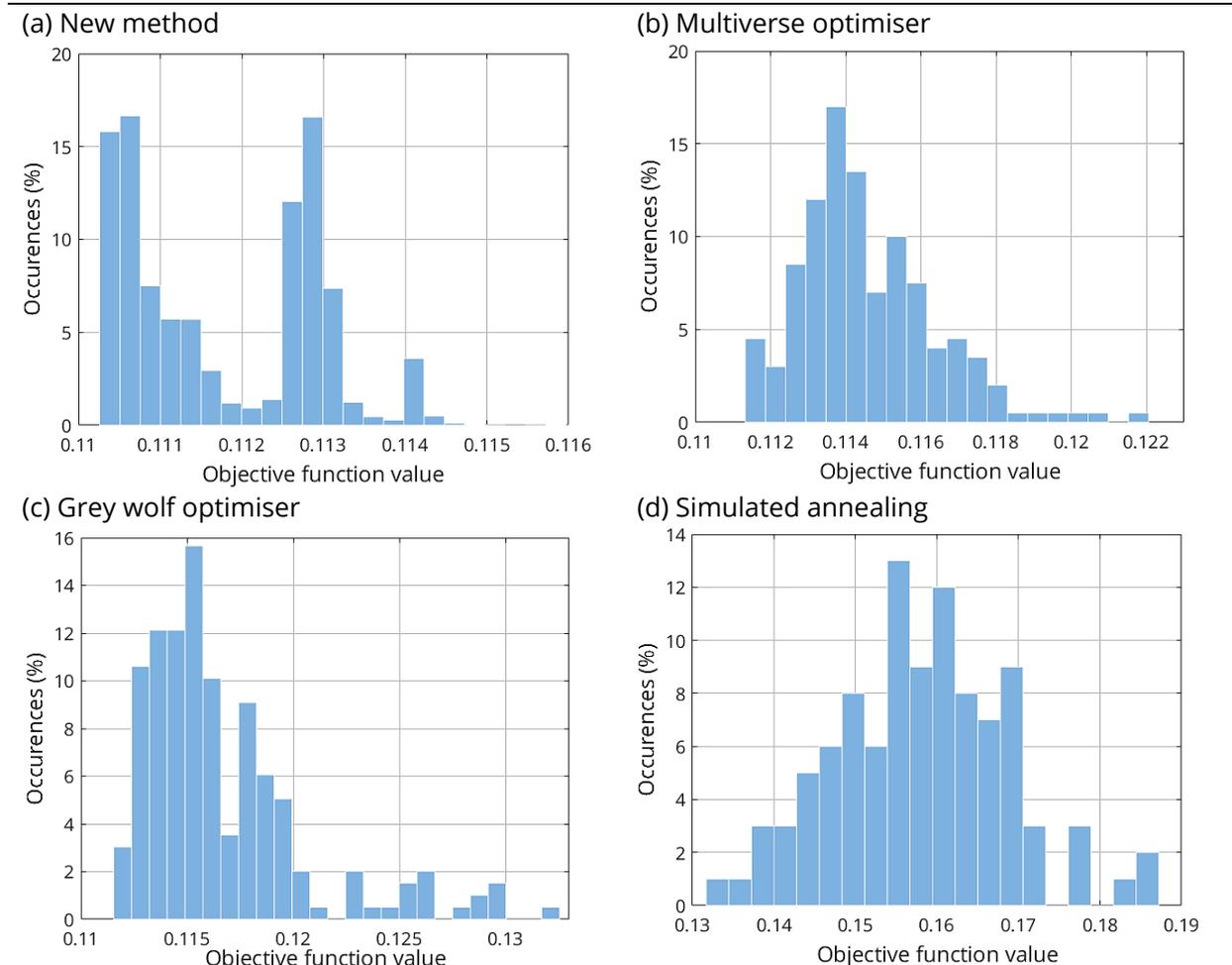
²⁶ The term $\max\{-w_i; 0\}$ can be replaced by $-\min\{w_i; 0\}$ with same effect.

To be more illustrative, fluctuations in results are shown in histograms in Figure 4. As can be seen, the new method repeatedly reached optimal values of around 0.1101 and 0.1130. This behaviour suggests that the optimised function has local minima there. Other algorithms were not able to identify these minima as strongly. General simulated annealing and randomised simplex were even not able to reach such low values. It means that the new algorithm is "strongly attracted" to these local minima. Moreover, histograms also show clearly that the results distribution for the new method is narrower in comparison with others. These observations can be considered as proof of superior performance of the new method in this task.

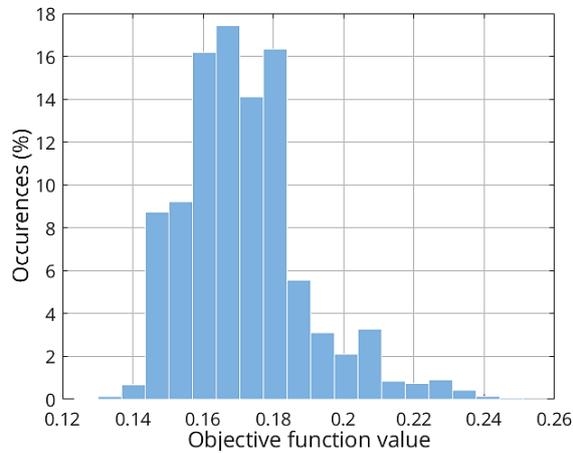
To sum it all up, the proposed method seems to be suitable for portfolio optimisation based on objective functions derived empirically. The method apparently shows better performance in comparison with others based on the statistics presented in Table 1.

Comparison of distributions of optimal values reached by different optimisation algorithms

Figure 4



(e) Simplex



Note: Histogram for the new method is trimmed because of outliers. Values higher than 0.117 were recorded in 0.5% of the cases.

Source: Author's calculations.

5. Practical task no 2: benchmark replication with fixed number of bonds

This section demonstrates the second practical test of the new method. A replication of a bond index by its subset with a significantly lower number of bonds in comparison with the total number of index members is carried out.

5.1 The task

Let us consider a bond index consisting of tens or even hundreds of bonds²⁷. An investor wants to replicate return and risk of this index by a simpler portfolio containing only units or, at maximum, lower tens of bonds. Let us consider that return is expressed by yield and risk is measured by modified duration and convexity. Apparently, only interest rate risk is taken into account. For the sake of simplicity, credit risk is considered to be same for the whole index (ie only one issuer is represented in the index) and there is no FX risk (ie bonds are denominated in one currency). Clearly, this simplification is not far from reality. For example the investor can be a central bank investing its US dollar reserve portfolio into US Treasuries only. So far the task is solvable by a goal programming since it is linear²⁸. However, let us

²⁷ For example, US government bonds indices or covered bonds indices.

²⁸ Goal programming is a special case of linear programming. Instead of minimising or maximising the objective function, it aims to make the difference between actual values of some metrics and their goals as small as possible. Technically, it uses a linear programming simplex method for problem solution finding. Note that here the simplex method for linear constrained problems is meant, not the Nelder-Mead simplex method introduced in Section 2.1.

introduce another constraint; maximal or even an exact number of bonds to be included in the investor's portfolio. Because of this special requirement, the task becomes non-linear with discontinuous constraint and it is not possible to solve it with the MS Excel Solver.

To solve the problem, it has to be translated into mathematical language. Return and risk constraints have the following form:

$$\sum_{i=1}^N y_i w_i = Y \quad (15)$$

$$\sum_{i=1}^N d_i w_i = D \quad (16)$$

$$\sum_{i=1}^N c_i w_i = C, \quad (17)$$

where N is number of bonds in an index; y_i , d_i and c_i are yield, modified duration and convexity of i^{th} bond in the index, respectively; w_i is weight of i^{th} bond in investor's portfolio and finally Y , D and C are yield, modified duration and convexity of the replicated index, respectively.

Equation (15) is a requirement to make portfolio return equal to an index return. Equations (16) and (17) stand for interest rate risk replication²⁹. To fulfil the last requirement, ie to have a specific number of bonds in the investor's portfolio (denote this number n), the following constraint has to be added

$$\sum_{i=1}^N \text{sgn}(w_i) = n. \quad (18)$$

Function $\text{sgn}(x)$ is a so-called sign function. It returns zero when its argument is equal to zero, one for positive argument and -1 for negative argument. Since short positions are not permitted, all weights are by definition non-negative and function $\text{sgn}(x)$ returns a value of zero for bonds not present in the portfolio and a value of one for bonds included. As a result, the sum in (18) represents the number of bonds in the portfolio. It is worth noting that constraint (18) is the only source of non-linearity and discontinuity in the task.

An objective function for the task is the sum of absolute differences between left and right sides of equations (15) - (18):

$$F(w) = \lambda_1 \text{abs} \left(\sum_{i=1}^N \text{sgn}(w_i) - n \right) + \lambda_2 \text{abs} \left(\sum_{i=1}^N y_i w_i - Y \right) + \lambda_3 \text{abs} \left(\sum_{i=1}^N d_i w_i - D \right) + \lambda_4 \text{abs} \left(\sum_{i=1}^N c_i w_i - C \right). \quad (19)$$

The function (19) is minimised.

²⁹ To better replicate yield curve risk of the benchmark, key rate duration could be used as a constraints.

5.2 Solution and results

The index used for demonstration comprises around 200 US Treasuries with maturity between one and 10 years. The number of bonds was set to five. Yield, modified duration and convexity were set to 2.11, 3.69 and 0.2, respectively.

To obtain results, lambda parameters in (19) had to be set. λ_4 was set to 10, other lambdas were equal to one. This setup represents constraints scaling. As can be seen from the values of parameters above, the convexity is in order of tenths, while other parameters in order of units. That is the reason for the ten times higher value of λ_4 . Clearly, lambdas can have different values; however, their ratios should be preserved to reach good results quickly.

Each run of the algorithm can lead to a different portfolio because the task does not have a unique solution and the used heuristic algorithm is probabilistic. Examples of four portfolios compatible with the imposed constraints are presented in Table 2. The table shows selected bonds in each portfolio and compares portfolios parameters with those of the index.

Possible portfolios for the US Treasury index replication

Table 2

Portfolio no 1

#	ISIN	Coupon	Maturity	Yield	Duration	Convexity	Weight		Portfolio	Difference
18	US912810EX29	6.75	15-08-26	2.402	6.776	0.565	11.15	Yield	2.1084	-0.08%
49	US912828A834	2.375	31-12-20	2.028	2.859	0.098	16.88	Duration	3.6902	0.01%
72	US912828J272	2	15-02-25	2.405	6.547	0.483	3.08	Convexity	0.2	0.00%
121	US912828RT95	1.375	30-11-18	1.78	0.912	0.013	22.44			
142	US912828TJ95	1.625	15-08-22	2.206	4.404	0.221	46.45			

Portfolio no 2

#	ISIN	Coupon	Maturity	Yield	Duration	Convexity	Weight		Portfolio	Difference
1	US912810EC81	8.875	15-02-19	1.668	1.065	0.017	16.6	Yield	2.1002	-0.46%
8	US912810EK08	8.125	15-08-21	2.083	3.152	0.124	34.1	Duration	3.6901	0.00%
11	US912810EN47	7.625	15-11-22	2.184	4.179	0.212	22.6	Convexity	0.2	-0.01%
13	US912810EQ77	6.25	15-08-23	2.268	4.775	0.276	18.7			
21	US912810FA17	6.375	15-08-27	2.437	7.475	0.688	8			

Portfolio no 3

#	ISIN	Coupon	Maturity	Yield	Duration	Convexity	Weight		Portfolio	Difference
1	US912810EC81	8.875	15-02-19	1.668	1.065	0.017	10.91	Yield	2.11	0.00%
40	US9128282Y56	2.125	30-09-24	2.384	6.218	0.436	25.93	Duration	3.6897	-0.01%
122	US912828RY80	1.375	31-12-18	1.813	0.989	0.015	6.52	Convexity	0.2	0.01%
134	US912828SX98	1.125	31-05-19	1.858	1.402	0.027	17.53			
206	US912828XR65	1.75	31-05-22	2.214	4.221	0.203	39.11			

Portfolio no 4

#	ISIN	Coupon	Maturity	Yield	Duration	Convexity	Weight		Portfolio	Difference
1	US912810EC81	8.875	15-02-19	1.668	1.065	0.017	15.5	Yield	2.11	0.00%
5	US912810EG95	8.75	15-08-20	1.95	2.342	0.071	17.16	Duration	3.6901	0.00%
12	US912810EP94	7.125	15-02-23	2.241	4.336	0.23	39.16	Convexity	0.1999	-0.03%
58	US912828D564	2.375	15-08-24	2.379	6.051	0.416	22.65			
154	US912828U998	1.25	31-12-18	1.816	0.99	0.015	5.53			

Source: CNB internal benchmark for November 2017; author's calculations.

As can be seen from Table 2, presented portfolios are different. Despite their differences, both modified duration and convexity are replicated well. The yield is lower for portfolio 1 and 2, otherwise the match is perfect.

The number of algorithm iterations was 30,000. A higher number of iterations did not contribute to the better constraints fulfilling while a lower number (around 10,000) sometimes led to breaching convexity constraint. Algorithm coded in MATLAB found the solution on average within 12 seconds on a standard office PC³⁰.

³⁰ Running time fluctuated between 11.59 and 13.63 seconds.

5.3 Some practical aspects

Due to the heuristic nature of the method and rounding errors, on occasion, some bonds in the portfolio had extremely low yet non-zero weights (in the order of hundred-thousandth). Although such values are, of course, economically irrelevant, they play an important technical role in equation (18). Despite their negligible values, these weights are positive numbers, hence the sign function returns a one for all of them. As a result, issues with fulfilling constraint (18) can occur. For example, the final solution fulfils constraints imposed on yield, duration and convexity, however, the number of bonds is significantly higher than the desired one, eg 15 instead of five often came up during tests. To avoid this behaviour, the algorithm proposed in Section 3.1 has to be altered. Firstly, a threshold value for a bond weight has to be stated³¹. In each iteration, weights lower than the threshold are set to zero.

In order to fulfil constraint $\sum_{i=1}^n w_i = 1$, a value defined as

$$E = \sum_{\text{tiny weight}} \bar{w}_i, \quad (20)$$

where \bar{w}_i are former values of weights set to zero, has to be added to any weight higher than the threshold. This weight is chosen randomly. However, constraint $0 \leq w_i \leq 1$ has to be preserved as well, so the value E is spread among many different weights higher than the threshold. It is the same approach as in the cycle in pseudocode step no 3.6.1 of the original algorithm. It is clear, that this modification can be considered as an introduction to the second noise operator.

The method was not compared with other optimising algorithms for this task. It serves only as a demonstration that the method can be used for a quick solution of common tasks an investor can face in practice. Moreover, it shows how to cope with rounding errors and economically insignificant values.

6. Conclusion

The main aim of this paper was to introduce a new heuristic method suitable for portfolio optimisation tasks with hard-to-optimise objective functions. This functions category comprises empirically derived objective functions, eg empirical VaR. Since these functions can lack mathematical elegance, employing heuristic methods could be a good option.

The proposed method is based on a simulated annealing heuristic. The general algorithm was adapted to portfolio without short positions optimisation. Its cooling schema is simpler than in the case of general simulated annealing algorithm, ie linear temperature decreasing is used instead of exponential, and there is no direct connection with thermodynamics. For these reasons, the new method is easily understandable, implementable and usable.

The new method was demonstrated on two practical tasks. The first one dealt with the optimisation of a portfolio currency composition. The objective function was

³¹ A threshold equal to 1% was ascertained as the best possible value for this particular task by trial-and-error procedure. Moreover, a 1% share of bond in portfolio can be considered economically meaningful. However, the threshold depends on the opinion of a decision-maker.

empirical VaR. The method was compared to randomised simplex, general simulated annealing, grey wolf optimisation and multiverse optimisation. It was demonstrated that the new method was able to find the lowest value of VaR. The dispersion of possible optimal results was minimal for the new method as well. Hence the method is more suitable for portfolio optimisation tasks in comparison with other employed heuristics.

The second task focused on the replication of a benchmark bond index with its significantly smaller subset. The biggest advantage of the new method in comparison with the MS Excel Solver is its ability to operate with discontinuous constraint imposed on the number of bonds in the subset. It was also shown how important it is to adapt a general algorithm to specific tasks. This case illustrates that despite the power of heuristic optimisation methods, they cannot be used as black boxes.

The secondary aim of this paper is to serve also as a guide for risk and portfolio managers looking for powerful tools for solving their portfolio optimisation problems.

Acknowledgments

I am very grateful to Daniel Krejci, my former manager, who brought to my attention the practical pitfalls of optimisation tasks with the objective function based on empirical VaR measure and who also read and reviewed an earlier draft of my manuscript.

I would like to express my appreciation to Shengting Pan for invaluable advice on how to improve the paper and make it more accessible. He also suggested simplification of the initial solution preparation method in Section 3.2.

My special thanks are extended to George Alexander Komrower for language and stylistic corrections.

