Essays in Finance



Ilona Mostipan

St Edmund Hall and Saïd Business School

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2018

This thesis is dedicated to

my family

for their love, encouragement, and supporting me in all my pursuits

Acknowledgements

First and foremost, I want to profoundly thank my two advisors, Mungo Wilson and Kevin Sheppard. It has been an enormous pleasure to work on this research under their guidance, and I appreciate all their contributions of time and ideas, as well as general support.

It has been a real honor and an incredible learning experience to work with Thomas Hellmann and Nir Vulkan on our joint paper.

I also tremendously appreciate the invaluable mentorship and financial support provided by Frank Hwang and the Oxford Chinese Economy Programme (OXCEP) at St Edmund Hall.

I am also very grateful to the faculty members at Saïd Business School (SBS), namely Pedro Bordalo, Thomas Hellmann, Ludovic Phalippou, Ilaria Piatti, Tim Jenkinson, Ken Okamura, Joel Shapiro, Oren Sussman, Dimitri Tsomocos, Tom Noe and Nir Vulkan for providing invaluable feedback during seminars and meetings.

I also extend my thanks to the faculty members at the Economics Department—Guido Ascari, Martin Ellison, Christopher Bowdler, Andrea Ferrero, Climent Quintana-Domeque—for their feedback during an early stage of the debt management paper.

The paper on information diversity has benefited from comments by Jennifer Conrad, Kevin Crotty, and Jeremy Large, and for these helpful comments, I am also very grateful.

I am also grateful for the academic advice and professional support provided by John Knight, Dominik Karos, Keith Gull, Helen Chen and Martin Slater of St Edmund Hall.

Words of appreciation also go out to Bige Kahraman and Dong Lou for their insightful comments during the viva voce examination.

Additionally, I would like to thank SEEDRS for providing their data.

I would like to thank the following Research Assistants for their excellent data gathering work: Samuel Gibb, Chris Johnson, Paul Schmidt-Engelbertz, Julia Selgrad, Shayna Servillas, Laura Smith, Chris Nottingham, Serge-Raymond Nzabandora, and Charlotte Winter.

I also thank the SBS Sainsbury Library and Chris Flegg for access to data that is part of this research.

I extend my appreciation to the SBS Research Assistance Fund Committee and Graduate Teaching Assistantship Award Committee at the Economics Department for fruitful employment opportunities.

I would like to also sincerely thank my doctoral colleagues, Johan Cassel, Beata Gafka, Anna Stepashova, Deborah Anderson, Tanja Collavo, Neroli Austin, Alex Montag, David Cardoso, Theofanis Papamichalis, Ji Yan, and Maria Balgova for their feedback and encouragements these last three years.

Last but not least, special words of appreciation go to my dog Newton.

Abstract

This thesis presents three empirical analyses that contribute to our understanding of how financial securities and intermediaries facilitate the allocation of resources between firms, consumers, and the government. Each paper investigates a distinct asset class and aspect of the resource allocation problem. The first paper explores the role of information in financial markets and the wisdom of crowds. The analysis uses a recently developed statistical technique to infer overlap in information sets of security analysts and develops a profitable trading strategy using analyst stock ratings. This finding highlights the prevalence of information diversity. The second paper focuses on government debt managers in national treasuries and finds that they actively choose the maturity structure of government bonds to minimize the expected costs of debt service while responding to demographic factors so as to enable risk-sharing across generations. The third paper leverages the new phenomenon of equity crowdfunding campaigns to disentangle entrepreneurial choices from investor preferences, differentiating the fundraising strategies that entrepreneurs pursue based on their experience and gender, and the funding outcomes they achieve. Taken together, these three studies contribute to our understanding of how intermediaries and securities allocate resources in various financial settings.

Contents

1 Introduction **1**

2 General Literature Overview **19**

2.1 Information diversity and security analysts 19

2.2 Debt management and catering to clienteles 28

2.3 Entrepreneurial finance and equity crowdfunding 32

3 Can Investors Profit from Information Diversity?
The Wisdom of Crowds in Security Analyst Recommendations **41**

3.1 Introduction 41

3.2 Statistical aggregation methods 48

3.2.1 Measurement error framework 49

3.2.2 Optimal combination of a time-series of forecasts 50

3.2.3 Novel approach to combine a cross-section of forecasts 52

3.3 Revealed aggregate recommendation score of security analysts 58

3.3.1 Desired data properties and descriptive statistics 60

3.3.2 Deducing the information sets: an example 63

3.4 Portfolios formed on consensus and revealed aggregate recommendation scores 67

3.4.1 Portfolio construction 68

3.4.2 Test for information diversity 70

3.4.3 Portfolio performance evaluation 73
3.4.4 With transaction costs 77
3.4.5 Information and skill 79
3.5 Simulated model with information diversity and noise 81
3.5.1 Information percolation 82
3.5.2 Calibrated simulation 84
3.5.3 Assumptions and sensitivity analysis 87
3.5.4 The context of security analysts 91
3.6 Conclusion 92

4 The Impact of Demographic Factors on Sovereign Debt Maturity 95
4.1 Introduction 95
4.2 Definitions, identities, and dynamics 103
4.3 Optimal maturity structure in debt management theories 105
4.3.1 Portfolio theory 106
4.3.2 Tax smoothing and fiscal insurance theory 110
4.3.3 Clientele and liquidity preference theories 114
4.3.4 Inflation credibility 119
4.4 Empirical model of maturity structure 121
4.4.1 Data and operationalized variables 122
4.5 Cross-sectional differences in advanced economies 127
4.5.1 Model and methodology 127
4.5.2 Estimation results 130
4.6 U.S. maturity structure dynamics over time 134
4.6.1 Model and methodology 135
4.6.2 Error Correction Mechanism results 137
4.6.3 Estimation results 141
4.7 Conclusion 142

5 Be Careful What You Ask For:
Fundraising Strategies in Equity Crowdfunding **145**
5.1 Introduction . . . 145
5.2 Data . . . 153
5.2.1 Data sources . . . 153
5.2.2 Variable definitions . . . 155
5.2.3 Descriptive statistics . . . 157
5.3 Empirical results . . . 169
5.3.1 Fundraising strategies . . . 169
5.3.2 Campaign success . . . 172
5.3.3 Fundraising amounts . . . 175
5.3.4 Stopping the campaign . . . 182
5.4 Robustness and extensions . . . 189
5.5 Conclusion . . . 191

6 Conclusion **195**
6.1 Diversity of information in the market for public equity . . . 196
6.2 Demographic clienteles in the market for long-term government bonds 199
6.3 Startup founder choices in the crowdfunding market for private equity 201

A Appendix: Information Diversity and Security Analysts **205**
A.1 Proof of proposition . . . 205
A.2 Derivation of implied weights . . . 207
A.3 When revealed aggregator nests the simple average . . . 208
A.4 Trimming algorithm . . . 210
A.5 Impact of an upgrade . . . 211
A.6 Further observations . . . 213
A.7 Using hard cut-offs . . . 214

A.8 Using historical percentile cut-offs 216
A.9 Using a larger sample 218
A.10 With transaction costs 221
A.11 Sensitivity analysis with the noisy information percolation model . . . 224

B Appendix: Debt Management and Demographics **225**
B.1 Optimal share of long-term debt under the tax smoothing approach . 227
B.2 Descriptive statistics of variables in panel estimation 230
B.3 Specification of GMM models 230
B.4 Selection of the panel model 232
B.5 Specification tests for the panel model 235
B.6 Unit root tests 238
B.7 Bivariate cointegration tests 242
B.8 Trivariate cointegration test 248
B.9 Alternative cointegration 250
B.10 Equilibrium correction mechanisms 251
B.11 Specification tests 253

Bibliography **254**

Chapter 1

Introduction

Financial markets allow resources to flow from where they are in surplus to where they are needed, and this is their key economic role (Merton and Bodie, 1995). Securities and intermediaries, the main components of financial markets, are therefore important means of allocating resources. Each of the three papers presented in this dissertation contributes to this theme by investigating a distinct asset class and a distinct aspect of the resource allocation problem. The first paper is about information in financial markets and is titled *"Can investors profit from information diversity? The wisdom of crowds in security analyst recommendations"*. The second paper examines government debt issuance strategies and is titled *"The impact of demographic factors on sovereign debt maturity"*. The third paper considers a fundamental question in entrepreneurial finance—what determines the amount of money raised—and it is titled *"Be careful what you ask for: Fundraising strategies in equity crowdfunding campaigns."* This introductory chapter sets the common ground between these three papers through the way in which securities and intermediaries allocate resources between firms, consumers, and the government.