I am also highly indebted to the Czech National Bank for providing me with the technical base necessary for programming and testing the new optimisation method.

All errors or mistakes are of course mine.

References

Corana A, M Marchesi, C Martini and S Ridella (1987): "Minimizing multimodal function of continuous variables with the simulated annealing algorithm", *ACM Transactions in Mathematical Software*, vol 13, no 3: 262–80.

Izrailev F and R Scharf (1990): "Dyson's coulomb gas on circle and intermediate eigenvalue statistics", *Journal of Physics A: Mathematical and General*, vol 23, no 6, pp 963–77.

Metropolis N, A Rosenbluth, M Rosenbluth and A Teller (1953): "Equation of state calculations by fast computing machines", *The Journal of Chemical Physics*, vol 21, no 6, pp 1087–92.

Mirjalili S, S M Mirjalili, A Lewis (2014): "Grey wolf optimizer", *Advances in Engineering Software*, vol 69, pp 46–61.

Mirjalili S, S M Mirjalili and A Hatamlou (2016): "Multi-verse optimizer: a nature-inspired algorithm for global optimization", *Neural Computing and Applications*, vol 27, no 2, pp 495–513.

Nelder J and R Mead (1965): "A simplex method for function minimization", *The Computer Journal*, vol 7, no 4, pp 308–13.

Rasheed K, H Hirsch and A Gelsey (1997): "A genetic algorithm for continuous design space search", *Artificial Intelligence in Engineering*, vol 11, no 3, pp 295–305.

Disclaimer

The author hereby grants a non-exclusive, unrestricted license to use and publish the paper in whatever form, provided that the rights of BIS in the paper remain unfettered. Opinions expressed in this paper are the author's own and do not necessarily represent the opinions and positions of the Czech National Bank (CNB), the central bank of Czech Republic, and Bank for International Settlements (BIS). The author declares that all ideas without citations and links to other articles and materials published in this paper are the author's own. Neither CNB nor BIS nor the author is responsible for any losses (including but not limited to financial and reputational ones) incurred in connection with using the methods described in this paper. The paper is neither an offer to buy or sell any security nor solicitation to do so. The paper does not serve as investment advice.

Part 4

Risk Management

Stealing others' strategies: lessons on model risk

David Jamieson Bolder¹

Abstract

Mathematical and statistical models are used extensively in decision-making across the financial industry. Complex, high-dimensional decisions are generally improved through the addition of mathematical structure and rigour. Models, as a simplified representation of an intricate reality, are nonetheless not without their own risks. Weak or unrealistic assumptions, poor calibration of parameters, erroneous implementation and improper interpretation of results can lead to poor decisions, financial losses and, in extreme cases, systemic shocks. Negative financial outcomes resulting from such events are referred to as model risk. Model risk is essentially a quality control problem. The finance industry is not alone; other areas of endeavour are confronted with similar challenges. Quality control issues arise in scientific discourse, the production of computer software and the construction of comprehensive reference material. Logical, empirical or theoretical errors in these areas are a constant danger. Such a possibility has led to the consequent development of strategies to manage those risks. Risk management professionals also draw numerous lessons from these other disciplines. Examples are peer review, actively seeking multiplicity of perspective, openness regarding technical choices, and independent verification of findings. This perspective provides not only useful motivation for the burgeoning field of model risk management, but similar measures might represent useful additions to the model risk manager's toolkit.

"Truth is the daughter of debate, not of sympathy"
– Gaston Bachelard

¹ Head of Capital and Portfolio Credit Risk, Risk Management Unit, Nordic Investment Bank.

Introduction

Mathematical and statistical models are used extensively in decision-making across the financial industry. They make an appearance in pricing and hedging financial instruments, the computation and attribution of market and credit risk, the estimation of economic capital requirements, the determination of strategic portfolios, and are increasingly employed in the assessment of operational risk. Complex, high-dimensional financial decisions are generally improved through the addition of mathematical structure and rigour.

Models, as a simplified representation of an intricate reality, are nonetheless not without their own risks. Weak or unrealistic assumptions, poor calibration of parameters, erroneous implementation, and improper interpretation of results can lead to poor decisions, damage to an institution's reputation, financial losses and, in extreme cases, to systemic shocks. Negative financial and reputational outcomes resulting from such events are referred to as model risk.

Model risk has, as a concept, been present in the world of finance and economics for the last 20 to 30 years. A famous critique by Lucas (1976) appears to have been something of a starting point, followed by Derman (1996) and, a few years later, by Rebonato (2003). Despite financial crises precipitated at least in part by financial modelling failures – such as the 1987 stock-market crash or the Long-Term Capital Management (LTCM) situation – only recently, as chronicled by Brown et al (2015), in the aftermath of the 2008 Great Financial Crisis, has model risk captured the collective attention of the financial community. OCC (2011) is an example of one of the first regulatory forays into this realm. It introduces the notion of effective challenge as well as some basic governance structures to mitigate model risk. Other comprehensive works, such as those by Morini (2011) and Danielsson et al (2014), have followed in recent years.

While OCC (2011) established – or, at least, initiated – the rules of the game for regulated entities, model risk remains a challenge for non-regulated entities or for model-based analysis performed outside of prescribed regulatory oversight. Many unregulated public institutions such as international organisations, central banks and sovereign wealth funds are still struggling with an appropriate response to this dimension of risk. Even within the regulatory realm, model risk activities appear to follow more the letter than the spirit of the law. Despite a broad-based understanding of the risk posed by poor use and the implementation of financial models, a general framework for its consideration eludes us.

This short paper argues that model risk is, in fact, not a new problem faced solely by financial-market participants. It is closely related to the notion of quality control in scientific discourse, software development, and compilation of reference materials. These disciplines are, as will be argued in the following discussion, fraught with issues closely aligned with the notion of model risk facing the financial community. Logical, empirical or theoretical errors are a constant danger. More importantly, strategies have been developed to minimise this danger and, upon closer inspection, there appear to be important commonalities among these disciplines.

Something can, therefore, be learned from examining the historic response of the academic, software developer, and encyclopaedia writer to these risks. This perspective offers not only ideas for the mitigation of these risks, but also a conceptual framework for their management. This work walks through the pertinent

elements in the development of these strategies. It uses these ideas to construct lessons for modern risk managers, which include the criticality of openness regarding assumptions, techniques, and parameter determination, the necessity of peer review, the essential role of documentation, the role of a healthy degree of scepticism, and the general need for an inclusive, open, and honest discussion of ideas. In a proprietary financial setting, attaining the requisite level of openness regarding modelling decisions can be challenging. Similar constraints, however, do not apply to public institutions. This may suggest slightly different policy responses for these different types of organisations, but the underlying challenges and mitigation techniques remain the same.

A key driver

When one thinks about risk management, it is natural to focus on market, credit, or operational risk. We have a reasonable understanding of the underlying factors driving these types of risk. Market risk is, by and large, well described as the potential for financial loss arising from changes in systematic variables such as interest, exchange rates and general levels of credit spreads. Credit risk stems principally from idiosyncratic or specific factors influencing the ability, or willingness, of credit counterparts to meet their obligations.² Operational risk, in contrast, is somewhat trickier to handle. Hemrit and Arab (2013) argue that operational risk “arises from lack of awareness and deficiencies of skill in detecting threats related to [operational] risk.” While certainly helpful, it lacks the concreteness of market and credit risks.

Model risk finds itself in a similar conundrum. Identification of its underlying drivers is not obvious. It is tempting to identify poor, or potentially dangerous, modelling choices such as linearity, time invariance or Gaussianity. This viewpoint certainly has some merit, but there are cases where such assumptions are entirely justifiable. A model needs to be fit for its purpose and since various models serve varying objectives, clear prescriptive, technical guidelines for model construction have limited value. Restricting our scope to these real, but purely technical, dimensions of model risk feels a bit like confounding the symptoms and sources of an illness.

An alternative logical path provides some insight into this problem. Let us pose the question: what are the optimal conditions for model risk? In other words, when are we most worried about model risk? While a definitive answer will be hard to find, we can nonetheless formulate a reasonable response. Model risk is, potentially at least, greatest when the details of a given model are unclear, unknown, unchallenged, or undocumented. By their very construction, models are approximations. We know a priori that they will often be wrong. It is thus foolish, and dangerous, to talk of a perfect model; such a thing does not exist. The argument is that model risk is most acute when model users have a limited understanding of key model choices and how they might lead the model astray.

² This is, of course, not perfect. Market risk has idiosyncratic elements, while systematic risk also arises in credit risk.



Photo by: Aleksey Gnilenkov / Flickr.

Model risk can thus be considered – only partly facetiously – like a mushroom: ie it grows best in the dark! What is the implication of this conclusion? Models constructed by individuals or small teams, with little external oversight or review, are consequently the most potentially dangerous. A key underlying driver of model risk is limited model knowledge, exposure and understanding. This is broadly consistent with a powerful statement from Morini (2011): “model misunderstanding is the core of model risk.” To abuse the English language somewhat, we might refer to this as “black-boxedness”.

A bit of caution is required. Open, well documented, thoroughly discussed models are, of course, not free of risk. Every player in a modelling process can have a thorough and nuanced appreciation of the key assumptions, but the implementation code could possess serious errors, or the parameters might be improperly estimated. It seems reasonable to conclude, however, that the model owners are better equipped to identify such shortcomings when they have a thorough understanding of the model. Black-boxedness is thus not an exhaustive descriptor of the sources of model risk, but rather an important one.

The point is that models have less risk when they are well understood. This is not a particularly novel statement and it certainly lies implicitly at the bottom of much, if not all, of the extant work on model risk. It is rarely, however, if ever, explicitly identified as a key driver of model risk. Such a statement has value, because as in the market and credit risk settings, identification of the drivers of risk is the first step in the development of strategies for its mitigation.

The main idea

Having identified a key driver of model risk, the next natural step is to seek mitigation strategies.³ Our model risk driver provides, in this regard, a benefit beyond the mere conceptualisation of model risk. It allows us to identify endeavours, or disciplines, that face a similar situation. To be clear, we are thinking of complex, multidimensional tasks – with clearly defined inputs and outputs – typically performed by individuals or small groups. The principal objective of these tasks is to enhance our understanding of some phenomena and take decisions about it. While perhaps not the formal definition of a model, this description gets to the heart of the matter.

What other endeavours follow this pattern? An obvious choice is scientific discourse. Science is principally about developing testable theories to contribute to our understanding of the world around us. On a dramatically less general level, software development also falls into this category. Effective software is certainly complex and multidimensional; moreover, it is often employed to solve problems or take decisions. A final example relates to the construction of comprehensive reference material; an encyclopaedia is a good example. These are three disciplines, of increasing specificity, that have strong qualitative similarities to the construction of financial models.

In these three undertakings – as in model construction – logical, empirical or theoretical errors are a constant danger. A variety of strategies have been employed in the aforementioned activities to mitigate such errors. Some of these solutions are long-standing, whereas others are more recent and innovative. They nonetheless share some important commonalities and, ultimately, model risk management can be inspired from examining their responses. Some of the advice is similar to that found in the current model risk literature, whereas some is, hopefully, somewhat original. In all cases, it appears to provide something of a foundation for discussion of the model risk problem.

In the following sections, the plan is to examine each of these three areas separately and attempt to identify the common aspects of their strategic response to the possibility of important and costly errors. The final section will pull these ideas together and offer something of a loose framework for model risk management. The paper concludes, in the spirit of the underlying thesis, with a careful examination and identification of potential shortcomings of this approach.

Scientific discourse

*“Take nobody’s word for it (Nullius in verba).”
– Motto of the Royal Society*

Let us begin with some brief historical perspective. In 1453, Constantinople fell to the Ottoman empire. This is commonly accepted, among historians, as the beginning of the Renaissance period. Around the same time, in the 1440s, Johannes Gutenberg developed the first functional printing press. These two fundamental elements set the

³ One might argue that risk measurement is a superior starting point, but a bit of reflection suggests that our notion of black-boxedness does not immediately lend itself to direct quantification.

stage for the Enlightenment in the 1600s and 1700s, where science and philosophy took centre stage. Leading thinkers – such as Locke, Leibniz, Descartes, Bacon, and Newton – solidified the scientific method and the notion of empiricism. Thus, in the lead up to the scientific revolution, a modern way of thinking was emerging. New, controversial, and world changing ideas were in circulation. All of these facts are well known.⁴ Let us ask an interesting question related to our model risk problem: how did they manage the quality of these ideas?

By all accounts, a free and ready exchange of ideas was essential. Hatch (1999) indicates that, in the first half of the 1600s, such an exchange of ideas occurred through private societies (ie salons) and correspondence networks. The mechanics are fascinating. In a time of potential church censorship, a clever approach was necessary. The response, it appears, was the so-called republic of letters with intelligencers acting as hubs and de facto editors. Letters, mixing mundane discussion of daily life with logical exposition, were employed to communicate, critique, and develop scientific ideas; this was something of an ink-and-paper precursor to the World Wide Web. In the latter part of the 1600s, there was a gradual movement to state-sponsored academies and printed journals. Both of these approaches were a response to a common problem: how to overcome geographical, cultural and political boundaries to communication.

Letter writing had its limits. In the mid 1600s, as chronicled by Connor and Robertson (2004) and others, a number of intellectuals gathered in London to discuss philosophy. Schonland (1958), among others, refer to this as the Invisible Society. This ultimately led to the establishment of the Royal Society and, within a few years, the first scientific journal entitled *Philosophical Transactions*.⁵ This was an organised and formal response to a priority of claim, a dissemination of ideas and the creation of repository for the accumulation of knowledge.

Spier (2002) indicates that it took about a hundred years before the Royal Society started to internally review publications. In 1831, the first peer-review reports were produced. Nevertheless, it was not until after WWII that double-blind peer review became the standard. Peer review is expensive, time-consuming, and places burdens on the academic community. It appears, however, to be a useful strategy against black-boxedness. That scientific ideas require replication and verification is a key tenet of the scientific method. Interestingly, given commercial and career incentives for scientists, there are ongoing attempts to circumvent these constraints. The Economist (2018a, 2018b) chronicles a trend towards predatory journals that seek to undermine these controls. The quality of these unreferenced journals naturally suffers and the risk of dissemination of misleading or erroneous ideas is heightened. This is an ongoing challenge for the scientific community and speaks to the importance of quality control.

Over the centuries, a number of concrete strategies have evolved to further scientific thought and exercise a degree of quality control. Critical, therefore, to the effective exchange of scientific ideas in academia are:

- discussion and formal presentation of one's ideas to qualified and discerning colleagues;

⁴ See, for example, Wootton (2015) for a detailed description of the scientific revolution.

⁵ It is simultaneously amazing and humbling to look at the list of ground-breaking ideas first published in the pages of this journal.

- publishing one's results in the public domain; and
- subjecting one's analysis to independent peer review.

Scientific thought certainly existed without these concepts, but all of these ideas seek to bring ideas out of the dark. The bright light of constructive criticism helps identify, but not eliminate, weak assumptions and logical inconsistencies.

It is relatively easy to see that all of these elements have links to model development by finance practitioners. The model risk analogue of letter-writing and societies and journals is discussion with colleagues. Publication would take the form of model documentation as well as internal or external publication of results. Peer-review, as already indicated, is quite obviously equivalent to performance of an independent validation exercise; this is, indeed, the idea behind OCC (2011)'s idea of effective challenge. The guiding principle is simple: the greater the audience of competent, discerning eyes, the higher the degree of quality control. This has a Darwinian dimension; those ideas that survive harsh but honest criticism, continue, while those that do not, fail. This is essentially the opposite of black-boxedness.

A potential challenge

"I often compare open source to science. To where science took this whole notion of developing ideas in the open and improving on other peoples' ideas and making it into what science is today and the incredible advances that we have had. And I compare that to witchcraft and alchemy, where openness was something you didn't do."
 – Linus Torvalds

Although the academic model is encouraging and enlightening, it is based on the open exchange of concepts and results. Publicly sharing ideas in the financial realm can have costs. Some models have proprietary value and sharing one's development could lead to free-riding and loss of economic rents. It might even, in the limit, reduce incentives to innovate. Public-policy models, if exposed, might change the behaviour of market participants. A classic example is the Bank of Canada's so-called monetary conditions index – Freedman (1995) is an excellent overview. The publication of specific parameters for this measure undermined, in this case, its effectiveness as a monetary policy tool. Transparency in public policy is a subject of ongoing debate. Economists (see eg Jensen (2002)) have established a trade-off between credibility and flexibility. The pendulum has certainly swung towards transparency over the last few decades, but full transparency does not appear to have been viewed as optimal.

There is an interesting counterexample to this argument hiding in plain sight. Innovation and progress need not necessarily require propriety interests. There are examples of the two living in cooperation within the software industry. Lerner and Tirole (2002) indicate that early operating systems and internet protocols were developed collaboratively in academic or research centres (ie Berkeley, MIT, Bell Labs, Xerox PARC). This involved extensive sharing between institutions. In the early 1980s, a desire to enforce intellectual property rights emerged. Richard Stallman and others, in response, launched the GNU Project and the Free Software Foundation. The open-source software license permits code to be used, modified, or shared by other developers. In this manner, many open-source software products can be viewed as collaboratively developed. The guiding idea behind this movement is thus to maintain the collective spirit found in early development circles.

Despite initial scepticism and even derision, the open approach to software development has proven highly successful. Many commercial entities not only employ open software tools – such as Linux, Apache, Perl, Python, or a broad variety of compilers – in their day-to-day activities, but actively encourage staff to participate in the development process. Against all expectation, open-source programming has proven a success. Von Krogh and von Hippel (2006) actively seek to highlight the utility of aspects of this movement for many other fields in terms of governance, organisation, and innovation.

Software development, like model construction, is a technical and complex undertaking, which can often be broken into smaller, modular tasks. The proprietary model, of course, also works reasonably well in this field. Perhaps surprisingly, the open-source approach works as well or, on occasion, even better. Furthermore, it appears to offer the strongest benefits when users are most sophisticated (ie Apache, Linux). This point makes it particularly interesting for complex financial applications. Key elements involve putting source code in the public domain, sharing of knowledge, and exposure of design to critical review. It also appears to foster creativity and innovation. This concrete precedent suggests that exposing, in a cautious and controlled manner, internal models to a larger group can thus potentially enjoy these benefits.

The power of crowdsourcing

*“You shouldn't use anything as the sole source for anything.”
– Jimmy Wales*

Creating an encyclopaedia is also a technical endeavour, which is certainly prone to error. Reference books incorporate a multitude of facts and are relied upon by professionals of all ilks for important background information. Wikipedia, the online encyclopaedia, was launched in 2001. Content creation, in this set-up, is placed under the control of registered volunteer editors.⁶ At first blush, this does not seem to be a particularly compelling organisational structure. Individual entries are often constructed by multiple individuals with oversight and a control structure involving the entire editing community. Over the last few decades, however, it has become a preeminent source of information.

Popularity is not, it should be stressed, a universal criterion for success. Although there are questions about its accuracy – and a few horror-stories – there is also evidence that it works quite well. Giles (2005) suggests, for the sciences, that it is roughly as accurate as its commercial competitor, Encyclopaedia Britannica. It is also significantly more reactive to new information, events, and ideas. Brown (2011) further holds that it demonstrates surprising accuracy for political science.

A word of clarification is warranted. This work is certainly not advocating Wikipedia as a research platform. Greenstein and Zhu (2012) raise important questions about its biasedness, which is particularly acute with controversial, subjective, or difficult to verify topics. It is also constructed to place equal weight on expert and non-expert contributions, which seems conceptually difficult to defend.

⁶ Interestingly, Richard Stallman was an early proponent of this idea. For more on this point, please refer to Stallman (2000).

These important caveats notwithstanding, it is a successful model. What can we learn to use in model risk management? Once again, openness of ideas, exposure to criticism, and engagement with a broader, discerning community is at play. Wide-ranging access and openness appear to work, in this case, at least as well as small groups of experts. The key lesson, it seems, is the usefulness of multiplicity of perspective. This argues, in the financial setting, for the incorporation of the viewpoints of both technical and non-technical staff into the modelling process.

Beginnings of a framework

We have identified three related, but impressively disparate, complex, technical and error-prone endeavours. Each of these activities possess important qualitative similarities to those involved in the construction of financial models. They cover the spectrum of high-level conceptual to low-level implementation details, which is precisely the scope of complexity found in most financial models. All also face significant quality control issues that can importantly undermine their success. Each of them, even in the face of proprietary interests, have adopted quality control strategies that involve a high degree of openness and exchange of central ideas.

Table 1 highlights a broad range of mitigants aimed at our key driver of model risk: the black box. Each of these elements is inspired from our examination of these conceptually similar disciplines. The scientist's quality control dilemma, in particular, is a rich source of inspiration. Publication, peer review, conferences, and formal degrees all play a role in ensuring a healthy scientific discussion. Model risk managers can draw from this example. Independent model validation is already a key pillar, but more effort can be made to publish methodologies and results and engage with other well trained modellers both inside and outside one's institution.

Organising parallels		Table 1
Parallel		Model risk
Publication		Documentation and white papers
Peer review		Independent model validation
Conferences and societies		Internal and external networks
Courses		Subject-matter training, knowledge transfer
Degrees and diplomas		Professional designations
Openness		Inter-institutional cooperation
		Code sharing and repositories
Crowdsourcing		Multiplicity of perspective

The underlying table attempts to draw a number of practical parallels between extent strategies in scientific discourse, software development and construction of reference material with possible responses to financial model risk.

Knowledge-sharing and transfer are popular concepts that appear, in recent years, to have attained buzzword status in business circles. It is, in the experience of the author, nonetheless quite rare to find model owners formally presenting and defending their work to a broad audience of peers. It would be unthinkable, by contrast, for academic discourse to proceed absent such formal presentation of results. There are certainly costs to both presenters and attendees. Distilling and

digesting complicated ideas is hard work, which requires consumption of scarce resources. It is, however, a key mitigant of model risk, which if we deem it serious, makes this a worthwhile undertaking.

Internal presentation is already a useful step forward, but widening the audience to external parties improves the situation further. The challenge is that, quite often, sharing modelling ideas can potentially create business conflict. Managing this proprietary element is, in general, difficult although this aspect is, for public institutions, somewhat attenuated. The software-development and encyclopaedia-construction examples nonetheless suggest that there is the possibility of overcoming this challenge. The open-software movement indicates that inter-institutional cooperation can work. The financial community could, for example, look to cooperate in a limited manner. Sharing of model code through repositories is another step towards mitigating the black-boxedness of one's internal models; this might be most potentially helpful for complex, non-proprietary modelling sub-tasks.

Even faced with tight proprietary constraints, there is a case for enhanced openness within one's organisation. Too often, in the author's experience, modelling choices are viewed as the technical domain of a select few. Rarely are such decisions questioned, not to mention discussed, in senior management circles. Such openness has, of course, costs for both non-technical senior and technical staff. Senior management must be encouraged, or incentivised, to be more demanding in their consumption of modelling outputs and discerning of their inputs. Indeed, more time should be allocated to key inputs, assumptions, and methods and less to final results. Technical staff, for this to succeed, also need to become better communicators. Taking both of these groups out of their respective comfort zones is entirely consistent with the general thesis of this work.

The Wikipedia case appears, along these lines, to stress the power of multiplicity of perspective. This idea has its origins in psychology. Campbell and Fiske (1959) formally introduced the idea of validating results predicated on having demonstrated it along a variety of different dimensions. Although not as formal, the strength of Wikipedia appears to stem, at least in part, from this same fundamental idea. It is a kind of diversification of viewpoint that also argues for bringing non-technical and technical players from a range of areas into the details of the modelling process. Model risk managers, with a finance professional's appreciation for the benefits of diversification, would be well advised to exploit this same idea.

The commonality among these three ideas is openness. It is easy, and common, to make important errors in the performance of technical tasks. Few are better placed to identify these errors – be they methodological, conceptual, or simply operational – than fellow finance practitioners. In some cases, deep technical knowledge is required – to opine, for example, on the reasonableness of a mathematical methodology. In other cases, it is conceptual understanding of the problem that matters. Non-technical portfolio managers can often provide invaluable insight into the veracity of key modelling assumptions. If one does not share one's results, these insights cannot be harvested. The consequence is black-boxedness and a commensurate increase in financial or reputational loss inherent in model risk.

Walking the talk

“You’re worse off thinking you have a model and relying on it than [...] simply realizing there isn’t one.”

– Emanuel Derman

Much of the preceding discussion is fairly commonsensical. The intent is to offer a conceptual structure to what is a difficult and important challenge for the financial community. Identifying a key driver, finding related disciplines, and extracting lessons hopefully has some value in this regard.

As a final step, it is useful to apply the previous appeal for openness and critique to this analysis. One could legitimately argue, for example, that a certain amount of cherry picking, or selective choice of examples, was involved in this argumentation. Openness seems to work in these settings, but one might argue that other areas of technical endeavour operate in the opposite direction. Law, medicine and engineering involve, on some dimensions, tasks that are, qualitatively at least, similar to the construction of financial models. Not being an expert in these fields, it is difficult to comment definitively. These conclusions could thus come with some important conditionality. All, however, are subject to varying degrees of oversight. The obligatory audit of financial statements and the appeal process for judicial decisions would appear – superficially, at least – to resemble the notions of openness, challenge and healthy criticism suggested in the previous analysis.

Another reasonable point of attack relates to one of the key assumptions inherent in this treatise: that black-boxedness is a key driver of model risk. If this is not true, then the logical arguments potentially collapse. An alternative driver is related to an important idea credited to Knight (1921); which is often referred to as Knightian uncertainty. These are the unknown unknowns famously entering the public consciousness and lexicon via former US Secretary of Defence Donald Rumsfeld. Such risk cannot be measured and, as such, cannot be mitigated. An explicit assumption of this analysis is that financial modelling deals principally with known unknowns. In a positive sense, this appears to be true. Most financial modelling involves the specification of value and loss distributions either estimated or calibrated to market conditions under different probability measures. This, by construction, leaves little scope for Knightian uncertainty; since such outcomes are out of the range of our current information sets, we cannot hope to include them in this way. From a normative perspective, however, this is thoroughly debatable. Taleb (2007), among others, criticise current modelling practices harshly for their failure to incorporate these risks. These are valid criticisms and, quite clearly, this challenge is not directly resolved within this framework. At the same time, unless one insists upon taking a very extreme stance on the underlying drivers of model risk, it also does not negate the underlying thesis. On the contrary, it opens the door for sceptical parties, such as Taleb (2007), to vehemently argue for the explicit incorporation of Knightian uncertainty.

Yet another competing driver of model risk is implicit in the work of Derman (1996, 2009, 2012): an over-reliance or overconfidence in modelling results. His work is logically compelling, practically useful and always colourful. Derman (2009) states, for example, that “the greatest danger in financial modelling is the age-old sin of idolatry.” By this, he means he views the key driver of model risk as a failure to appreciate the inherent fallibility of a simplified mathematical construct (ie a model) attempting to describe an unknowable reality. Opening one’s models to organised

external and internal critique, along the previously discussed lines, does not directly address this issue. An argument could even be made that an intense process of ongoing discussion and refinement might even accentuate this risk by lulling model users into a sense of complacency. This would be false. The spirit of the preceding arguments is that a healthy degree of scepticism on the part of the modeller – and his/her actively solicited colleagues in different domains – acts not only to mitigate logical and practical errors, but works to avoid the very hubris Derman (2009) so rightly cautions us against. It is not a one-off exercise to rubber stamp a specific model, but rather an ongoing commitment. Thus, although this paper differs slightly from Derman (2009) in terms of the key underlying driver of model risk, the recommended strategic response appears appropriate in both cases.

Concluding thoughts

“One special advantage of the sceptical attitude of mind is that a man is never vexed to find that after all he has been in the wrong.”

– William Osler

Through identifying a key driver of model risk and highlighting a number of qualitatively similar activities facing similar risk drivers, a prescriptive set of strategies has been identified for its management. Openness, sharing of knowledge, and explicitly exposing one’s ideas and implementation to criticism form the foundation of the recommended strategic response. Independent validation of one’s models is one important step in this direction, but one can, and should, seek to go much further. Formal internal and external presentation of ideas, code sharing and repositories, limited publication of results, inter-institutional cooperation, and subject-matter training are additional tools. Based on the previous logical arguments, these policies or initiatives could usefully be added to the model risk manager’s toolkit. Like all elements of risk management, however, the ultimate success will rely upon the hard work and vigilance of all those involved.

References

- Brown, A (2011): "Wikipedia as a data source for political scientists: accuracy and completeness of coverage", *Political Science and Politics*, vol 44, no 2, pp 339–43.
- Brown, J, B McGourty and T Schuermann (2015): "Model risk and the great financial crisis: the rise of modern model risk management", *Wharton Financial Institutions Center Working Paper*, vol 15, no 1.
- Campbell, D and D Fiske (1959): "Convergent and discriminant validation by the multitrait-multimethod matrix", *Psychological Bulletin*, vol 56, no 2, pp 81–105.
- Danielsson, J, K James, M Valenzuela and I Zer (2014): "Model risk of risk models", *FEDS Working Paper*, no 2014-34.
- Derman, E (1996): *Model risk*, Goldman Sachs quantitative strategies research notes, April.
- (2009): "Models", *Financial Analysts Journal*, vol 65, no 1, pp 28–33.
- (2012): *Models behaving badly: why confusing illusion with reality can lead to disaster, on wall street and in life*, Free Press, New York.
- Freedman, C (1995): "The role of monetary conditions and the monetary conditions index in the conduct of policy", *Bank of Canada Review*, Autumn.
- Giles, J (2005): "Internet encyclopedias go head to head", *Nature*, vol 438, no 15, pp 900–1.
- Greenstein, S and F Zhu (2012): "Is Wikipedia biased?", *The American Economic Review*, vol 102, no 3, pp 343–8.
- Hatch, R (1999): "Correspondence networks", in W Appelbaum (eds) *The Scientific Revolution*, Garland Publications.
- Hemrit, W and M Arab (2013): "The major sources of operational risk and the potential benefits of its management", *Journal of Operational Risk*, vol 7, no 4, winter, pp 71-92.
- Jensen, H (2002): "Optimal degrees of transparency in monetary policymaking", *The Scandinavian Journal of Economics*, vol 104, no 3, pp 399–422.
- Knight, F (1921): *Risk, Uncertainty and Profit*, Houghton Mifflin Co, Boston.
- Lerner, J and J Tirole (2002): "Some simple economics of open source", *The Journal of Industrial Economics*, vol 50, no 2, pp 197–234.
- Lucas, R (1976): "Econometric policy evaluation: a critique", in K Brunner and A Meltzer (eds) *The phillips curve and labor markets*, pp 19–46, Carnegie-Rochester Conference Series on Public Policy.
- Morini, M (2011): *Understanding and managing model risk: a practical guide for quants, traders, and validators*, Wiley.
- OCC (2011): *Supervisory guidance on model risk management*, technical report, Office of the Comptroller of the Currency, April.

O'Connor, J and E Robertson (2004): *The royal society*, <http://www-history.mcs.st-and.ac.uk/Societies/RS.html>, School of Mathematics and Statistics, University of St Andrews, Scotland.

Rebonato, R (2003): "Theory and practice of model risk management", *Quantitative Research Centre (QUARC) Technical Paper*.

Schonland, B (1958): "The invisible college", *Royal Institute of International Affairs 1944-*, vol 35, no 2, pp 141–50.

Spier, R (2002): "The history of the peer-review process", *Trends in Biotechnology*, vol 20, no 8, pp 357–8.

Stallman, R (2000): *The free universal encyclopedia and learning resource*, <https://www.gnu.org/encyclopedia/anencyc.txt>.

Taleb, N (2007): *The black swan: the impact of the highly improbable*, Random House, New York.

The Economist (2018a): *Some science journals that claim to peer review papers do not do so*, June.

——— (2018b). *What are "predatory" academic journals?* July.

Von Krogh, G and E von Hippel (2006): "The promise of research on open source software", *Management Science*, vol 52, no 7, pp 975–83.

Wootton, D (2015): *The invention of science: a new history of the scientific revolution*, Harper Perennial, New York.

Model risk: a novel approach using a category-oriented framework⁺

Paulo Cacella*

Abstract

Amidst today's rapidly changing business environment, most government entities and private enterprises support their decision-making process with data, projections and analyses using reality-based models. Misuse, model errors and a misunderstanding of model limitations, however, can lead to huge losses and strategical misdirection. This paper presents an innovative way to reduce model risks, proposing a framework based on categories that allow for diversifying risks and, importantly, grant decision-makers the ability to adjust models in line with the changing environment. The main objective of this work is to establish a viable framework for a holistic view of model risk and explain how it can be applied to strategic decision-making.

Keywords: model, model risk, risk management.

⁺ The views, thoughts and opinions expressed in this paper are those of its author and do not necessarily reflect those of the Central Bank of Brazil or its members.

* Corporate Risk Dept, Banco Central do Brasil, Brasília, DF, Brazil; Corresponding email: paulo.cacella@bcb.gov.br

1. Introduction

A model is defined by Merriam-Webster (2019) as: “a simplified representation of an object; a system of postulates, data and inferences presented as a mathematical description of an entity or state of affairs”. Models capture a wide variety of representations and world views, and are used frequently in finance, decision-making, risk and resource allocation.