Securities and intermediaries as the means of allocating resources between firms, consumers, and the government

One example of how financial securities play a role in resource allocation is when firms issue securities to finance their business projects. Firms most commonly issue debt and equity, either on publicly traded exchanges or through private placements. By issuing such securities, firms can raise capital today against future earnings. This allows firms to undertake investments today that would not be possible if firms were limited to using only cash flows from operations. In case of early-stage companies, these constraints are particularly binding. The availability of entrepreneurial finance is often critical for investment, innovation, and business growth of new firms, which contribute to overall economic development.

Investors can also use securities traded in financial markets—equity, debt, derivatives, etc.—to transfer their wealth from today to the future. By allowing individuals to shift purchasing power, financial markets enable individuals to smooth their consumption over time. One such example is saving for retirement, which shifts purchasing power from high-income periods to low-income periods. Moreover, investors can also use securities to transfer their wealth across different states of the world in the future. Securities like derivatives allow hedging of (at least some) risk and are commonly used as risk-management tools. More generally, the financial market allows investors with different risk preferences to select security types with the risk-return characteristics they prefer.

It is not only individual consumers who use the financial markets. Governments and firms can also engage in investment activities to reallocate their wealth and risk across time and states of the world. When government spending exceeds government revenues, governments can either increase taxes or issue debt. When there is a budget surplus, governments can either lower taxes, buy assets, or pay off debt (if any). Under certain assumptions about the behavior of individuals and the structure of the

economy, the exact method of financing a government's budget deficit or surplus—be it through taxation or assets—does not really matter. This is when new government debt today is likely to translate into higher taxes in the future to cover debt repayments, which would imply that debt today and taxes tomorrow are equivalent. This is the Ricardian Equivalence result. Regardless of whether debt and taxes are equivalent in reality, governments can and do use financial securities. Securities issued by governments transfer and redistribute wealth, across individuals, time, and possibly even states of the world.

Besides mobilizing savings from investors, financial markets also enable individual investors to subdivide and to pool their resources and risk. For instance, by issuing securities, the ownership of the firm is subdivided. Financial securities are the means by which individuals hold their claim on income generated by real assets. Hence the owners of the firm need not be the same individuals as the ones managing it. This ownership-management separation can lead to agency problems that arise when managers do not share the same objectives as shareholders and creditors.

Financial intermediaries facilitate pooling and dissemination of resources, and may also help to overcome some agency problems. Frictions in the financial markets theoretically justify the reason financial intermediaries exist. The most prominent candidate frictions include transaction costs, costs of information acquisition, and costly monitoring of managers which is undertaken to mitigate agency problems. Financial intermediaries provide economies of scale with respect to all three of these frictions since investors no longer need to incur these fixed costs individually but rather pool their costs and share information. Therefore, by minimizing such costs, financial intermediaries could improve resource allocation. The category of financial intermediaries includes banks, insurance companies, investment trusts, mutual funds, brokerage houses, pension funds, and so forth.

As such, financial securities in all main asset classes—equities, debt, derivatives—

facilitate the allocation of resources, as do financial intermediaries. Information is a critical input in the decisions of investors, issuers of securities, and intermediaries. Information essentially determines where resources flow. The first paper presented in this dissertation explores the question of information in financial markets and addresses the information aspect of the resource allocation problem. The analysis focuses on security analysts that work in brokerage houses and the information that they might have. The second paper of this dissertation contributes to the theme of resource allocation from the perspective of the government debt issuer in a national treasury and considers the allocation of risk between debt holders from different generations. This second paper explores debt management with an empirical study of the sovereign debt maturity structure in advanced economies. The third paper of this dissertation expands our understanding of how investors allocate resources to different start-up projects by considering the case of equity crowdfunding. The analysis in the third paper examines the microeconomic interplay between investor preferences and fundraising choices of start-up founders. It highlights the role that each factor plays in fundraising outcomes.

The role of information in allocating resources

For publicly-traded securities, prices are the primary transmission mechanism of information (Hayek, 1945). This is because prices of traded securities are signals that reflect information available at the time, and financial resources flow based on these price signals. Moreover, if prices of traded securities fully reflect all available information, then prices are "efficient" and provide accurate signals for the resulting allocation of resources, according to the efficient market hypothesis (Fama, 1970).

Prices that fully reflect all available information facilitate efficient allocation of resources. For example, consider firms issuing securities in the financial market to finance business projects. Suppose there are two companies with similar risk profiles

but different expected cash flows of their available investments. Investors will bid up the price of a company that at the moment appears to have higher expected cash flows. This allows management to easily finance expanding operations and further investment by borrowing funds or issuing new shares since the company represents a good investment opportunity. A company that appears to be a worse investment opportunity will have lower priced securities, might have to downsize eventually, and may even go out of business. Financial markets with efficient prices that correctly reflect the information about companies' prospects, in this example, allocate savers' resources to investment opportunities that can use these resources best.

Since efficient allocation of resources is a desirable outcome, the question of security prices and understanding whether the financial market is efficient, or inefficient, and when, has received significant attention in finance. In the presence of even a slight amount of information frictions, like costly information, the theoretical notion that prices fully reflect all available information no longer holds: the price system would publicly reveal some private information that informed traders gather, but only partially (Grossman and Stiglitz, 1980). The empirical findings of insider gains, which have been thoroughly documented in the literature, also support the argument that prices do not fully contain all possible information (see Aboody and Lev (2000)). Further evidence is the fact that there is a market reaction to corporate earnings announcements, which suggests that private information on earnings, known to the insiders, was not fully incorporated into the stock's price (see the compilation in Jensen (1978) as another example). Last but not least, it is illegal to trade on insider information. This theoretical and empirical evidence suggests that prices imperfectly reveal all available information, especially private information.

Understanding the interplay between information, financial intermediaries, and how information is reflected in prices of traded securities is thus a prominent area of research in finance. This is the topic that I contribute to with my first paper, by

shedding new light on the information content of stock ratings provided by sell-side equity analysts. I also consider the broader question of how to aggregate information that different individuals might have. When individuals produce predictions, rather than trade on their information, there are many ways to aggregate them. Security analysts produce predictions that are based on the information they have. Hence this paper is also about methods of aggregating heterogeneous forecasts, and how to infer the underlying information that different analysts might have.

Information diversity, the wisdom of crowds, and security analyst recommendations

Security analysts are professionals that analyze companies and the industries in which these companies operate, and attempt to assess what the fundamental value of a firm is. As an aggregate measure of their assessment, they issue recommendations to investors as to whether they should buy, hold, or sell company stock. These recommendations generally reflect the analysts' predictions of the firm's future performance. There is often disagreement amongst analysts, and it is common for some analysts to issue a buy recommendation, and for another analyst(s) to issue a sell recommendation, for the same company. There is a need to aggregate individual predictions, for practical and theoretical reasons.

Aggregating individual predictions is theoretically justified by the wisdom of crowds. Going back to Galton (1907), the wisdom of crowds result is that an aggregate of everyone's predictions will be better than any one individual prediction. When individuals observe equally informative signals of the truth, such that deviations only represent measurement error, then the best way to combine heterogeneous predictions is the simple average.

The simple average (the consensus view) is pervasively used as an aggregate of individual analysts' stock ratings. Yet a simple average is not necessarily an optimal

way to combine individual predictions if some analysts know more than others, i.e., if analysts have different information sets and hence individual analysts' predictions are not necessarily equally informative. I utilize a recently developed statistical technique which allows the researcher to combine predictions optimally. The technique infers overlap in information across agents and produces an aggregate based on the union of everyone's information. It is called the revealed aggregator, developed by Satopää, Jensen, Pemantle, and Ungar (2016). The revealed aggregator requires that agents predict several different outcomes in order to infer the information sets of agents. The key to inferring information overlap lies in the cross-sectional covariance of individuals' estimates across these outcomes. The aggregator will consider individuals who always make the same predictions to essentially count as only one individual since they do not add anything new. In contrast, individuals whose predictions are independent of others, and individuals that at the same time tend to be more bullish or more bearish, will be viewed as bringing new information. This aggregator nests the simple average if individuals are equally well-informed.

In this chapter, I apply this statistical technique to stock recommendations of sell-side analysts. I then test two hypotheses. First, I test whether differences in analyst recommendations stem from information diversity by comparing returns from two portfolios, one formed on consensus and one formed on the revealed aggregator. I find that trading on the revealed aggregator beats trading on the consensus, which provides evidence that analysts are not equally well-informed. If they were equally well-informed, the two aggregating methods would be the same and would not yield different portfolio returns. For the analysis, I use the data from I/B/E/S (Institutional Brokers' Estimate System), for the period Jan 1994–Dec 2016, and rebalance the portfolios weekly.

Subsequently, I also test for abnormal returns in the context of semi-strong market efficiency. If diversity of information concerns public information, which is efficiently

aggregated by the market, then trading on the revealed aggregate should not deliver abnormal returns. What I find is that a long-only trading strategy delivers an alpha of 3-4% on an annualized basis. Net of transaction costs, the abnormal return is 2-3%, statistically and economically significant. Since the stocks in my portfolio have large market capitalization and good liquidity, the result of abnormal returns is unlikely to be the consequence of inefficient aggregation of public information by the market. The finding of abnormal returns instead suggests that some of the differences in information among security analysts could reflect private information that some analysts have and others do not. The novel aggregation method allows better inference of analysts' information sets, potentially enabling trading on private information.

Since there are several additional identifying assumptions behind this revealed aggregation technique, I also examine the sensitivity to these assumptions with a calibrated simulation. This allows me to verify that these assumptions are appropriate for the setting of analyst recommendations. In the simulation, I model the spread of noisy information following Duffie and Manso (2007), generate agents that are not equally well-informed, let them predict a cross-section of outcomes, and compare the performance of the consensus and revealed aggregator when the identifying assumptions are violated. I conclude that likely departures from identifying assumptions in case of security analysts are not critical to the results of the analysis, and can in some cases even be exploited with portfolio design.

As such, this paper sheds new light on the information value of analysts recommendations. It contributes to the literature on the information function of financial intermediaries, which is an important element in facilitating the flow of resources in the financial system. This chapter also highlights the prevalence and economic significance of information diversity as a phenomenon, stresses the usefulness of disaggregated individual predictions, and contributes to the wisdom of crowds idea with a finance application.

Government's role in allocating resources

The question of allocating scarce resources in the economy has been at the heart of the economics discipline for centuries. Market forces, together with the flow of information, facilitate allocation of resources. However, there are instances when the market forces are unable to do so efficiently, leading to a social welfare loss. The standard examples involve monopoly power, externalities, provision of public goods, and so forth. In the financial markets, the well-studied cases of market failure include the problem of moral hazard, information asymmetries, principal-agent relationships, and conflicts of interest. These market failures justify much of the regulation and governance of financial institutions that various branches of the government enforce. Thus there is a theoretical role for the government to facilitate allocation of scarce resources because the resulting allocation can be better than the one obtained through market forces.

This role of resource allocation also matters in the context of governments issuing fixed-income securities. The national Treasury makes several choices for any given level of debt that needs to be issued, and these include the currency, indexation (such as to inflation), and the length of time until repayment of debt, i.e., the bond's maturity. The maturity profile determines the cost of debt service through debt repayment profile. For example, a shorter maturity means a more heavily front-loaded debt redemption profile and a higher rollover risk because any unexpected rise in interest rates will much more swiftly translate into higher costs of debt service for the borrower.

Therefore, one obvious strategy for the national Treasury is to choose the maturity structure of government bonds so as to minimize the cost of servicing the debt, while avoiding excessive rollover risk. This is the flip side of the problem faced by a risk-averse investor choosing his/her securities to maximize the expected return, for a given level of risk. Following the Markowitz (1952) portfolio theory is one way to find

the optimal asset allocation to finance the government.

However, the government does not operate in partial equilibrium where the prices of securities are determined outside the system. Under mild deviations from the Ricardian equivalence result of Barro (1974), such as distortionary taxation rather than lump-sum taxation, there is theoretical scope for maturity structure to affect the real economy by smoothing taxation over time, which in turn leads to a smooth consumption profile of a representative agent and is welfare-improving.

In general equilibrium models, besides assuming lump-sum taxation, we typically also assume infinitely-lived agents. Deviating from this assumption, consider an environment with two or more overlapping generations. The young generation saves by purchasing new securities, which generate income later in retirement. In this environment, the two generations don't share risk perfectly. Young savers are the ones who buy newly-issued bonds and thus purchase the risk of an uncertain tax burden in the future, while the old retirees do not purchase newly-issued bonds nor this risk. For the government, issuing debt imposes more risk on the young generation. Hence it is possible for the debt issuer to internalize this risk and improve welfare by designing its debt issuance with intergenerational risk-sharing in mind.

Furthermore, debt issuance is fraught with a time-inconsistency problem since debt holders face the risk of surprise inflation and the possibility that debt will be inflated away. Most debt issued is in fact nominal rather than inflation-indexed. It may be welfare-improving for the debt manager to show commitment to maintaining low and stable inflation and the maturity choice of debt is one way of achieving this.

As such, government debt issuance and the choice of maturity structure can, in theory, be used to improve the allocation of resources in the presence of frictions like distortionary taxes, multiple overlapping generations, and surprise inflation. Just as the government can regulate financial institutions by stepping in to counteract market failure in the financial system, there is also theoretical scope for the government to

use financial instruments to affect the real economy in a welfare-improving manner, besides merely selecting assets that optimize the risk-return profile of outstanding debt. What do debt managers in national treasuries do in practice? The second paper of this dissertation conducts an empirical study of the sovereign debt maturity structure to examine this question.

The government debt asset class is also important because of its economic size and popularity among investors. In the U.S. alone, the Treasury market as of 2016 was $14tr, comparable to the $40tr corporate equity market and $8tr corporate bond market (SIFMA, 2017). Short- and long-term government debt is commonly used in the investment management practice as the asset with the lowest risk, which helps explain the size of the market.