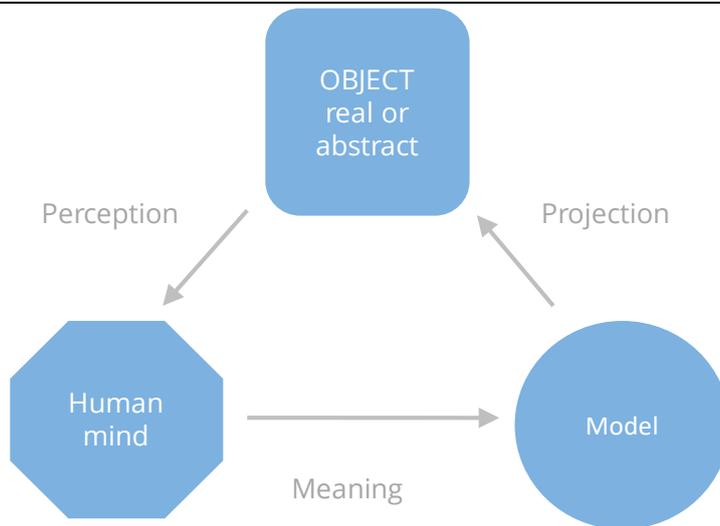
In this paper, and using Merriam-Webster as a foundation, I simplify the meaning of a model as “a human construct that represents a view of the world”.

As a human construction, we often try to explain possible relationships among objects, the human mind and the model. We try to define categories based on models’ shared characteristics. Models, in this sense, are representations of an object: real or abstract, that may have a meaning and must be projected from our minds into reality. Figure 1 shows a pictorial representation of the definition of model as an entity, ie as a conceptual reality.

In such a representation, the existence of an object, real or abstract, is accepted, as is the human mind as a processor that obtains data from its senses before building explanatory models of the object. This is done through perception, which maps sensory input data to environmental awareness. This awareness is subjective and high in entropy. To make things understandable, ie capable of generating objective and rational actions, we search for a meaning. Then we build models that can be projected into our conceptual reality. The reasons for designing a model are to forecast future events, comprehend past ones and to allow a rational view of the world with cause-and-effect relationships.

Model as an entity

Figure 1



Source: Author's elaboration.

1.1. Epistemic model categories

To better understand the role of models in our connection with reality, I introduce them as abstract entities that can be grouped into epistemic categories. These categories represent the object's nature itself, in contrast to applied categories, which are oriented to techniques targeting the problem-solution format. I propose an approach where each of the following epistemic model categories is related to an object type that we intend to replicate or understand:

Pure abstract model: a model based on an abstract conception of the world. The model's projection represents an abstract object, eg pure algebraic mathematical models. It is not intended to forecast real quantities or future events; it is a framework rather than an imitation of a real object.

Natural model: a model based on an understanding of how nature works, eg physical sciences models. The model's projection represents a real object. It is intended to forecast future events and other data through a rational view of the world. This model relies heavily on empirical sensory data.

Emergent model: a model based on emergent sciences linked to behaviour or emergent phenomena, eg social sciences. The model's projection may represent abstract or real objects. It is intended to forecast future events and data, or to mirror rational perspectives of the world and/or an ideology.

1.2. Understanding the problem of model risk

The scope of model risk is usually limited to the model itself and not to the whole framework and set of models. Model risk is mainly related to validation aspects such as an incorrectly designed model, insufficient or flawed input data, a misunderstanding of model limitations etc. This process is called model validation, and though it is present in almost all large organisations worldwide, it lacks in several important aspects related to the construction of a model.

Firstly, a model is designed to represent a schematic, partial and limited view of an object. It has an inherent error that it will lead to an approximate representation of the reality. In the case of an emergent epistemic category, like sociology and economics, a model is usually prone to generating misconceptions and misrepresentations as the external environment changes. A practical example is the use of value-at-risk (VaR) and credit risk ratings during and after the 2008 Great Financial Crisis. At the time, standard models that relied solely on historical data collapsed as the economic and financial system moved to a new equilibrium. A crisis can be explained in a systems engineering perspective as a transitional phase between two equilibria plateaus. In this way, environmental conditions are real and important factors to consider when managing model risks.

Another relevant aspect to consider is that even if a complete set of alternative or complementary models exist, they lack the synergy needed to be used as a single strategy in the decision-making process. This is because a preferred model usually deprecates older models, concentrating risks as it is used as a single model of reality. A thorough interpretation of the breadth of model risk clearly goes beyond the analysis of each single model. However, it is nevertheless necessary to have a holistic view, ie a connecting framework among models, to reduce misuse and enhance the quality of decision-making.

Modelling practice is strongly related to the amount of data and processing power available. Natural epistemic models, in general, due to their stability in time and the low number of significant explaining variables, are the easiest to implement with limited data and processing, adding to the fact that they are easy to comprehend. These models are based on the assumptions that nature has no teleological basis, that reality behaves as in the so-called "A Series"¹ (McTaggart (1908)) and that models' structures remain stable even if parameters change. That is why they are used as an initial approach to solve problems that are related to emergent epistemic models such as those in social sciences. Because emergent phenomena are from a distinct nature, with the human factor as a central issue, we understand that this approach of using only a natural epistemic model set is inadequate, imprecise and may lead to ideology manipulation and calibration/time window bias.

With the recent availability of big data and cheap processing power, new opportunities are arising to tackle complex problems related to social sciences and behaviour. A new revolution of computer-based models using artificial intelligence, new statistical models and a less comprehensible but more precise view of the world is what will set new standards in our era. The handling of human-based, decision-making models is also, somewhat, put in a distinct basket, creating an opportunity for us to redesign the way we see and use models in an organisation. A new approach is necessary to tackle the problem of model management and model risk. This is the idea I propose and discuss in this paper.

2. Discussion

Risk is often discussed in terms of taking preventative actions or making course-changing decisions that could minimise losses or amplify opportunities. In terms of loss control, model risk offers a procedure known as model validation, ie a model risk mitigation tool that is usually implemented in every model in an effort to address several sources of model risk.

In this paper, I present a novel approach on how to handle model management and model risk within an organisation. I begin by outlining the current view on model risk, followed by my proposal, which includes an abstract framework, decision-making support and, finally, governance.

¹ "A Series" are series that allow the existence of past, present and future, in contrast with "B Series" where flow of time is an illusion.

2.1. Model risk

Model risk can generally be defined as a loss that an organisation may incur as a result of decisions made based on errors in internal models (KPMG (2016)). It is usually considered a subset of operational risk. Even if this can be accepted conceptually, I consider this classification to be inadequate, as operational risk is strongly related to a business's process control activities, while model risk can be strongly impacted by non-control elements, such as changes in the external environment and intersubjectivity of players. Also, model risk has a higher impact on strategic decisions than on operational ones. Because the main nature of models is to support decision-making, it is also, to a high degree, related to strategic actions and strategic risks. In this sense, model risk is better managed when viewed outside an operational risk framework and distinctively from any other major type of risk. Thus, it must be considered in the highest levels of an enterprise's risk framework.

2.2. Sources of model risk

Model risks may have internal and external factors that can be analysed during model validation, such as the following:

Internal sources

- incorrect implementation
- incorrect model specification
- misunderstanding of model limitations
- flawed input data
- errors in calibration
- inaccurate numerical approach
- weakness in backtesting

External sources

- changes in the environment that affect the world model explanation
- advances in technology

2.3. Regulatory initiatives

Although the proposed framework is not related to any one specific activity, the following example from the financial sector provides insight into model risk validation.

Several regulatory initiatives arose after the 2008 Great Financial Crisis with the objective of minimising model risks and its nefarious consequences for the decision-making process. The US Federal Reserve and the Office of Comptroller of the Currency published in 2011 supervisory guidance on model risk management with reference SR 11-7 (Federal Reserve (2019)). This turned into a regulatory standard for model risk management. The focus of the guidance is to challenge models using a thorough understanding of all relevant sources of model risk and mitigation initiatives covering the end-to-end model lifecycle. The United Kingdom and the European Central Bank (ECB) also have in place regulatory initiatives on model risk. The ECB released the *Guide for the Targeted Review of Internal Models (TRIM)* in 2017 (ECB (2017)).

One key aspect in all regulatory frameworks is that models are viewed as an inventory and that model validation is individually managed.

2.4. Applied categories

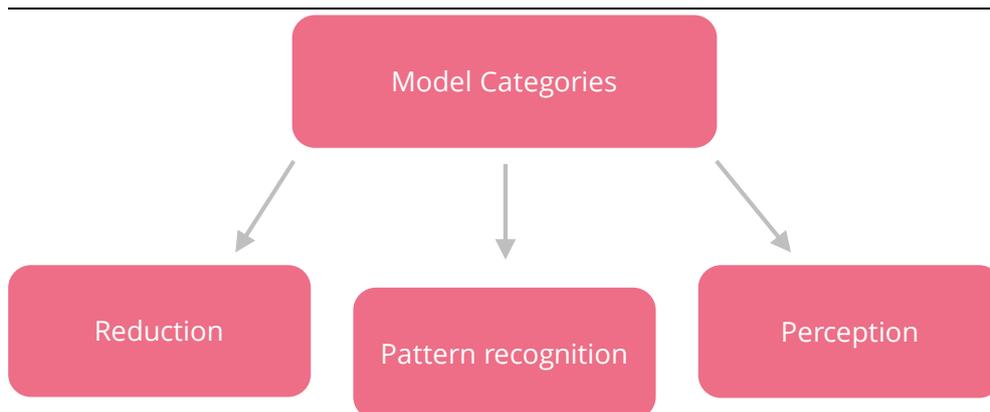
In order to build a workable, comprehensive model risk framework as an effective support for decision-making, I argue that it is not sufficient to handle models as an inventory and manage model validation individually. My belief is grounded in the following key aspects:

- big data and processing power are available in staggering volumes nowadays;
- models have different categories that must be handled distinctly; and
- decision-making processes must be fed with more structured data to allow decision-makers to select models that best match the present environment and/or adapt to changing conditions.

To solve this conundrum, I propose a set of categories that are applied to model implementation and diversification. Specifically, I recommend three applied categories based on a model as an entity that represents distinct views of the world: reduction, pattern recognition and perception (see Figure 2).

Model-applied categories

Figure 2



Source: Author's elaboration.

This setup matches perfectly with the concepts of meaning, projection and perception as shown in Figure 1 (ie the definition of a model), covering all aspects of its epistemological classes.

2.4.1. Reduction models

Reduction models aim at capturing a meaning represented by a small number of representative variables that express, in a simplified way, how the object behaves. The model is usually an equation or a set of equations with variables that could explain future events and the behaviour of the object at a significant level. Reduction models are oriented to meaning. Natural science models are usually of this kind. The modelling formalism can be defined as in system theory, using differential equations, difference equations and discrete events and/or dates (Zeigler et al (2019)). A classic example in economics is the dynamic stochastic general equilibrium (DSGE) model types. In finance, the Black-Scholes equation is another classic example. Reduction models may or may not be deterministic, being frequentist or Bayesian, but certainly

they can be understood and are defined by a manageable and comprehensible set of explanation variables.

2.4.2. Pattern recognition models

Pattern recognition is a scientific discipline whose goal is the classification of objects into a number of categories or classes (Tolk (2013)). A pattern recognition model is agnostic to a specific world view defined by a small set of understandable variables. This type of model is used when you cannot get a meaning from observations in a simplified way. In several circumstances the problem is so complex that each variable, individually, explains very little of the whole system behaviour. Worse than that, they usually are not understandable together, putting the modeller in a tentative position to reduce the model and possibly rendering it worthless. In cases like this, it is necessary to use computational power to find patterns that could be used to forecast observations, even if these patterns cannot be comprehended as a whole or in parts. These models are oriented to projection. Principal component analysis and machine learning, supervised or unsupervised, are examples of pattern recognition models. They frequently need a large amount of data and processing power.

The question of inadequacy of a numerical model regarding low predictive power, overfitting and other regression diseases is related to the validation process of the specific model under study. The decision-making process requires necessary information including the confidence level and the type of environment where it can be used securely.

2.4.3. Perception models

Perception models are oriented, as expected, to human perception. We, as a biological species, are very complex information processors. We absorb an enormous amount of information through our senses and process it in a way that the most complex computers in the world cannot yet handle. Also, human beings have objectives, some of which are hidden from others. Pattern recognition and reduction models are dependent on past observations and pre-conceived world views, respectively. They are not efficient during changing environmental conditions, like a crisis, which represents a transient step between two equilibria plateaus.

Perception models use real time information in a much more effective way when there is a rupture in system behaviour and the regime change generates a transient response. Perception models are more adaptable to a changing environment. Consensus methodologies like Delphi (Brown (1968)) and expert elicitation (Colson and Cooke (2018)), and others briefly explained below, are examples of perception models. Because they are pivotal in this proposal and not used often in organisations, I will discuss some of them in more detail.

Delphi: a technique used for the elicitation of opinions with the object of obtaining a group response of a panel of experts (Brown (1968)). The main idea behind this technique is to replace confrontation and brainstorming with a planned sequence of questionnaires among a panel of experts. This process may be repeated several times until a consensus among the group is reached on how to address all problems raised. The opinions are anonymous and questionnaires are conducted individually. The selected panel (usually totalling five to 20 people) are experts in the respective areas under review. An interesting approach to Delphi is the technique Real-time Delphi,

which tries to speed up the process and is well suited for fast financial decision-making (Gordon and Pease (2006)).

Expert elicitation: a systematic approach to synthesise subjective judgments of experts on a subject where there is uncertainty due to insufficient data, ie when such data is unattainable because of physical constraints or lack of resources (Slotje et al (2008)). This method relies on subjective statistical distributions agreed upon by experts and considers lagged observables when taking into account delayed effects. As Morgan (2014) stated: “We humans are not equipped with a competent mental statistical processor. Rather, in making judgments in the face of uncertainty, we unconsciously use a variety of cognitive heuristics. As a consequence, when asked to make probabilistic judgments, either in a formal elicitation or in any less formal setting, people’s judgments are often biased. Two of the cognitive heuristics that are most relevant to expert elicitation are called “availability” and “anchoring and adjustment.””

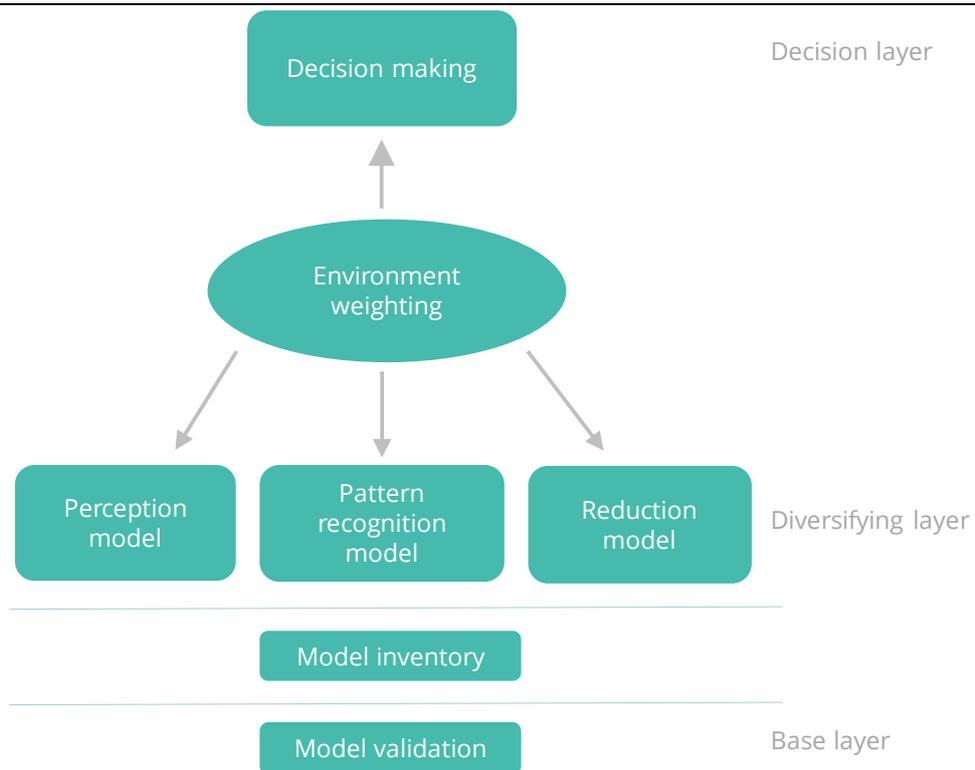
The Analytic Hierarchy Process (AHP): a general theory of measurement. The AHP is used to derive ratio scales from both discrete and continuous paired comparisons in multi-level hierarchic structures. These comparisons may be taken from actual measurements or from a fundamental scale that reflects the relative strength of preferences and feelings. The AHP has a special focus on departures from consistency and the measurement of those departures, as well as with dependence within and between the groups of elements of its structure. It has found its widest applications in multi-criteria decision-making, in planning and resource allocation and in conflict resolution” (Saaty and Vargas (2001)). The main objective of the technique is to reach a goal using the best group information focused in the problem’s objective. It can be used to attach actions to strategies, which is a paramount step in the decision-making process.

Multiple-criteria decision-making (MCDM): a subfield of operations research that has been succinctly defined as making decisions in the face of multiple conflicting objectives (Ramanathan et al (2017)). Where there are multiple objectives with the possibility of one optimal solution to the problem, a performance measure needs to be defined and an MCMD optimisation implemented. There are several distinct heuristic approaches to a problem like genetic algorithms. Applications in finance can solve complex problems that go beyond simple convex optimisations (Cacella et al (2010)). The main challenge of this methodology is defining the performance function. However, another important aspect is that because the solution will be a Pareto frontier, usually the decision-maker is not limited by a single choice and can select the solution that has the best probability of reaching the specific goal.

2.5. Proposing a model risk governance

Figure 3 highlights a proposed model risk governance that covers all model-applied categories, which can also be applied when working with emergent models like social models (economics, sociological, financial etc) or other models involving the use of human behaviour.

The base layer is already well developed in current model risk frameworks. Usually, model validation is applied to individual models in the inventory. The innovative approach here is to create a diversifying layer that incorporates, for each decision-making process, a set of models covering, if possible, all three categories: perception, pattern recognition and reduction.



Source: Author's elaboration.

2.5.1. Diversifying layer

In emergent models related to social sciences, the human behaviour is the game-changing variable. In natural models like those for classic physics, we can be sure that certain events will happen – such as an eclipse or lightning – regardless of human behaviour. Even when dealing with complex problems like a delayed-choice double slit quantum experiment or deterministic chaotic dynamic systems, there is no human behaviour factor. In emergent models of social sciences, however, everything may change without warning. Belief, will, compassion and other human attributes can change in the blink of an eye. The complexity is so high that we cannot simplify the explanation of the phenomena in a meaningful way. Precisely because the changing environment is associated with the impossibility of reducing complexity, it is important that all three categories of models can be used simultaneously. In this case, the decision-makers can state, subjectively, the weight of each model in their decision-making according to the environment.

One might ask if such weighting is a superimposed perception model over the diversifying layer. The answer is both yes and no. The very nature of the human decision-making process is a perception model. However, there are some subtle differences regarding approach in the cases at hand. The perception model in the diversifying layer is setup to minimise personal bias and preferences and to collect filtered information from experts in each field. The decision-making process, in another way, relates to a small number of people, like those comprising a board, who usually are not technical experts in each specific field and are more prone to be

affected by a principal-agent problem. It is not easy to impose at this level, without strong governance, the correct incentives for a sound decision-making process. However, I believe the availability of a set of models that can show aspects of the same problem in distinct viewpoints may be a restraint regarding agent behaviour.

The diversifying layer has the objective of mitigating risks associated with a single viewpoint/solution for a problem. By contrast, in model validation, mitigation is usually restricted to factors like uncertainties in estimates, inadequate model use, lack of data etc.

One additional advantage of the proposed model is that it is designed to embed the treatment of emerging risks. The most dangerous emerging risks are those that may occur in the transient phases of regime changes. These kind of risks are poorly assessed with models that use historical data as a single source of information. The diversifying layer provides a way to tackle these risks through a sound decision-making process.

2.5.2. Decision-making process

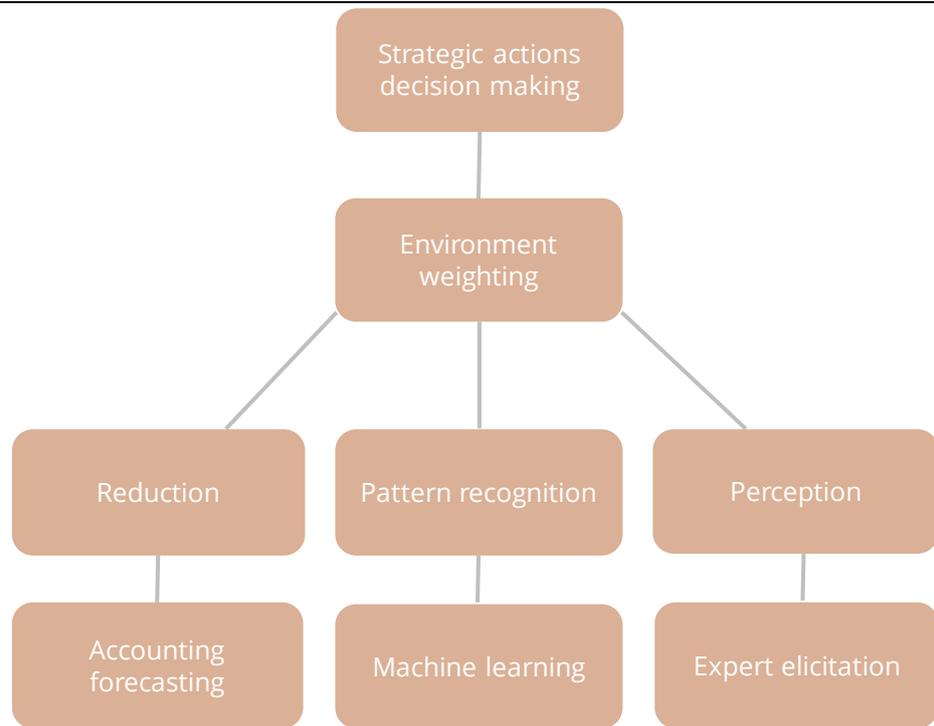
Decision rules under uncertainty is a classical problem in decision theory. One of the key aspects regarding human decision-making is that human bias may render a process inefficient or even ineffective. This was pointed out in a simple way in prospect theory (Tversky and Kahneman (1974)). Beyond techniques like decision trees, decision matrices and others, we consider multi-objective trade-off analysis (Haimes (2009)) as an important reference point to consider when designing a model risk framework in an organisation. Multi-objective methods largely depend on a subjective assessment of weights that will generate a Pareto frontier of available solutions. Even if there is scepticism in using models as a tool to generate final answers in a decision-making process, it is reasonable to use them, at a minimum, to guarantee that decision-makers' perceptions and available information are not self-contradictory, which would lead to inefficiencies in the process of mapping intentions to actions.

One further factor to consider when using models that have inputs from other models is the uncertainty propagation. Special care must be taken to control these effects in order to avoid presenting incorrect or biased data to the decision-makers.

Finally, although automated decision-making is a very convenient approach to some operational and repetitive processes, it is not adequate for strategic decisions, where human factors, bias and preferences are not yet mapped to support the changing environment.

3. Framework in practice

To exemplify the theoretical approach of the proposal, I present a practical implementation of the framework. A standard case is used as a hypothetical example: decision-making on strategic actions in a company (see Figure 4).



Source: Author's elaboration.

The environmental conditions include:

- changing economic environment, with huge uncertainties in profits and geopolitical aspects; and
- a company with budget restrictions in information technology.

The models available include:

- a reduction model based on historical data forecasting earnings and cashflow behaviour in the next year;
- a pattern recognition model based on machine learning with some sentiment analysis showing that the market is set to grow; and
- a perception model based on an expert elicitation round with the main executives of the company requesting a shift on the main product line.

With the information of all three model categories and the environmental conditions, the decision-makers set actions based on their trust of each model output and the applicability to the specific situation. This process is subjective by nature and cannot be automated.

In our sample case, the decision-makers, based on environment data, decided to begin a speedup study on shifting the product line, do not trust in machine learning because of changing environment and are conservative about earnings for the next year based on the same reason.

4. Conclusions

I have outlined how the current framework of model risk is oriented to model inventories and individual model validation practices, and how my proposed framework tries to take advantage of modern technologies, perception models and big data to enhance the strategic decision-making process in emergent phenomena like social sciences. I also introduced the concept of model-applied categories and environmental weighting. I believe this approach brings more flexibility, accuracy and precision in supporting the decision-making process of an enterprise.

References

Brown, B (1968): *Delphi process: a methodology used for the elicitation of opinions of experts*, RAND Corporation.

Cacella, P, I Damaso and A da Silva (2010): "A strategic asset allocation methodology using variable time horizon", in Berkelaar A, Coche J and Nyholm K (eds) *Interest rate models, asset allocation and quantitative techniques for central banks and sovereign wealth funds*, Palgrave Macmillan, pp 93–111.

Colson, A and R Cooke (2018): "Expert elicitation: using the classical model to validate experts' judgments." *Review of Environmental Economics and Policy*, vol 12, no 1, winter, pp 113–32.

ECB (2017): *TRIM Guide for the Targeted Review of Internal Models*, February, https://www.bankingsupervision.europa.eu/ecb/pub/pdf/trim_guide.en.pdf.

Federal Reserve (2019): *SR 11-7 Guidance on Model Risk Management*, <https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm>.

Gordon, T and A Pease (2006): "RT Delphi: an efficient, "round-less" almost real time Delphi method", *Technological Forecasting and Social Change*, vol 73, no 4, May, pp 321–33.

Haimes, Y (2009): *Risk modeling, assessment and management*, Wiley.

KPMG (2016): *Model risk management: A sound practice for meeting current challenges*.

McTaggart, J (1908): "The unreality of time" *Mind*, vol 17, pp 457–73.

Merriam-Webster (2019): *Merriam-Webster dictionary online*, <https://www.merriam-webster.com/dictionary/model>.

Morgan, M (2014): "Use (and abuse) of expert elicitation in support of decision making for public policy", *Proceedings of the National Academy of Sciences of the United States of America*, vol 111, no 20, May, pp 7176–84.

Ramanathan, Ramakrishnan, M Mathirajan, and A Ravindran (2017): *Using multiple criteria decision-making models*, London: CRC Press.

Saaty, T and L Vargas (2001): *Models, methods, concepts and applications of the analytic hierarchy process*, Springer US.

Slotje, P, J van der Slujs and A Knol (2008): *Expert elicitation: methodological suggestions for its use in environmental health impact assessments*, RIVM Letter report 630004001/2008, The Netherlands National Institute for Public Health and the Environment (RIVM).

Theodoridis, S and K Koutroumbas (2009): *Pattern recognition*, Academic Press.

Tolk, A (2013): *Ontology, epistemology, and teleology for modeling and simulation philosophical foundations for intelligent M&S applications*, Springer.

Tversky, A and D Kahneman (1974): "Judgment under uncertainty: heuristics and biases", *Science*, pp 1124–31.

Zeigler, B, A Muzy and E Kofman (2019): *Theory of modeling and simulation*, Academic Press.

An alternative approach to measuring the liquidity risk of public investors' investment assets

David Doran, Steve Kilkenny, Šarūnas Ramanaukas and Alex Shablov¹

Abstract

Public investors – particularly central banks – often apply different criteria, compared with private investors, when deciding how best to allocate their reserves across the range of assets eligible for investment in their own portfolios. While return is undoubtedly an important factor, other criteria, in particular liquidity, often take on greater importance when managing public funds. Furthermore, where the public investor's mandate extends to managing reserves in order to facilitate effective management of exchange rates, the investment criteria can take an even more specific focus, with liquidity risk of particular importance. In recent times, the measurement of liquidity risk has attracted far more attention given the lessons of the Great Financial Crisis; both in terms of understanding actual liquidity when a crisis occurs, and in terms of perceived liquidity in the context of the subsequent expansionary central bank policies and large-scale asset purchase programmes. Given these developments, and evolving market conditions, this paper considers an alternative approach to measuring the liquidity risk of public investors' investment assets. The approach set out in the paper allows for a consideration of liquidity risk in the context of risk assessments and asset allocation decisions that are specific to the mandate and policy objectives of public investors, and central banks in particular. To demonstrate the approach, the paper applies the methodology to the euro area sovereign market and, in doing so, appears to track the deterioration in liquidity during a period of political or fundamental uncertainty while also indicating an increase in liquidity (or a reduction in liquidity risk) following the announcement of the ECB's asset purchase programme. Other potential applications of the methodology, from a central bank risk management perspective, are also considered.

¹ The authors are Head of Financial Risk Management and Senior Risk Analysts, 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 Ruth Gleeson for helpful comments as well as Steve Flanagan and Naoise Metadger for their assistance with the data. Any remaining errors or omissions are our own.

1. Introduction

Public investors – particularly central banks – often use a different set of criteria, compared with private investors, when deciding how best to allocate their funds across the range of assets eligible for investment in their own portfolios. While return is undoubtedly an important factor, other concerns, such as liquidity and capital preservation, often take on greater importance when managing public funds. Furthermore, where the public investor's mandate extends to managing reserves in order to facilitate effective management of exchange rates, the investment criteria can take an even more specific focus, with liquidity risk being of particular importance.

For the purposes of this paper, a security can be considered as liquid if its acquisition or disposal can be executed relatively quickly and with a minimal effect on the market price. There are a number of factors that may determine the liquidity of a particular security; for example, fixed income securities are generally less liquid compared with equities, since the latter are normally traded on open exchanges and the former largely in less transparent over-the-counter (OTC) markets (Laganá et al (2006)). Given that public institutions primarily invest in fixed income securities, liquidity risk should therefore be of particular importance from a risk management perspective. Within the fixed income asset class, bond liquidity is further determined by the subsector to which the issued security is assigned; for example, sovereign bonds are generally considered more liquid than corporate, covered or other non-government bonds (Galliani et al (2014)).

In addition, specific security features such as the amount issued, rating, duration and time to maturity (Galliani et al (2014)), as well as whether they were issued as part of a private placement (Amihud and Mendelson (1988)), further distinguishes between the actual and perceived liquidity of fixed income securities. It therefore follows that different bonds issued by the same issuer may have significantly different liquidity risk profiles. Additionally, the concept of "on-the-run" or benchmark bonds also determine a bond's relative liquidity; bonds that have most recently been issued should be relatively more liquid (Pasquariello and Vega (2009)). Securities that exhibit characteristics of illiquidity will usually experience a direct and visible impact on their trading patterns. Illiquid securities generally have lower traded volumes and display wider bid-ask spreads (Favero et al (2010)).

Heightened liquidity risk, in the form of an inability to dispose of a security within a required time frame and with limited impact on the price, is a risk to which public investors may have differing degrees of sensitivity. Firstly, there might be a negative impact on the security's price, resulting in potential loss of capital. Secondly, an inability to dispose of a security within the required timeframe might result in an inability to meet liquidity cash-flow requirements. However, it is important to note that one cannot consider risk in isolation of potential investment returns, and illiquid securities should command a liquidity premium. Therefore, there is some potential to earn higher returns if there is an appetite to take on more liquidity risk in a manner consistent with an investor's risk appetite. Private placements and hold-to-maturity (HTM) portfolios are examples of this.

However, the understanding of liquidity risk measurement has arguably changed as a result of the Great Financial Crisis (GFC); both in terms of assessing actual liquidity when a crisis occurs, and also in terms of perceived liquidity following the subsequent emergence of a benign environment driven by expansionary central bank

policies and large scale asset purchase programmes. Regarding the latter, measures of liquidity risk can be clouded by fewer market participants, in some respects, yet offset by the presence of large-scale central bank purchases, which has provided a guaranteed market purchaser for many fixed income assets. It is noteworthy to recall that conditions prior to crisis, wherein issuer ratings and an overly benign outlook proved not to be representative of the underlying risks and subsequent liquidity risks when market conditions deteriorated. As such, in considering the development of market conditions, public investors should question the accuracy of current measures of liquidity risks, particularly in relation to the universe of assets in which they typically invest. Given the lessons of the crisis, and the effect of non-standard central bank monetary policies on market conditions, this paper considers an alternative approach to measuring the liquidity risk of public investors' investment assets. The outlined approach advances the assessment of liquidity risk in the context of asset allocation decisions that are specific to the mandate and policy objectives of public investors, and central banks in particular.

The paper is set out as follows: Section 2 discusses liquidity risk in the context of the aftermath of the GFC and the liquidity risk considerations for public investors that inform the approach proposed in this paper. Section 3 discusses the importance of the data in this approach and the algorithmic methods that can be employed to ensure its robustness. Section 4 introduces a scoring model and how it is constructed, while Section 5 discusses some of the potential uses of a scoring approach, including both analytical and risk management applications. Section 6 concludes.

2. Liquidity risk considerations for public investors

2.1 Liquidity risk since the crisis

Since the GFC, there has been an increasing focus on liquidity² risk as a risk factor in its own right. For example, a number of new regulatory measures were introduced in response to the crisis, such as the *Liquidity Coverage Ratio* and the *Net Stable Funding Ratio* (Basel Committee (2010)), which aim to ensure that commercial banks can better withstand stressed liquidity events. Separately, other regulations have impinged on the ability of global investment banks to act as market-makers and provide two-way markets in fixed income securities (ESRB (2016)). This is more apparent in the non-sovereign segments of bond markets and has affected liquidity in corporate bonds (see eg CGFS (2016)). It is in the context of this increased recognition of liquidity risk that all investors, including public investors, have become more interested in the liquidity risk of discretionary investment asset portfolios.

Two additional developments affecting liquidity risk since the GFC are noteworthy. Firstly, in recent years, global financial markets have experienced a historically low interest rate environment – particularly in the euro area³, but also in

² Liquidity can be divided into two types. Funding liquidity is the ease with which a market participant can meet its obligations as they fall due. Market liquidity is the ease with which an asset can be bought or sold, at something close to its current market price. For the purposes of this paper, our main focus is on market liquidity.