Government debt maturity structure and demographic factors

In this paper, I review formal models of optimal maturity structure and provide reduced-form specifications for determinants of maturity structure. This allows me to comprehensively test alternative theories as competing hypotheses: (1) Are government debt managers choosing the maturity structure so as to minimize the cost of servicing the debt? (2) Do government debt managers attempt to smooth consumption and taxes with their choice of the maturity structure of debt? (3) Do government debt managers pick the maturity structure so as to improve sharing of risk between the generation of savers and the generation of retirees? (4) Are government debt managers using the maturity structure to maintain an anti-inflationary stance? It is possible that government debt managers follow any of these theoretical approaches and associated actions when selecting the maturity structure of government debt.

I use the panel setting with advanced economies that are members of the OECD (Organisation for Economic Cooperation and Development), 2000-2010, to best match the theoretical assumption of steady-state maturity structure. I thus estimate what

explains cross-country differences. I also use the time-series setting to model time-varying maturity structure, using the data for U.S., 1973-2013. I leverage unit root properties of the maturity structure and estimate theoretically-consistent long-run and short-run relationships with an equilibrium correction mechanism.

What I find is that debt managers appear to be pursuing cost minimization, in line with standard portfolio theory. However, as part of the empirical analysis, I also examine the role of demographic factors and institutional investors, such as pension funds and insurance corporations. In line with the theory of optimal maturity by Guibaud et al. (2013), modeled in an environment with multiple overlapping generations, I find that debt managers cater to demographic clienteles through their choice of maturity structure. This catering is meant to enable intergenerational risk-sharing.

The idea of efficiently sharing risk across generations through different financial securities is also representative of the overall role of financial securities for the real economy: efficient allocation of resources from where they are in surplus to where they are in a shortfall. In the case of government debt management and the risk of an uncertain future tax burden, the task of facilitating risk-sharing across generations is in the theoretical remit of the central planner, rather than the task of a decentralized security price. The second chapter thus contributes to the theme of resource allocation by means of financial instruments from the perspective of the government debt manager in a national treasury in the context of an economically significant asset class.

Entrepreneurial finance, innovation, and growth

So far we have considered how financial instruments and intermediaries facilitate allocation of resources in the context of public equities and government bonds. The third chapter of this dissertation is about the allocation of resources in the private equity asset class, which has grown significantly in recent years. This chapter sheds new

light on how investment resources are allocated to early-stage companies by looking at the new phenomenon of equity crowdfunding. In the UK, equity crowdfunding is now the second most popular type of financing for privately-listed companies, after the traditional venture capital (Halmari et al., 2017). Equity crowdfunding platforms are Internet-based electronic platforms that allow entrepreneurs to raise funding from the "crowd", i.e., investors who can individually invest in a company's campaign, in return for equity shares.

Besides the fact that this is a growing market and a recent phenomenon, financing early-stage business ventures is an area where the availability of finance and types of financial systems arguably have the most significant potential impact on the real economy. This idea goes back to Schumpeter (1934), who contended that financial intermediaries provide services that are essential for technological innovation and economic growth. These services—mobilizing savings, evaluating projects, managing risk, monitoring managers, facilitating transactions—channel the capital to entrepreneurs that are most likely to provide innovative products and processes. There is ample evidence to support this contention, in favor of financial systems promoting economic development through entrepreneurs and innovation.

Entrepreneurial finance therefore plays a critical role in the allocation of capital to different investments. Financing methods that are available to entrepreneurs that start new business ventures traditionally include debt and equity (in the form of bank loans, angel investments, and venture capital). Clearly, the outcomes of early-stage finance are important: how much money different projects raise, what are the success rates, how much do the companies innovate and contribute to economic growth, is the funding being allocated to projects with the best prospects, and so forth. However, the observed outcomes are not singularly due to market forces from the side of investors. Entrepreneurial finance is a market where many parties meet—the investors and the founders of the company—and they interact. The microeconomic aspect of the

resource allocation problem in entrepreneurial finance is the interaction between the choices that investors and the founders make, and how investors and founder infer information from the actions of the other party.

Start-up founders choose how much money to raise, when to raise it, the type and the method of financing, and so forth. Some of these choices might also be strategic, in anticipation of investors' interest. While we know quite a bit about outcomes in entrepreneurial finance, relatively little is known about how these outcomes are affected by founders, investors, or both, and the factors that can be attributed to each. For example, vast literature examines the role of prior entrepreneurial experience and gender in raising funds. We know that experienced founders receive more funding, while female founders receive less. Yet an analysis of these questions is limited by the fundamental challenge of investigating causality. If we observe that a particular founder characteristic is associated with lower investment amounts, is that because the investors wanted to invest less, or because the entrepreneurs asked for less?

The third paper contributes to this topic by looking at the microeconomic interaction between company founders and investors on a major UK crowdfunding platform. This analysis sheds new light on how resources are allocated to early-stage companies. Placed in this economically significant area of entrepreneurial finance, this chapter provides new insights into entrepreneurs' fundraising strategies: how much money do founders ask for and how long do they keep their fundraising campaigns open.

The analysis is a joint paper with Professors Thomas Hellmann and Nir Vulkan, both at Saïd Business School, University of Oxford. I have made a substantial contribution to research design, data collection, analytical work, and writing of this co-authored paper. The order of authors is alphabetical, which is the standard practice in the area of finance. Throughout the dissertation, the word "we" refers to our collective work.

Investors' choices and entrepreneurs' fundraising strategies in equity crowdfunding campaigns

Equity crowdfunding provides an environment where we can observe the interplay between market selection of start-up ventures and founder choices. We don't observe this kind of interplay in venture capital financing, where the negotiation is over the valuation of the business and how much equity the founders will be giving away, for a set investment amount. In venture capital, the financing strategies that founders pursue are difficult to pin down.

In this third paper, we consider the relationship between characteristics of entrepreneurial teams, their fundraising strategies, and outcomes of crowdfunding campaigns. We ask the following three questions: (1) What is the specific role of experience and gender? (2) How do entrepreneurs make strategic fundraising choices? (3) To what extent are fundraising success and investment amount driven by the founder or investor choices? The equity crowdfunding setting and the empirical research design allows us to disentangle founder choices in terms of the initial ask and when to stop the campaign, from investors' preferences for shares in different companies. This also allows us to consider how founders might be designing their fundraising strategies in anticipation of investor interest, and how they adapt their strategy once they see the realized investment flows.

To address these questions, we use proprietary data from SEEDRS, a UK-based equity crowdfunding platform that is considered one of the first successful platforms in the newly emerging area of FinTech. The data contains details on the entire fundraising process for the period July 2012 (when SEEDRS started) to June 2017. To obtain further information on founder characteristics, we augment the data with information available online on Linked-In, resulting in complete information on 823 campaigns of which 339 (41%) were successful in raising at least as much money as the campaign goal they asked for. An important aspect of SEEDRS campaigns is

that, after reaching a minimum campaign goal, entrepreneurs decide when to close the campaign, thus determining the amount of money they eventually raise. This stopping decision provides a unique empirical setting for studying the interplay between entrepreneurial choices and investor interest, something that cannot be studied in more traditional entrepreneurial finance datasets.

The analysis suggests that the initial ask of the founders is a crucial determinant of final investment outcomes. In terms of different founders pursuing different strategies, we find that experience and gender makes a difference. We find that experienced teams pursue more ambitious fundraising strategies by asking for more, are more likely to be successful, and raise more money, while female teams (especially female-only teams) pursue safer fundraising strategies by asking for less, are equally successful, but raise less money. We control for variation due to different industries and capital intensity of business ventures, and a host of other calendar time factors. We also find evidence that founders anticipate investor interest but accommodate it in different ways. For example, all-female teams not only pursue safer fundraising strategies, but they also hold out for more money while facing lower investment flows after they have met their campaign goal target. As such, we can address an old question about which entrepreneurs raise more funding by disentangling causality. The analysis also points to the importance of fundraising strategies for determining campaign outcomes. "Be careful what you ask for", you might just get it!

Although each paper of this dissertation investigates a distinct asset class and a distinct aspect of the resource allocation problem, all three papers contribute to our understanding of how financial securities and intermediaries facilitate the allocation of resources between firms, consumers, and the government. The first paper does so by exploring the role of information and financial intermediaries. The second paper does so by exploring the role of government, and the risks that different generations of debt holders bear. The third paper examines the equity crowdfunding market

and how supply and demand factors interact to determine the eventual allocation of investment resources.

Chapter 2

General Literature Overview

As discussed in the preceding chapter, each of the three papers in this thesis extends our understanding of how financial intermediaries and securities facilitate the allocation of resources in the economy. This chapter provides an overview of the pertinent literature. Since each paper investigates a distinct asset class and aspect of the resource allocation problem, this chapter sequentially presents the general literature overview for each paper.

2.1 Information diversity and security analysts

The question of how agents form expectations with respect to the information around them has received substantial attention in the disparate literature. In macroeconomics, the work of Lucas (1973) laid the foundations for rational expectations using the idea that agents face a signal extraction problem when figuring out information from prices. Yet there are many other conceptual frameworks that break with the assumption of rational and computationally unconstrained agents, in the behavioral, learning, and robust control literature. These rely on a variety of information frictions. For example, one commonly-used conceptual framework that explains the expectation formation process is sticky information, i.e., information delay—information is

available without error but with a time lag (Mankiw and Reis, 2002).

An alternative information friction is that of rational inattention (Sims, 2003, 2006). The rational inattention literature relies on the notion that people have a limited capacity to process information and models it with a Shannon filter. Intuitively, some of the information gets filtered out. The notion of a capacity filter is also supported by the psychology and science literature. For example, according to Marois and Ivanoff (2005) and Moscoso del Prado Martín (2011), the human brain has a hundred billion neurons and several hundred trillion synaptic connections that can process and exchange information in a matter of milliseconds, but we cannot easily multi-task or understand more than two people talking at the same time. Sims (2010) further utilizes a Shannon capacity filter in order to explain why some agents do not perfectly use freely available information.

A large body of finance research also finds evidence that investors have limited resources and capacity to collect, analyze and trade on relevant information, which leads to asset prices incorporating this information gradually, rather than instantaneously (Cohen and Frazzini, 2008; Barber et al., 2008; Dellavigna et al., 2008). The idea of cognitive costs is also the central theme behind the findings of return predictability from easy-to-analyze firms to complicated firms (Cohen and Lou, 2012). These asset-pricing implications are supported by theoretical work by Merton (1987) and Hong and Stein (1999). Acquisition of information with limited frequency and accuracy may be rational in presence of information costs (Huang and Liu, 2007). In an extension of the work on rational inattention, a growing literature has also incorporated other variants of noise and information costs in order to better reconcile theoretical predictions with the micro and macro evidence on prices (Woodford, 2009; Mackowiak and Wiederholt, 2009).

In centralized markets, noisy dispersed information is aggregated through prices (Hayek, 1945; Fama, 1970; Grossman, 1981; Diamond and Verrecchia, 1981). Yet

in reality, many of the markets are decentralized. As a result, another strand of research examines information transmission in decentralized markets. One theory is of information percolation, akin to word-of-mount communication, where different agents reveal information to each other through random bilateral meetings (Duffie and Manso, 2007; Duffie et al., 2009). This gives rise to some agents that know more than others at any given point in time.

The way information is aggregated in decentralized markets has also been explored in the context of common value auctions, voting and information cascades. Sometimes this literature relies on the notion that signals are generated by a process, whereas other times this literature relies on the idea that agents construct predictive models that produce signals. In other words, signals are interpreted by agents, and two agents might interpret the same information differently (Hong and Page, 2009). These are the sorts of information frictions and information segmentation that have been explored in the literature. Building on these studies, the first paper presented in this dissertation, in Chapter 3, contributes an analysis that is conceptually motivated by the information frictions described above—that some individuals potentially know more than others.

On the empirical side, we know that information frictions and information segmentation is a common occurrence. Coibion and Gorodnichenko (2015), for example, show that information rigidities (due to rational inattention, sticky information, or noisy information) lead to the rejection of the full-information rational expectations hypothesis. Veldkamp (2006) shows that costly information and information markets can explain the observed cross-sectional and time-series properties of co-movement in different asset prices. Van Nieuwerburgh and Veldkamp (2009) uses information frictions to explain the home bias puzzle. Other literature has documented abnormal returns around delayed or masked information disclosure using the example of institutional investors' confidential holdings (Agarwal et al., 2013) and the example of

asset managers cloaking trades to mask truly informed positions (Cohen et al., 2016), in support of information frictions and segmentation. In a similar vein, Andrei and Cujean (2017) examine word of mouth communication and argues that it can explain the presence of both short-term momentum and long-term reversal, a phenomenon that is pervasive in finance but difficult to reconcile without incorporating behavioral biases. Hong et al. (2005) also provide evidence in support of word-of-mouth communication. Huang et al. (2016) quantify the speed of such communication by considering cross-industry mergers and acquisitions as a plausibly exogenous event.

In more general terms, the question of information aggregation in the presence of dispersed information and multiple actors sits squarely in the literature that examines the wisdom of crowds phenomenon. Going back to Galton (1907), the wisdom of crowds is a finding that an aggregate of everyone's predictions will be better than any one individual prediction, so long as (1) people form opinions independently, (2) there is diversity of information among individuals, and (3) they use this information when making their predictions. One of the main questions is about the mechanism to aggregate individual predictions (Ray, 2006; Surowiecki, 2005). There is substantial evidence in favour of crowd choices being "wise" in reality. Chalmers et al. (2013) examine mutual fund holdings data and how individual investors collectively react to economic signals when determining asset allocations. They find that individual investors react in a sensible manner and should not be dismissed as mere noise traders. Mollick and Nanda (2016) examine funding decisions for proposed theatre projects and find that crowd judgment was not substantially different from expert judgment. Nonetheless, it might be possible to select a meaningful sub-set from the crowd by identifying experts through past performance, as in Budescu and Chen (2015). In a related vein, Prelec et al. (2017) show how to extract the opinion of minority experts when the majority vote is incorrect.

The analysis presented my first paper is also complementary to the literature

which documents disagreement in beliefs and explores asset-pricing implications. The premise of this literature is that differences of information cannot exist in equilibrium, based on the no-trade theorem. However, differences of beliefs can, and are important to understand the non-zero trading volume (Varian, 1989). The analysis in my paper primarily relies on differences of information due to information frictions, which might be short-lived. Furthermore, while the literature on heterogeneous beliefs primarily explores asset-pricing implication, my analysis tackles the question of aggregating disagreeing predictions in a manner that is not through the price mechanism.