³ For example, at the end of 2018, the German sovereign yield curve was negative out to seven years.

other major jurisdictions. This environment raises significant income challenges for fixed income focused investors such as central banks or other public investment institutions. In response to these challenges, public investors have explored alternative means of generating investment returns, albeit in a manner consistent with the typically conservative risk appetite associated with public investors.

This trend can be observed in the public investor survey data, such as UBS's *Annual Reserve Management Survey*⁴, which shows that the share of reserve managers that are investing in a broader range of asset classes (other than sovereigns and supranationals) has been increasing in recent years. In doing so, public investors may enter into less populated parts of the financial markets universe. Hence, it becomes even more important for these institutions to have a means of monitoring the liquidity risk profile of the investment assets in a manner commensurate with their specific risk tolerance levels and investment criteria.

Secondly, central bank policies in recent years, in the form of quantitative easing and large-scale asset purchase programmes, have stabilised various fixed income markets, which has helped contribute to a lower financial market volatility backdrop⁵. Coupled with a search for yield in the low interest rate environment, this can induce risk-taking behaviour by financial market participants – perhaps even in a complacent manner – and this can sometimes be a precursor to a financial crisis, in the form of a so-called “Minsky moment” (Danielsson et al (2018)). The realisation of the effects of a financial crisis as a consequence of the moral hazard of a Minsky Moment may come too late for investors to exit particular areas of the market⁶, due to the heightened liquidity risk at that time. This highlights the importance of a forward-looking approach to assessing liquidity, at an asset allocation level, as part of a holistic risk assessment of new investment proposals and asset allocation strategies in the context of overall risk appetite.

2.2 Liquidity risk management of public investors

Public investors can use a number of liquidity proxies to indirectly manage liquidity risk. As noted above, their traditionally conservative risk appetite has meant that they have tended to invest in large, highly rated issuers such as sovereigns, whose debt will always likely be relatively liquid and in high demand. In addition, concentration risk in particular holdings can be managed, as a proxy for liquidity risk, by applying limits on how much of a particular bond or issuer is held.

Other market risk management measurements, such as duration targets with limited deviation parameters, also seek to reduce the nominal amount of illiquid investments. In addition, the articulation of a certain level of liquidity risk appetite can

⁴ See www.ubs.com/global/en/asset-management/insights/asset-class-research/asset-allocation/2018/annual-reserve-management-survey.html.

⁵ For example, the VIX index reached an all-time low in November 2017.

⁶ It may be that other aspects of a central bank's mandate, such as maintaining financial stability, may lead to a decision not to contribute to pro-cyclicality in markets through the disposal of assets, although investment decisions of one's own reserve assets are normally viewed through the lens of normal investment principles where practicable.

facilitate a hold-to-maturity⁷ accounting approach on a portion of a public investor's reserves. Despite such risk management measures that seek to indirectly limit liquidity risk, it is not the same as specifically measuring liquidity risk in a manner that reflects the investment criteria associated with management of public sector investments and consistent with its risk appetite. Given the developments set out in the foregoing section, it can be useful for public investors to initiate a detailed evaluation of various liquidity risk management tools while considering the bespoke requirements for liquidity risk measurement appropriate for its own investment criteria and requirements.

2.3 Liquidity risk measurement requirements – some considerations

It is important that any means of measuring and managing risk (including the liquidity risk scoring approach considered in this paper) is appropriate for an institution's own needs. More specifically, it should be consistent with both its mandate and its risk appetite in the context of its investment policy principles and criteria. For example, a central bank with an explicit requirement to intervene in the foreign exchange market to manage the national currency would likely have heightened sensitivity to liquidity needs. Similarly, a public pension fund with regular cash outflow requirements would have equivalent liquidity sensitivities.

In the case of many public investors, its investment policy principles remain relatively conservative with an emphasis on avoiding investment losses taking primacy, such that the generation of return is subject to adherence to capital preservation priorities. In accordance with this, they might invest predominantly in sovereign and sovereign-like fixed income bonds. Therefore, some market available liquidity scoring methodologies that are calibrated to the broader bond market universe may not provide the best fit for public investor's investment management requirements. More specifically, when using such broadly calibrated metrics, some instruments or asset classes could be considered highly liquid, but, from a traditional central bank's (or public investors') perspective, could be considered less liquid when evaluated against its self-determined investable universe. It is therefore clear that, given its investment objectives and risk appetite, a tailored liquidity scoring methodology that could be specifically calibrated to a public investor's own investment universe might be more appropriate.

In this regard, when considering whether to develop a tailored liquidity scoring approach, the following characteristics might be considered important:

- **Transparent and easily explainable to internal, senior stakeholders.** There are many different ways to measure and manage liquidity risk, which can range from a relatively simple scoring approach to a more opaque "black box" approach, which may include machine learning techniques. Each approach has pros and cons, and can incorporate both quantitative and qualitative aspects.

7 While such an accounting treatment does not preclude selling an HTM bond, it is possible that such an action could, in certain circumstances, require marking all HTM assets to market, which would not be desirable in almost all circumstances.

A possible approach to take could be to develop a relatively transparent mechanism that could be easily explained to senior stakeholders for it to be utilised credibly in the context of risk monitoring and risk assessment considerations. Therefore, one could create a scoring model that includes a number of different observable and well understood indicators of a bond's liquidity, which combine an overall measurement of liquidity.

- **Flexible to changing requirements and risk appetite.** It is important that any risk measurement approach is fit for purpose and is robust over time. While commercial providers of liquidity risk solutions offer many benefits, they are often calibrated in a standardised way and not easily customisable.

It might be considered important to have an approach that allows for some flexibility in how a public investor ultimately measures liquidity risk. For example, a changing liquidity environment in financial markets may necessitate a change in how the different components are calibrated – particularly during periods of systemic change. In addition, it could be considered important to have an approach that could respond to potential changes in the liquidity risk appetite of a public investor. This could be implemented, for example, by changing the composition of the institution's "investable universe" in order to generate a more representative liquidity score distribution over time.

- **Sufficiently granular for the public investor's holdings.** It is important that a liquidity measurement approach generates useful information about the liquidity of investment assets, both in and of itself, but also in comparison to other investment assets. This allows judgements to be made about each asset and its relative liquidity characteristics. In this way, the methodology should allow for the comparison of liquidity characteristics among various factors of the investment portfolio. For example, comparison among asset classes, sectors and issuers.

Moreover, to allow for a more dynamic monitoring, the methodology should be sufficiently granular to capture the daily movements in market liquidity, as having a score or metric that is too rigid might result in larger, more volatile shifts in the time series. Finally, the methodology should allow for the differentiation among the various aspects of market liquidity, as this would help to identify the key drivers behind any changes in the market conditions.

- **Built and maintained in-house.** As with any risk management tool, consideration must be given as to whether a product could be procured from a commercial provider or developed in-house. Important considerations relate to the cost of buying the solution, versus the internal resources required to build and maintain a model and tool, as well as some of the considerations related to flexibility discussed above.

Ultimately, there can be a number of strong benefits to developing the model in-house. Firstly, many institutions have been developing their data analytics capabilities in recent years, such that in-house development of a liquidity scoring measure could leverage from these capabilities. Secondly, in-house development provides a good opportunity to enhance internal knowledge of liquidity risk, both narrowly in the area of the public investor's financial assets, but perhaps in a broader sense, over time, in other areas of its mandate, such as prudential supervision or financial stability.

It is important to note that, while a methodology can be developed in-house, such an analysis can be informed by studying the methodologies and approaches already available elsewhere. While other solutions and approaches provide many benefits, they do not necessarily align with the public investor's requirements discussed above, mainly in terms of the transparency, flexibility and granularity in alignment with the public investor's risk appetite and investment mandate.

3. Liquidity scoring data

In this section, the paper outlines an approach to sourcing, specifying and augmenting the data used for scoring liquidity.

3.1 Data source and specifications

As opposed to financial instruments traded on public exchanges, fixed income securities are predominantly traded through decentralised OTC markets. Since there is no physical trading location, the participants conduct transactions via various modes of communication such as telephone, Bloomberg chat or proprietary trading systems. As a result, trades are often completed without other participants being aware of the transaction details, which normally makes OTC markets a lot less transparent and the degree of liquidity somewhat difficult to gauge. In order to calibrate a model of market liquidity scoring or liquidity risk, a detailed and granular data set is required, covering a range of variables relating to the underlying bond transactions. In the case of OTC transactions, it can be difficult to obtain reliable and consistent bond trading data from public sources and, therefore, specialist data providers must be used.

The data set used in the model presented is procured from IHS Markit, a professional market data provider⁸. The data set used is an extract from the corporate and sovereign bond universe, which, in turn, is a subset of Markit's fixed income data universe. The extract includes data since 2012 and contains daily data collected from over 300 market makers – covering over 100,000 corporate and sovereign bonds globally⁹.

The data includes hundreds of bond-level fields, ranging from the instruments' basic parameters, such as coupon frequency and issuer region – to various liquidity measures, such as shadow liquidity¹⁰. For the liquidity analysis, the proposed model includes measures related to market depth, trading volumes, bid-ask (yield) spread and maturity. Table 1 outlines selected data fields and provides a short field description.

The proxies for market depth are given as (i) number of data sources and (ii) number of dealers, and both are relevant. For example, if the bond's data is available from numerous different sources, but only a limited number of dealers are quoting it,

⁸ See <https://ihsmarkit.com>.

⁹ See <https://cdn.ih.com/www/pdf/Pricing-and-Reference-Data.pdf>.

¹⁰ Shadow liquidity is applied when an instrument does not receive any market data but shares characteristics with other instruments that are liquid.

it could be that the same dealers are quoting the bond in different places, in which case the market could not be considered deep. Alternatively, if there were many dealers quoting, but these were only available in a certain area of the market, this would not be considered a deep market either.

Bond parameters selected for liquidity scoring tool Table 1

	Field alias	Field description
	Maturity	The date on which the principal amount of the security becomes due and payable, as stated in the terms of the security.
	Bid-ask yield spread	The difference between the ask yield and bid yield.
Market depth	Number of data sources	The number of distinct data sources that were received for the instrument set on a given business day.
	Number of dealers	The number of dealers quoting the bond averaged over a number of business days.
	Trading volumes	Trade volumes observed per specific bond, issuer or market sector.

Source: Markit.

3.2 Data Cleaning

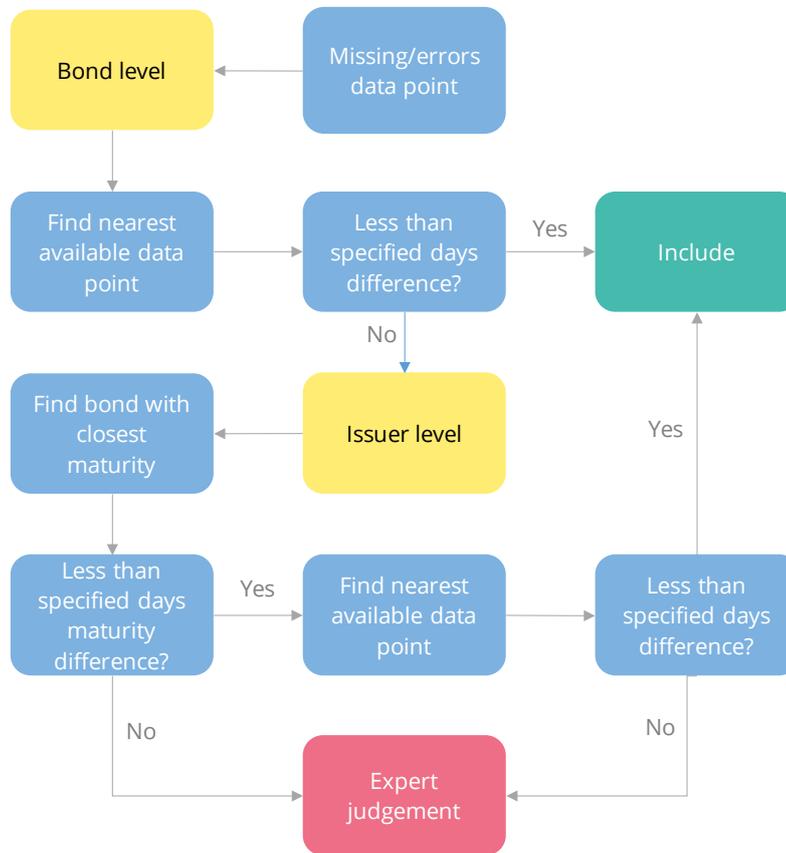
Large data sets rarely contain complete information, creating a number of challenges for researchers and practitioners. According to Kofman and Sharp (2003), 28% of publications in finance for the period 1995–1999 used data sets where, on average, 20% of values were missing. The missing data in the data set available for the model presented made up about 15% of the data utilised, which is in line with the statistics provided above. Most data gaps or consistency issues encountered were addressed by either interpolation techniques or manual replacement using expert judgement.

Given the large volume of data used in the exercise, algorithmic tools and techniques are preferable to manual interventions, which were kept to a minimum. However, outlier data points can distort the efficient calculation of a liquidity score and such data required some form of intervention. For instance, the bid-ask yield spread field had a number of data points that were either a number of standard deviations from the mean or, in some cases, the entry displayed negative spreads. Nevertheless, the vast majority of the outlier data points were normalised using an internally developed interpolation algorithm – a key element of the data processing phase.

As opposed to deleting records, data correction using the interpolation algorithm was chosen as the preferred method, since deleting records will affect other fields with valid observations and, as a result, leave gaps in the time series. The interpolation method involves three major processing steps: (i) identification of missing/outlier data points, (ii) interpolation using the bond’s own time series data and (iii) interpolation at an issuer level.

In the first part, the algorithm identifies missing data points in each data set field. Next, the nearest available data points are identified based on the bond's time series. The empty/error element is then populated by the nearest available data point if it is within a pre-defined specified date range, otherwise the algorithm proceeds to the issuer level matching. The matching on the issuer level is somewhat more complex. Firstly, a bond with the nearest maturity is identified from the same issuer. In the event that the maturities differ by an amount greater than a pre-defined specified range, the record is marked for review; otherwise, the missing point is matched to a nearest point in the time-series of the bond selected in the previous step. The nearest available point must be within a certain date range or the record is marked for review. Figure 1 provides a visual representation of the interpolation algorithm steps. This interpolation method proves to be a reliable data cleansing mechanism that allows detecting, fixing or removing errors and inconsistencies from the data set, and thereby substantially improves the usability of the data set.

The interpolation algorithm is by no means static and is, indeed, an area for potential enhancement. For instance, there are number of alternative methods to be explored to deal with missing data points such as multiple imputation or predictive imputation techniques. One example is a statistical framework-based method (Yuan (2010)) that allows an analyst to sample missing points from a deduced joint distribution function and preserve data statistical properties such as mean and variance. Another approach could be a range of methods, including machine learning techniques, that essentially imputes missing values from the observed ones (Bertsimas et al (2017)).



Source: Author's elaboration.

4. Calibration of a bespoke liquidity scoring tool

This section outlines the calibration of a potential liquidity scoring approach for monitoring and assessing the liquidity risk of fixed income investment exposures. The methodology ranks individual fixed income securities based on a number of liquidity indicators and aggregates these into a final liquidity score. The calculated scores allow comparison of exposures, not only at the individual security level, but also at issuer and portfolio levels. The tool provides a point-in-time estimate of liquidity, together with the ability to estimate liquidity for a historical time period, as well as the potential for more forward-looking analysis. This contributes to assessing trend and event analysis, with potential risk management applications that are considered in subsequent sections.

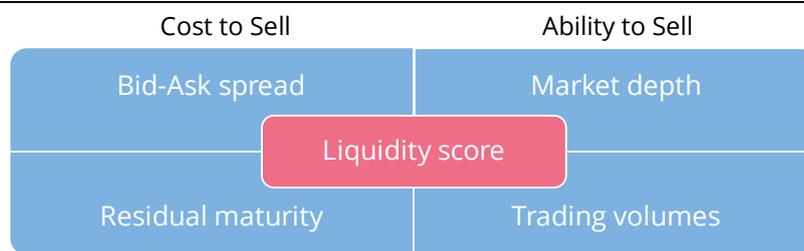
4.1 Methodology

To ensure that the approach was transparent and easily explainable to stakeholders, the methodology to estimate the liquidity score includes liquidity metrics that are

widely accepted as affecting bonds' liquidity. Metrics such as the bid-ask spread of the bond, maturity, trading volumes and market depth (as defined in Section 3) were aggregated using a weighted scoring methodology, which is outlined in the subsequent sections. Through a combination of all these metrics, the calibrated liquidity score provides a robust liquidity measure for individual securities, from a public investors' perspective, as it incorporates the different aspects of liquidity that can be experienced in the financial markets. These four metrics can be grouped into the two key aspects of liquidity risk, which we have labelled "cost to sell" and "ability to sell".