The theoretical modeling of differences in beliefs includes discrete-time techniques (Varian, 1985; Harrison and Kreps, 1978) as well as continuous-time models first developed by Williams (1977) and most recent developments including Basak (2005), where agents' incomplete and symmetric information gives rise to heterogeneous beliefs, which have an impact on security prices. There is also support for these theories in that differences of opinion seem to matter for price drift post-earnings announcements (Zhang, 2006; Verardo, 2009), options prices and volume (Buraschi and Jiltsov, 2006), as well as international stock prices and exchange rates (Pavlova and Rigobon, 2007). Heterogeneous belief models require an additional assumption that investors "agree to disagree" or have "differences of opinions" about these beliefs so as justify the empirical phenomenon of a price drift (Banerjee et al., 2009), and disagreement about the interpretation of public information so as to justify the observed trading volume Banerjee and Kremer (2010). Nonetheless, empirical evidence suggests that investors do condition on prices to infer the private information of others efficiently, and thus investors agree on the interpretation of signals, rather than "agree to disagree" (Banerjee, 2011), and there is support for heterogeneous beliefs models in the bond markets as well (Buraschi et al., 2017). As such, the analysis in my first paper in Chapter 3 is complementary to this literature that documents disagreement in beliefs.

Additionally, my analysis contributes to the large literature examining security

analysts (see Bradshaw (2011) for a most recent survey overview). By now, there is a fairly strong consensus in the literature that analyst earnings forecasts are better than time-series forecasting models of corporate earnings (Fried and Givoly, 1982). Analyst earnings forecasts have been commonly measured to calculate an earnings surprise, which has been found to be correlated with publicly available information. Evidence of both under-reaction and over-reaction has been documented, supporting the conjecture that analysts do not fully use all available information (De Bondt and Thaler, 1990; Easterwood and Nutt, 1999). Besides the examination of corporate earnings, more recent literature has looked at other analyst outputs, such as stock picks, as well as analyst activities within the context of individual and employment incentives.

One strand of this literature has focused on the information content of analyst stock recommendations, which are public information. The motivation is the seeming contrast between academic theory and Wall street practice: according to semi-strong market efficiency hypothesis, investors should not be able to trade profitably on publicly available information such as the analyst stock recommendations; yet research departments of brokerage houses continue employing many security analysts. There are many ways to consider trading on security analyst recommendations—level, changes (upgrades, downgrades), and so forth. The literature is fairly conclusive that trading on the level of consensus recommendation scores of sell-side analysts is not profitable once transaction costs are taken into account. For example, using portfolio analysis with daily rebalancing, Barber and Odean (2001) show that annual abnormal gross returns in excess of 4% are possible, but once the large trading volume is taken into account, abnormal net returns are not reliably greater than zero. In a similar vein, Jegadeesh et al. (2004) show that after controlling for return predictability of other signals, the marginal predictive ability of the level of analyst recommendation is not significant since stocks recommended by analysts are positive momentum, high

volume, and high growth type of stocks. There may of course be positive (negative) returns associated with favorable (unfavorable) changes in analyst recommendations, including a long-term price drift following recommendation changes, as documented by Stickel (1995) and Womack (1996). In fact, there is some evidence that quarterly change (not level) in consensus recommendations is a robust return predictor that appears to contain information orthogonal to a large range of other predictive variables, as shown by Jegadeesh et al. (2004).

Overall, it remains a puzzle whether and how analysts add trade value through their stock recommendations—either because they bring information to market through recommendation changes, or because they create their own short-lived price momentum as opinion makers. Ivkovic and Jegadeesh (2004) study revisions of one-quarter ahead stock earnings that analysts forecast, finding that: (1) revisions are least informative in the week after earnings announcements; (2) upward revisions are very informative in the week before earnings announcements; and (3) downward revisions and recommendation downgrades are not. In a similar vein that highlights the information analysts might bring is the finding that analysts with school ties to senior corporate officers have an information advantage (Cohen et al., 2010). Specifically, this amounted to a 9.9% annual premium before Regulation FD in the U.S., when selective disclosure by firms to a subset of market participants was possible. On the other hand, Loh and Stulz (2011) study stock-price reactions following analyst recommendation changes that are major and visible—calling such a recommendation change influential—and find that influential changes come from analyst leaders, stars, and previously influential analysts.

Recommendation changes that are issued away from consensus, accompanied by earnings forecasts, and issued on growth, small, high institutional ownership, or high forecast dispersion firms have a visible stock-price reaction (Loh and Stulz, 2011). The finding of a larger stock price drift following recommendation changes that are issued

away from consensus links to the large literature on herding behavior. For example, in a recent analysis, Jegadeesh and Kim (2010) propose a new test of whether analysts herd around the consensus recommendation, and find that they do herd, and this is not exclusively due to common information that analysts have. The fact that there is a larger stock price drift following recommendation changes that are further away from the consensus suggests that the market recognizes analysts' tendency to herd. In a related vein, Sadka and Scherbina (2007) investigate mispricing that results from mistaken beliefs and contends that the reason mispricing persists in the long-run is due to high trading costs, a contention that is illustrated using analyst disagreement.

Another strand of literature on analysts considers the incentives they face as individuals and employees, including motivations pertaining to sales, conflict of interest, and career concerns. For example, studies have documented a bias in analyst recommendations towards positive "Buy" and away from "Sell". Various analyses suggest that the resulting loss in forecast accuracy from optimism is intentional, representing self-selection (McNichols and Brien, 1997), so as to curry favor with managers (Francis et al., 1993). Other analyses contend that this bias is linked to sales incentives and investment banking services and relationships (Lin and Mcnichols, 1998; Michaely and Womack, 1999; Hong and Kubik, 2003). Other scholars in this literature consider other aspects of career concerns, such as forecasting ability. We know that analyst accuracy improves with firm-specific experience (Mikhail et al., 1997), and is negatively related to job turnover (Mikhail et al., 1999), so is an important element to analysts. Furthermore, the stock-picking ability is persistent (Mikhail et al., 2004). However, Hong and Kubik (2003) show that optimism matters more than accuracy for promotions and demotions at investment banks, underscoring the role of the broader financial intermediation context that analysts are employed in.

As such, there is an extensive literature on security analysts, which finds that it is not profitable to trade on the level of their publicly-issued consensus recommendation,

and that analyst recommendations are biased. It is interesting to re-examine this standard result in the context of information diversity, with a recently developed forecast aggregator that infers differences in information across agents and combines them more efficiently than a consensus (simple average) method. Thus, my analysis in the first paper also contributes to the literature on the combination of forecasts with a finance application that utilizes a recently developed framework of partial information and revealed aggregation (Satopää et al., 2014, 2015, 2016). This technique relies on ideas that are similar to the established method for optimally combining time-series forecasts by Timmermann (2006). One noteworthy application of Timmermann's optimal time-series forecast combination method is forecasting the equity premium with improved out-of-sample predictability (Rapach et al., 2010; Rapach and Zhou, 2013). The main difference with the revealed aggregator, however, is that this is a cross-sectional setting: there are numerous agents predicting numerous outcomes, rather than predicting the same outcome repeatedly over time. The technique relies on overweighing more bullish and bearish agents and also shares the sample intuition as Active Share measure for mutual fund performance developed by Cremers and Petajisto (2009).

As discussed above, the first paper in this dissertation is presented in Chapter 3, and it contributes to the large literature in many different fields of economics and finance. This literature ranges from expectations, information frictions, information segmentation and heterogeneous beliefs to the literature on security analysts and the environment they operate in, to another large literature on econometric and statistical techniques for forecast combination, statistical aggregation. Last but not least, the analysis broadly complements the literature on the wisdom of crowds, and also the informational role of security prices.

2.2 Debt management and catering to clienteles

While my first paper on information diversity and security analysis touches upon a range of topics within an established or flourishing literature, the second paper on debt management and demographic clienteles sits in a more niche area, between macroeconomics and finance. The existing literature on government debt management is fairly scarce. Many papers have included occasional stylized facts about the maturity structure of debt (Eichengreen and Hausmann, 1999; Broeck and Guscina, 2011; Abbas et al., 2014) or have focused on individual determinants in line with specific theories (Guibaud et al., 2013; Missale and Blanchard, 1994; Missale, 1999). To date, there has been no comprehensive empirical research on determinants of maturity structure in advanced economies. My second paper, in Chapter 4, is the first to include such a comprehensive analysis that treats structural specifications seriously when considering many competing hypotheses about optimal maturity structure, filling this gap in the literature.

The empirical papers with analysis closest to mine are Goudswaard (1990) and Gonzalez-Fernandez and Gonzalez-Velasco (2014). Goudswaard (1990) was the first to estimate determinants of maturity structure, with the scope of analysis limited to the Netherlands in 1960-1985 and without structural foundations of determinants. Gonzalez-Fernandez and Gonzalez-Velasco (2014) build on Goudswaard (1990) and Guibaud et al. (2013) so as to focus on the demographic determinants of maturity structure that are linked to the savings life-cycle. However, their paper treats interest rates, inflation, and debt as control variables rather than as competing hypotheses.

In order to derive reduced-form determinants of maturity, I rely on several formal theories, which have been briefly discussed in Chapter 1. The first theory is portfolio theory, set in partial equilibrium, where debt manager is risk averse and chooses maturity structure so as to minimize the cost of servicing the debt while avoiding excessive rollover risk. This has been formalised by Dottori and Manna (2016) and

is the flip side of the portfolio choice problem faced by a risk-averse investor who wishes to maximize the expected return subject to a certain level of risk, seminally introduced by Markowitz (1952).

However, any starting point for discussing debt manager's choice of maturity structure and its effects on the economy as a whole has to deal with failures of the Ricardian equivalence result of Barro (1974). In a representative-agent model with lump-sum taxes, the level of debt and its maturity structure do not matter for the real economy. However, without the representative agent assumption, or with distortionary taxation, both the level of debt and its maturity structure can be used to smooth taxation and therefore consumption profile of agents in the economy over time, and over states of nature. A smooth consumption profile of a representative agent is desirable for a welfare-maximizing planner. The literature on optimal debt structure that provides tax smoothing shows that the first-best option is state-contingent debt (for example, contingent on expenditure or GDP), according to Barro (2002). The second-best option is indexed perpetuities, i.e., consols of infinite maturity.

Curiously, neither state-contingent debt nor perpetuities are currently extensively issued by governments, if at all. Some subsequent literature has focused on the optimal policy when governments cannot issue state-contingent debt. Yet there are some limitations when it comes to practical application of the theories. For example, Lucas and Stokey (1983) as well as Angeletos (2002) and Buera and Nicolini (2004) demonstrate that state-contingent debt can be synthetically replicated with non-contingent debt. However, the practical policy application is to issue long-term debt and invest in short-term assets (Nosbusch, 2008), which requires enormous debt positions in order to work as a hedge, and is implausible. The policy prescriptions from this literature are also not robust to informational imperfections and financial frictions (Faraglia et al., 2010), and lack of commitment by government (Debortoli et al., 2014).

Extending the intuition in Barro (1979), Bohn (1990) follows with a general equilibrium model that features distortionary taxation and a stochastic environment. He shows that Barro's approach—whereby the government should smooth tax rates over time—generalizes to tax smoothing over states of nature, by means of maturity structure instead of the debt level. In this paper, I contribute to this strand of the theory on optimal maturity structure by testing whether this tax-smoothing consideration is a statistically significant determinant of the maturity structure of government debt in advanced economies.

Besides the environment with distortionary taxation, Ricardian equivalence also fails in models with overlapping generations. Prior literature has shown that in an overlapping generations setting, issuing debt can be beneficial for inter-generational sharing of risk (Fischer, 1983; Gale, 1990; Guibaud et al., 2013). The notion that institutional investors—particularly pension funds and insurance corporations—are clienteles in the long-term government bond market goes back to Modigliani and Sutch (1966) and Culbertson (1957). The work by Guibaud, Nosbusch, and Vayanos (2013) builds on it and is central to the findings of the paper presented in this dissertation.

A similar strand of literature considers preferred habitats and clienteles without normative prescriptions and examines asset-pricing implications instead. Vayanos and Vila (2009), for example, develop a formal model where limits to arbitrage arise from risk-averse arbitrageurs that substitute assets across maturities, and each maturity has its own clientele. They show that shocks to demand of particular clienteles affect the term structure and investigate implications for bond risk premia, transmission of monetary policy, and other questions. Further motivated by the question of monetary policy transmission, Greenwood and Vayanos (2010) underscore the relevance of the preferred-habitat view with the case studies of the 2004 UK pension reform and the US Treasury's buyback program which took place in 2000-2001. Subsequently, Greenwood and Vayanos (2014) examine how changes in debt supply affect bond

yields and returns with a formal term-structure habitat model. The model predicts that an increase in supply should raise bond yields and expected returns, holding the short rate constant. The authors confirm this empirically by looking at high excess returns of long-term bonds that follow increases in maturity-weighted debt-to-GDP.

My analysis of the government debt maturity structure and the extent to which debt managers cater to demographic clienteles also contributes to the recent literature on corporate bond issuance and catering to clienteles in the market for long-term government bonds. Stein (1996) contends that company managers should cater to investor sentiment. For example, in a certain environment, when a stock is overvalued, firms will optimally choose to be over-levered (Stein, 1996, p.443). There is empirical evidence that companies cater to investor sentiment and possible market segmentation. Firms tend to borrow long-term when subsequent long-term bond returns are expected to be low (Baker et al., 2002), consistent with debt managers trying to time the market. Moreover, firms tend to react to shocks in the supply of government debt, issuing more long-term debt when the government issues more short-term debt (Greenwood et al., 2010), which illustrates that firms time the market with their capital structure choices and also act as macro liquidity-providers.

Lastly, this paper also complements the literature on the time-inconsistency aspect of debt management (the fact that debt holders face the risk of surprise inflation, and the possibility that debt will be inflated away) by relying on the optimal maturity structure theory by Missale and Blanchard (1994). This concludes all the theories of optimal debt management that I consider as competing hypotheses in Chapter 4, as well as the related literature on catering to investor sentiment and clienteles that exist in many parts of the financial market. Overall, my second paper on government debt management contributes to several strands of literature at the relatively niche intersection of macroeconomics and finance, ranging from optimal public debt policy and the welfare-optimizing role of the central planner to term structure theory and

pricing of government and corporate bonds. The analysis and findings of this paper are also broadly related to capital structure choices of corporations.

2.3 Entrepreneurial finance and equity crowdfunding

The third paper presented in this dissertation is in Chapter 5. This third paper is situated squarely in the well-established entrepreneurial finance literature. Following the review in Wallmeroth et al. (2018), entrepreneurial finance can be split into several streams: (1) pre-investment and selection evaluation, (2) investment ownership structure and contracting, and (3) post-investment success and failure. Venture capitalists (VCs) were the first to enter the field of entrepreneurial equity finance in the 1940s and have since become the most prominent vehicle. The literature on venture capital is well-established (see Da Rin et al. (2013) for an overview), as is the literature on angel investors, and their interplay. In terms of impact, for example, we know that VCs are particularly important for firms that are innovating (Hellmann et al., 2000), professionalize startup firms in a positive manner (Hellmann and Puri, 2002), and frequently bring benefits through active investment (Bottazzi et al., 2008). Angel financing also has positive effects that are not just monetary (Kerr et al., 2014).