Cost to sell identifies how much it will cost to sell any given security based on the observable data available. The approach captures this liquidity aspect by incorporating the bid-ask yield spread of the bond, as well as its residual maturity. With regard to residual maturity, the liquidity of a bond increases as the bond nears its redemption date, as the exposure becomes closer to cash the nearer it gets to maturity. This aspect of the score best maps to the common description of the "tightness" of market liquidity (CGFS (1999)).

Ability to sell identifies whether there is a ready market available for this bond, should a bondholder wish to sell it, and to what extent a bondholder might "move the market" if trying to enter the market as a seller (something that public investors generally try to avoid). This indicator comprises two elements, namely; trading volumes and market depth. Firstly, a measure of the market depth, which combines available dealer quotes and the number of pricing sources, aims to identify whether there is a buyer for the bond. Secondly, a measure of trading volumes has also been incorporated, to try to capture how much of a given bond could be sold over a particular period of time, without significantly impacting the bond's price. Figure 2 displays the summary of the composition of the presented liquidity score. This aspect of the score maps to the common description of the "depth" of market liquidity (CGFS (1999)).



Source: Author's elaboration.

The final liquidity score for an individual bond is calculated by using the following simple approach. A security receives a score for each of the four key liquidity metrics. The scores for the four key metrics are determined by a number of predefined thresholds. For example, the bid-ask spread universe is divided into a number of thresholds or bands, where the band with the lowest bid-ask spread receives the highest liquidity score. The scores decrease in line with an increase in spreads. A similar methodology is also employed for the remaining liquidity components¹¹.

Next, the scores are aggregated to provide cost-to-sell and ability-to-sell scores, which are then combined to produce a final liquidity score for the instrument. The final score ranges from one to 100, with 100 being the most liquid. Once the unique securities are scored with an individual liquidity score, the final portfolio or issuer score is aggregated by using a simple weighted average approach, which is based on the relevant exposure size.

Furthermore, as well as calculating a score for all marketable fixed income exposures, such as bonds and treasury bills, other types of exposures may also be considered, such as repurchase agreements¹² and uncollateralised money market deposits. However, these exposures cannot be liquidated in the same way as one could do with a bond; hence a modified approach would be necessary. Given that the potential methodology for marketable and non-marketable exposures would not be homogenous, and would be based on different assumptions and factors, there would likely be difficulties in justifying the aggregation of these types of exposures in terms of an overall score. It is important that such differences are well calibrated, as well as being well communicated to senior stakeholders. A number of examples of how to incorporate these instruments into a liquidity score are outlined below, although it is acknowledged that they do not necessarily provide a perfect solution to the challenges raised by attempting to calibrate a portfolio level score.

As a starting point for looking at transactions such as these, an assumption could be made that all instruments with a residual maturity of less than one month are treated as equivalent to cash. Given that we assume such securities to be perfectly liquid in this scoring approach, an automatic score of 100 would be applied once any security reaches a residual maturity of less than 30 days. It is important to note that

¹¹ See Section 4.2 for more detailed information on the calibrations of the thresholds.

¹² It is true to say that an offsetting repo transaction could be conducted to offset the liquidity impact. However, it could also be argued that a repo transaction could be used to meet any liquidity need, which may not be possible in all circumstances, and may only have a limited term in any event.

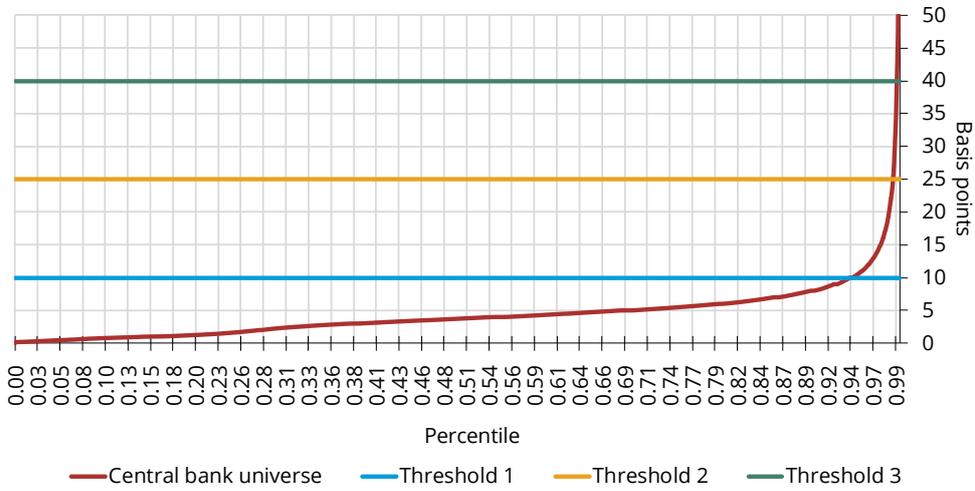
this would be based on likely liquidity requirements, which are assumed here to mainly revolve around needing to sell securities to rebalance a portfolio or to transfer from one portfolio to another. Of course, this assumption is heavily dependent on the specific mandate of a central bank or public investor. For example, a central bank that has an explicit mandate to be ready to regularly intervene in the foreign currency markets may need to have a more restrictive view of what “cash-like” means to them and their requirements. Such an active currency intervention strategy may require a residual maturity to fall within a week in order to be considered cash equivalent, for example.

Having established this assumption, a simple scoring approach could be applied to all non-fixed income exposures, on a sliding scale from 100 at one-month maturity, down to 50, that would be aligned with the institution’s maximum allowed maturity for such collateralised and uncollateralised exposures. The scoring methodology for these instruments can be informed by an analysis of the likely overall size and compositions of the exposures, primarily focusing on maturities and position concentrations. This can also be supplemented by an overlay of expert judgement. Another alternative methodology could include the liquidity score estimated based on the collateral that is being utilised in the transaction, usually a sovereign bond. Once a liquidity score on the collateral pledged is estimated, a discount factor could be applied to the score, to decrease the overall score due to non-marketability.

4.2 Calibration approach to score components

To meet the previously outlined granularity requirements, the model inputs and settings have to be appropriately calibrated. For example, it is important to ensure that the bid-ask spread and trading volume thresholds are granular enough to capture the daily liquidity changes of a particular bond. Separately, indicators such as market depth and bond maturity are more stable – hence the calibration for these variables can be less granular.

The calibration of the bid-ask thresholds should be based on the analysis of the bid-ask spreads of the public investor’s investable universe. This analysis indicated that more than 90% of all the observable spreads of bonds within a traditional central bank’s investable universe would fall within a very tight range of between zero and 10 basis points (see Graph 1). This meant that, if wider spread bands or thresholds were calibrated, the model would be less able to capture the changes in the bond’s daily liquidity. Therefore, the calibration of bid-ask scoring bands were calibrated in a way that allowed analysts to better rank the bid-ask spreads within a public investor’s investment universe. Interpolation was used within the bands to avoid discontinuous jumps in the scores, which might be based on small changes in variables.

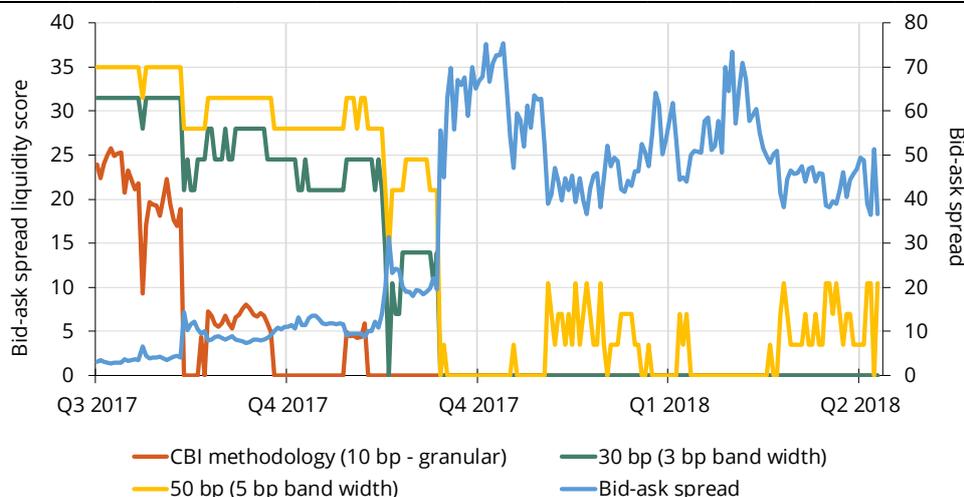


Source: Markit; Authors' calculations.

Graph 2 demonstrates how wider bid-ask bands and higher maximum allowable bid-ask spreads (once the maximum spread is reached, the security receives a score of zero for this aspect of its liquidity) would not be as responsive to daily liquidity changes of the bond. These types of wider calibrations would be more aligned to some market liquidity risk providers, which are typically calibrated to a wider bond market universe than a traditional central bank reserve manager, as discussed earlier. For this example, we use data for a specific corporate bond that experienced significant deterioration in risk profile and caused the bond's bid-ask spread to increase substantially over the period displayed. As can be seen by the graph, the significant changes in the bond's bid-ask spread, and the subsequent deterioration in liquidity, would not have been captured promptly by a liquidity metric that employs relatively wide bid-ask bands. For instance, a score based on the 50 bp wide band, depicted in the graph, does not indicate a significant drop in the bond's liquidity until November 2017. On the other hand, a more granular methodology, such as that described in this paper, displays greater sensitivity to developments in the bid-ask spread. The methodology presented captures substantial deterioration from July to November 2017, until the bond's bid-ask spread level increases beyond the final threshold range and is assigned the lowest score of zero, making such measurement more adequate for a public investor's investable universe.

Liquidity deterioration of a bond – bid-ask spread threshold comparison

Graph 2



Source: Markit; Authors' calculations.

The calibration of the other variable thresholds can be based on a more qualitative analysis, where expert judgement was applied. For example, the calibration of maturity scoring bands can be linked with the public investor's target investment horizons and duration targets. The calibration of the trading volume thresholds can be estimated by reference to the standard composition of a public investor's exposures and nominal amounts of holdings.

5. Potential applications of the liquidity scoring approach

Incorporating liquidity risk monitoring into a holistic risk monitoring framework with other risks such as market risk, credit risk and interest rate mismatch risk, may allow for the identification of previously unidentified risks and/or a more complete view of the overall interaction of the investment risks faced by the investor. Armed with these insights, risk managers would be in a better position to provide more granular information to portfolio managers, and the institutional governance bodies, in support of risk-informed decision-making.

However, there are a number of additional, potentially beneficial, approaches that a central bank, or other public institution, might take to utilise liquidity tools – such as the one presented in this paper – to better incorporate liquidity risk into their overall risk management framework. A number of such approaches are proposed in the subsequent sections, beginning with the simplest applications, such as market analysis, followed by more complex uses, such as portfolio liquidity benchmarking and optimisation approaches that can be built upon the scoring approach.

5.1 Market liquidity analysis

To complement risk monitoring and reporting of its own investment exposures, risk managers must also perform and communicate risk intelligence to the institution's senior management. Liquidity tools, such as the one presented in this paper, can be valuable assets that allow for monitoring of financial market liquidity more broadly. In particular, the tool could aid in analysing market liquidity trends or conducting event analyses.

Euro area benchmark liquidity score (2012–2018)

Graph 3



Source: Markit; Authors' calculations.

For example, using the liquidity scoring model, a simple analysis was performed to estimate whether liquidity in the euro area has changed in a meaningful manner during the period 2012–2018, which incorporates the introduction of the various ECB asset purchase programmes. An equally weighted euro area benchmark was developed, consisting of six countries (France, Germany, Ireland, Italy, Portugal and Spain) and four tenors (two-year, five-year, 10-year and 30-year). Graph 3 provides the summary snapshot of the analysis.

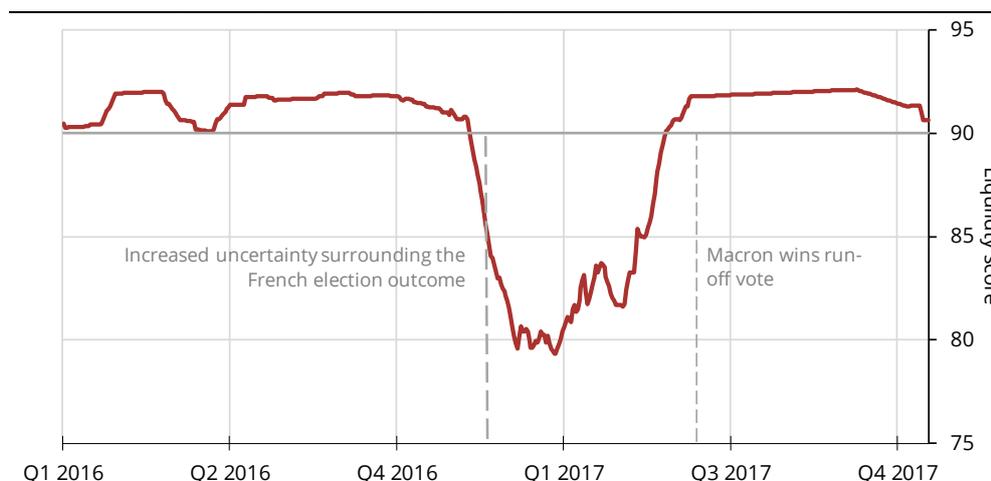
As can be seen, the score indicates that market liquidity increased significantly during the period 2012–2014, while the period 2014–2018 exhibited no major changes in market liquidity. The key driver in the increased liquidity in the presented model was the decrease in euro area bid-ask spreads. While this analysis is consistent with a number of recent publications (Jurkšas et al (2018), Larkin et al (2019)), there is some mixed evidence on the topic (ESMA (2018)). There are a number of reasons offered for the increase in liquidity in the European bond market, including the introduction of quantitative easing by the ECB and the overall increase in the credit quality of euro area sovereign issuers.

In addition to the longer-term trend analysis, the tool can also be utilised when evaluating liquidity changes in specific issuers surrounding a risk event. Graph 4 presents the evolution of the French 10-year bond around the time of the 2017 election in France, with the newly developed tool used to produce a liquidity score of between one and 100 in the manner described previously. As can be seen from the increased uncertainty surrounding the French election outcome in late 2016, the

liquidity of the bond began to decline as expectations for a Marine Le Pen victory grew¹³. It is important to note that the bond was still considered very liquid according to the model; however, there were some notable changes. Once the outcome of the election became clearer, the bond returned to its previous liquidity levels. Analysis of historical events such as these provides useful information on how market liquidity may react to similar events in the future. For example, if the uncertainty persisted for a longer period of time, or if the outcome of the elections was not perceived to be market friendly, the liquidity of the bond may have decreased further.

Liquidity score of French 10-year bond around the 2017 election in France

Graph 4



Source: Markit; Authors' calculations.

5.2 Liquidity benchmarking

Once the monitoring and analytical skills and expertise are developed from consistent use of a scoring tool, a more formal introduction of liquidity risk management in the investment management process, in the form of liquidity benchmarking, can be contemplated.

As with performance evaluation or market risk measurement, absolute liquidity risk figures provide useful information, but may not tell the full story. Therefore, liquidity must also be evaluated on a relative basis; for example, the liquidity risk of an issuer might be compared with its peers, a fixed income class such as corporate bonds might be compared with other classes such as sovereign fixed income, or active portfolio management decisions might be compared with a more neutral portfolio liquidity risk score. Therefore, there is a strong argument to use liquidity benchmarking when evaluating the relative liquidity of a portfolio manager's holdings or, indeed, the overall asset allocation of a public investor.

A role for portfolio liquidity benchmarking is supported by two observations. Firstly, a central bank or any other public institution must invest in something, as public funds left idle would forego potential returns (notwithstanding negative yields

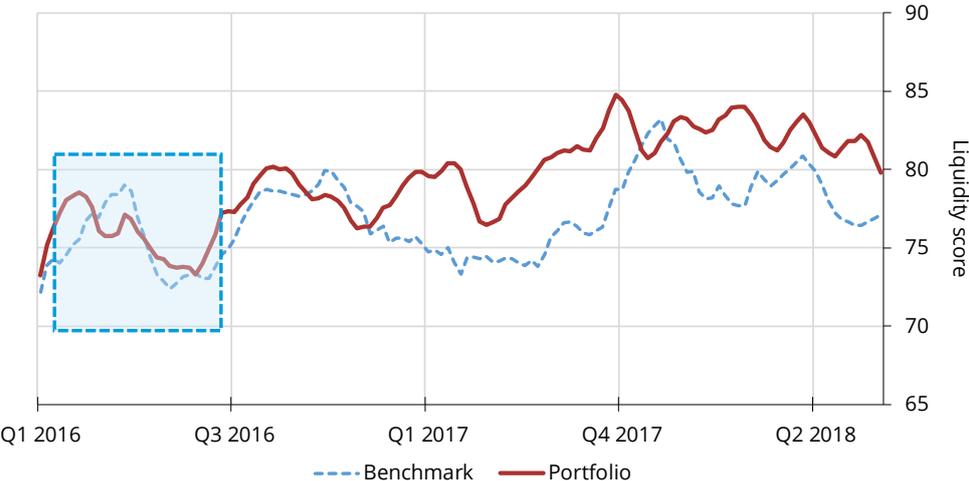
¹³ For an example of the sentiment at that time, see Financial Times (2017).

on some investment options in the current low interest rate environment). Secondly, the overall liquidity of the market inevitably changes over time caused, for instance, by various structural and regulatory factors, which are outside an investor’s control. In this way, at a portfolio level, an institution’s risk managers should not be “punishing” portfolio managers for systemic trends that are not within their control. Hence, similar to a performance benchmark used for portfolio manager’s investment performance evaluation, a liquidity benchmark might be used to evaluate the investment portfolio’s liquidity relative to a benchmark. This would allow the risk manager to determine whether changes in portfolio liquidity is being driven by the portfolio manager’s actions or by the overall change in liquidity of the market or market sector to which the portfolio has to invest, according to their mandate.

There are a number of options that can be considered when choosing the appropriate liquidity benchmark. Ideally, the chosen benchmark should be aligned to the investment manager’s return and risk objectives. Hence, the benchmark that is currently employed for the purpose of relative performance and risk measurement might be the best candidate. However, if risk managers want to evaluate liquidity risk of the overall market rather than the specified sector, a broader market benchmark might be more preferable.

Following the construction of the appropriate liquidity benchmark, the starting point might be the estimation of liquidity of both the individual portfolios and the specific benchmarks. An example of a liquidity benchmark and portfolio liquidity score are illustrated in Graph 5, where the decrease/increase in the liquidity score indicates an increase/decrease in associated liquidity risk.

Tracking portfolio liquidity versus a benchmark liquidity Graph 5



Source: Markit; Authors’ calculations.

As can be seen from this example, the portfolio manager is tracking the benchmark relatively well; in fact, during significant periods of time, the portfolio manager’s liquidity score is higher than the assigned benchmark. This information could be interpreted in two ways; positively, from a risk management perspective, as the portfolio manager is taking on less risk than the benchmark, or it could also have potentially negative aspects from an investment perspective, as the portfolio

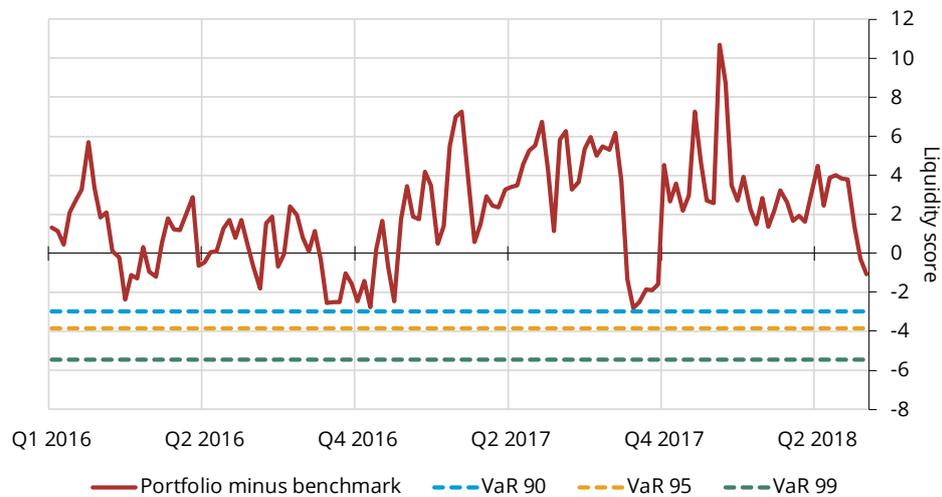
manager may not be fully utilising and capturing potential liquidity premia allocated to them by the benchmark.

Graph 5 also highlights a number of other observations relating to market developments. For example, from February to July 2016 (in the run up to the UK Brexit vote), portfolio liquidity risk increased noticeably (as the overall portfolio liquidity score decreased), which might raise some questions for a risk manager if portfolio liquidity risk was estimated in isolation. However, analysing this information in tandem with the benchmark liquidity metrics indicates that benchmark liquidity risk also increased (as the overall benchmark liquidity score decreased). Given that the benchmark tracks the portfolio's investable universe, a conclusion can be made that the overall liquidity of the invested portfolio decreased due to more systemic factors such as political uncertainties, rather than portfolio managers' actions.

The natural follow up question is how to calibrate a means of limiting the extent that a portfolio manager (or the overall investment assets) is permitted to deviate from the relevant benchmark? One way of answering this might be to allocate a deviation budget to the portfolio manager, whereby deviation beyond this budget would be limited. This would be similar to the type of active risk budget that is allocated to investment managers in respect of market and/or credit risk.

Benchmark deviation thresholds

Graph 6



Source: Markit; Authors' calculations.

This can be achieved by examining the volatility of the liquidity score of the benchmark itself, and the extent to which the benchmark liquidity score deviates from its mean value. This could be used as a proxy for the acceptable deviations of the portfolio liquidity score from the benchmark liquidity score. Portfolio liquidity deviations are estimated by subtracting the overall portfolio liquidity score from the overall benchmark liquidity score. Subsequently, these estimates might be employed to set limits against which the portfolio manager might be evaluated. For example, the limits could be estimated by choosing an appropriate confidence level (eg VaR 90), on either a historical or parametric basis. Graph 6 illustrates potential threshold levels. The choice of percentile can be guided by, amongst other things, the liquidity risk appetite of the investor.

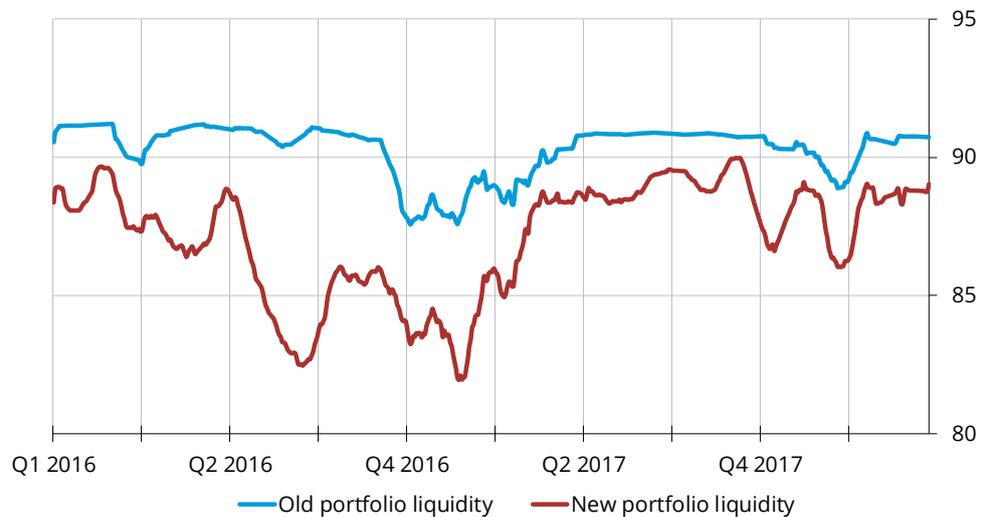
As can be seen from the example in Graph 6, the portfolio manager would not have breached any of the assigned limits during the past number of years. In order to operationalise this methodology, there are additional considerations that may need to be incorporated. It may be necessary to include some sense of the persistency of the limit breaches. For example, a portfolio manager might be informed about the breaches and corrective action might be requested only if the limits breach is persistent rather than a once-off occurrence. Again, these types of additional methodological approaches would have to be calibrated based on the institution's risk appetite¹⁴.

5.3 Liquidity optimisation

There are additional approaches that can be borrowed from other risk domains when designing risk management applications utilising the liquidity score. One such approach is to incorporate liquidity risk scoring into an optimisation exercise. This could be used in the context of performing risk assessments of new investment proposals. It should be noted, however, that such an exercise could be incorporated as part of a holistic risk assessment that also considers other types of risks (eg credit and market risks). It is, therefore, not proposed that asset allocation decisions are calibrated with respect to an optimisation exercise in terms of liquidity risk only.

In order to illustrate this approach, it is useful to look at a practical example. Consider a portfolio that consists of four issuers, which are equally weighted in terms of holdings (this example could easily be adapted to that of asset classes or currencies etc.). This portfolio has an associated liquidity score that can be tracked and monitored over time. Consider that an investment proposal is produced that wants to add a fifth issuer, with a subsequent reduction in the holdings of the first four issuers so that all issuers are once again equally weighted. A hypothetical new portfolio can then be created, which includes the benchmark holdings of the proposed new issuer together with the holdings already in the portfolio, and which measures its hypothetical historical liquidity score over the same period as the old portfolio. These two portfolios can be seen in Graph 7.

¹⁴ The benchmarking approach can also be applied at a global level, when making asset allocation decisions such as entering new asset classes, geographical areas or currencies. In this case, it is necessary to construct a global liquidity benchmark, which would include a combination of financial instruments from the most liquid currencies. For example, such a benchmark might include euro area and US dollar sovereign and sovereign-like securities, as well as other currencies, if appropriate.

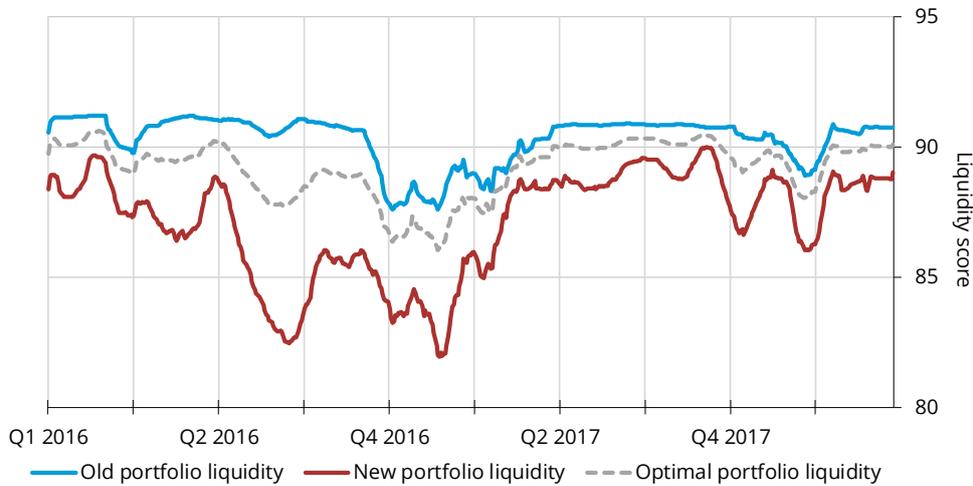


Source: Markit; Authors' calculations.

The graph shows that, on a historical basis, the effect of the addition of the new issuer would be to both reduce the overall level of the liquidity score, as well as increase its volatility over time. Similar exercises can be performed on a more forward-looking basis, such as in stress scenario analysis, if required. The question arises as to whether a risk manager would be happy with the outcome of changing the liquidity profile in this way. If not, it is possible to perform an optimisation of the liquidity score when considering an appropriate liquidity risk exposure.

There are a number of different bases on which to perform such an optimisation. For example, it is possible to construct an asset allocation that exhibits a generally higher liquidity score (either maximising or aiming for a particular value) or a lower score volatility (either minimising or aiming for a particular value). Another option is to limit the extent of deviation from a constructed global liquidity benchmark, similar to that described in Section 5.2. Examples of the output of such an optimised approach are illustrated in Graph 8.

As mentioned above, such an exercise could be performed as part of a comprehensive analysis of any potential new investments. Additional constraints could also be added to specific issuers that warrant an overweighting based on a consideration of the risk/return trade-off and, as such, many different liquidity profiles can be constructed and assessed in accordance with the public investor's specific risk appetite.



Source: Markit; Authors' calculations.

Similar to the use of mean-variance optimisations in asset allocation decisions, the assumption that the past history of an asset's liquidity is a good guide to the future is a significant assumption to make, and comes with some risks. Therefore, it may be useful to consider other means of calibrating such a liquidity optimisation approach, such as attempting to add a forward-looking element. This could be achieved by estimating the forward-looking parameters, with the addition of a simulation approach to the components of the liquidity score. This can also be complemented through the use of scenario analysis, where liquidity conditions of relevant asset classes in certain reference time periods could be utilised to derive potential outcomes for new asset classes. Finally, the utilisation of stressed liquidity factor correlations, either hypothetical or based on historical data, could also be incorporated in the optimisation process.

6. Conclusions

Since the GFC, there has been an increasing focus on liquidity risk as a risk factor in its own right. It is in the context of this increased recognition of liquidity risk that investors, including public investors such as central banks, have become more interested in measuring and monitoring the liquidity risk of discretionary investment asset portfolios.

The development process of the liquidity risk management framework, for a public investor, might identify that the nature of public investors' holdings necessitates a tailored liquidity scoring methodology, calibrated to an institution's own investment universe and risk appetite. Some characteristics that might be considered when developing such frameworks are transparency, flexibility and granularity.

This paper presents an approach to measuring and monitoring liquidity risk that can be tailored specifically to a public investors' investment assets. The methodology

presented in the paper was used to estimate whether liquidity in the euro area has changed in a meaningful manner during the period 2012–2018. Analysis using this approach shows that liquidity in the European sovereign bond market increased significantly between 2012 and 2014, while the period 2014–2018 exhibited no major changes in market liquidity.

Moreover, the paper considers a number of approaches that a central bank, or other public institution, might take to utilise liquidity risk scoring tools – such as the one presented in this paper – to better incorporate liquidity risk into their overall risk management framework. These approaches to managing liquidity risk, including concepts such as liquidity benchmarking and liquidity optimisation, can offer a number of potential benefits, such as a better evaluation of the portfolio liquidity position relative to the market, and a comprehensive assessment of the impact on the liquidity profile from potential changes in portfolio composition.

Finally, some next steps to develop practical applications of the methodology could range from enhancements of data pre-processing techniques to refining the applicable threshold and calibration approaches. The latter could include further development of the liquidity optimisation approach or the development of the look-through approach to assets linked to repo transactions. For public investors, the use of liquidity risk measurement and monitoring approaches is likely to become a standard feature of their risk management toolbox. By developing a better understanding and management approach for liquidity risk, alongside market and credit risk, public investors can enhance investment decision-making and ultimately limit potential losses. This paper has sought to contribute to this ongoing development path by presenting a liquidity risk scoring approach along with potential applications.

References

- Amihud, Y and H Mendelson (1988): "Liquidity and asset prices: financial management implications", *Financial Management*, pp 5–15.
- Basel Committee on Banking Supervision (2010): *Basel III – International Framework for Liquidity Risk Measurement, Standards and Monitoring*.
- Bertsimas, D, C Pawlowski and Y Zhuo (2017): "From predictive methods to missing data imputation: an optimization approach", *Journal of Machine Learning Research*, no 18, pp 1–39.
- Committee on the Global Financial System (1999): "Market liquidity: research findings and selected policy implications", *CGFS Papers*, no 11, May.
- (2016): "Fixed income market liquidity", *CGFS Papers*, no 55, January.
- 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.
- European Securities and Markets Authority (2018): *Liquidity in EU Fixed Income Markets – Risk Indicators and EU Evidence*, September.
- European Systemic Risk Board (2016): *Market Liquidity and Market-making*, European System of Financial Supervision, October.
- Financial Times (2017): "Could France's Marine Le Pen deliver Frexit?" www.ft.com/content/d37b6d90-fdd1-11e6-8d8e-a5e3738f9ae4.
- Favero, C, M Pagano and E Von Thadden (2010): "How does liquidity affect government bond yields?" *Journal of financial and quantitative analysis*, vol 45, no 1, pp 107–34.
- Galliani, C, G Petrella and A Resti (2014): "The liquidity of European corporate and government bonds: drivers and sensitivity to different market conditions", *JRC Technical Reports*, European Commission.
- Jurkšas, L, D Kapp, K Nyholm and J Von Landesberger (2018): "Euro area sovereign bond market liquidity since the start of the PSPP", *ECB Economic Bulletin Boxes*, no 2.
- Kofman, P and I Sharpe (2003): "Using multiple imputation in the analysis of incomplete observations in finance", *Journal of Financial Econometrics*, vol 1, no 2, pp 216–49.
- Laganá, M, M Perina, I von Köppen-Mertes and A Persaud (2006): "Implications for liquidity from innovation and transparency in the European corporate bond market", *ECB Occasional Paper*, no 50.
- Larkin, J, P Anderson and S Furlong (2019): "The Irish government bond market and quantitative easing", *Central Bank of Ireland Quarterly Bulletin*, no 2, April.
- Pasquariello, P and C Vega (2009): "The on-the-run liquidity phenomenon", *Journal of Financial Economics*, vol 92, no 1, pp 1–24.
- Yuan, Y C (2010): *Multiple imputation for missing data: concepts and new development (Version 9.0)*, SAS Institute Inc, Rockville, MD.

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 Conference. The volume is the first 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