In addition to VC and angel investors, another type of entrepreneurial financing is crowdfunding. The general phenomenon is well established by now, and the academic inquiry into the field of crowdfunding has been growing with the phenomenon itself. It is based on an electronic platform that enables a wide variety of individuals and organizations to raise funding from the general public (the crowd), for a wide variety of causes. Broadly speaking, there are three types of crowdfunding models. First, there is crowdfunding based on donations and rewards. Prominent examples of such platforms include Kickstarter, IndieGoGo, GoFundMe, and CrowdRise. The work of

Agrawal et al. (2014) examines the fundamental economics of this type of financing, and numerous other scholars have looked into various aspects of rewards-based crowdfunding. Second, there are so-called peer-to-peer lending platforms, where individuals and organizations seek debt funding from the crowd. Prominent examples include Lending Club, Proper, and Funding Circle, which are explored by Havrylchyk and Verdier (2018) and Mohammadi and Shafi (2016). Third, there is equity crowdfunding, which is solely available to companies. In equity crowdfunding, the companies are mostly entrepreneurial ventures that are developing novel products and services. The crowd invests on the basis of receiving equity in the company. Examples include AngelList (which is for accredited investors only) in the US and CrowdCube and SEEDRS (which entails no investor accreditation) in the UK, as the main platforms.

One important question in the literature has been whether equity crowdfunding (ECF) is complementing or competing with angel investment that takes places outside of online platforms. Estrin et al. (2017) and Hornuf and Schwienbacher (2016) conclude that ECF provides additionality to sources of entrepreneurial finance. The choice between rewards-based crowdfunding and equity crowdfunding, in turn, is likely driven by the amount of capital that the project requires. Belleflamme et al. (2014), for example, theoretically model this choice and show that pre-ordering is optimal if the initial capital requirement is relatively small compared with market size, while profit-sharing through equity is preferred if the initial capital requirement is large.

In terms of pre-investment and selection evaluation, there is an established theoretical argument that crowdfunding allows entrepreneurs to learn about project quality and estimate the demand for the product based on investor interest. It is not necessarily obvious that crowdfunding is welfare-improving in terms of resource allocation. A large number of investors present in crowdfunding by design will face problems like free-riding, duplication of costs, and an agency problem (entrepreneurial moral

hazard), as suggested in the seminal paper by Diamond (1984). Recent work by Strausz (2017) demonstrates that under aggregate demand uncertainty crowdfunding improves screening for valuable projects, while after-markets allow investors to mitigate the entrepreneurial moral hazard. In other words, the gain in efficiency persists despite the presence of moral hazard and private cost information. Chemla and Tinn (2017) and Ellman and Hurkens (2015), among others, also model the way in which entrepreneurs learn about the value of the project in rewards-based crowdfunding. On the empirical side, Xu (2017) shows that entrepreneurs who receive positive feedback on Kickstarter are later more likely to commercialize their projects. This supports the hypothesis that crowdfunding adds value through early feedback. Havrylchyk and Verdier (2018) also examine the financial intermediation role of peer-to-peer lending platforms through a similar perspective.

Nonetheless, perils of crowdfunding include individual investors underestimating default rates when compared to professional investors, as documented by Mohammadi and Shafi (2016). Thus another important question in the literature is how rational the investors are. Using the data from Crowdcube, a UK equity crowdfunding platform, Estrin and Khavul (2015) explore whether investors exploit the complementarity of signaling and network effects in a reduced transactions costs environment that is possible due to the architecture of equity crowdfunding platforms. The authors find that the dynamic process in the supply of funds is responsive and adaptive to new information but not explosive, suggesting that entrepreneurs and investors exchange soft and hard information.

While pursuing a similar question about the rationality of investors, Åstebro et al. (2017) examine investor herding, whether it is rational, and whether it leads to information cascades in the sense that investors might be ignoring personal information in favour of observed choices of other investors on the platform. The authors use the data from SEEDRS, another major UK equity crowdfunding platform. They

show that, both in the model and in the data, the size and likelihood of a pledge are affected positively by the size of the most recent pledges and negatively by the time elapsed since the most recent pledge. They contend this is inconsistent with naive herding, independent investments, or exogenously correlated investments. This evidence supports their conjecture that investor herding in equity crowdfunding is rational so long as there are some investments coming; however, social learning does stop once investments have stopped due to herding. Another study of information cascades using the equity crowdfunding setting (Vismara, 2016b) empirically finds a prominent role of investors with a public profile. This analysis complements the work by Chen et al. (2016), who highlight the importance of hybrid investor crowds (where members are led by an expert investor) in the presence of information asymmetry and high project risk.

Similarly, numerous scholars focus on signalling from the entrepreneur's side and the impact of networks. Ahlers et al. (2015), for example, use the campaigns from Australian Small Scale Offerings Board and find that when entrepreneurs retain equity and provide more detailed information about risks, this can positively impact the probability of funding success since these aspects constitute effective signals, whereas social capital and intellectual capital, by contrast, do not. Numerous other scholars also find social and public networks and geography to be important determinants of success (Lukkarinen et al., 2016; Vismara, 2016a; Agrawal et al., 2015; Mollick, 2014). For example, Vismara (2016a) finds that campaign success was associated with selling a smaller fraction of the company and having more social capital, which complements the findings of Ahlers et al. (2015), using the UK rather than Australian campaign data.

The analysis presented in Chapter 5, which looks at the microeconomic interaction between entrepreneurs and investors, builds on this nascent literature in crowdfunding and the large entrepreneurial finance literature. Concerning the entrepreneurs

themselves, an important early contribution to understanding equity crowdfunding is the paper by Bernstein et al. (2017), which uses data from AngelList. The authors use a randomized field experiment and find that the average investor responds strongly to information about the founding team, but not to firm traction or existing lead investors. This underscores the importance of the founder team to investors in early-stage ventures, and our analysis in the paper consequently is focused on understanding entrepreneurial strategies and disentangling them from investor interest.

The question of entrepreneurial experience and human capital has received significant attention in the larger literature on entrepreneurial finance. Our analysis complements and extends our understanding of the role of experience for early-stage finance. We know that there are benefits to being a serial entrepreneur: Gompers et al. (2010) find that around 10% of all VC-backed founders are serial entrepreneurs, and that entrepreneurial success breeds more success, while Hsu (2007) finds that serial entrepreneurs have better chances to obtain venture finance, and obtain better valuations. Colombo and Grilli (2005) look at a broader spectrum of human capital characteristics and find prior entrepreneurial experience to be positively related to company growth. Distinguishing between different types of human capital in greater detail, Gimmon and Levie (2010) highlight the role of experience over education for venture survival, based on the data of high-technology startups in Israel. See Unger et al. (2011) for a useful meta-analysis of the broader related literature. More generally, the analyses of Wernerfelt (1984), Rajan and Zingales (2001), and Rajan (2012) speak to the broader role of human assets in startups. Agarwal and Gort (2002) consider learning by doing and endowments as key characteristics of startup quality in the model of firm life cycle survival. Kerr et al. (2014) and Manso (2016) have looked at how experimentation and innovation plays a central role, the idea that goes back to Schumpeter (1934).

The third paper of this disseration also features an analysis of fundraising strate-

gies by gender, which touches on the vast literature about gender differences. We know from experimental evidence in the literature on attitudes towards risk that women are more risk-averse than men (consider Eckel and Grossman, 2008; Croson and Gneezy, 2009). Using the data on common stock investments of men and women from a large brokerage house, Barber and Odean (2001) document that men trade more than women, which reduces men's net returns by 2.65 percentage points a year as opposed to 1.72 percentage points for women, supporting their conjecture borrowed from psychological research that men are overconfident. Yet there are important nuances regarding selection, as women working in the field of finance may not be as risk-averse as women in the general population. Adams and Ragunathan (2017) actually find that the gender gap in risk aversion reverses in finance: financial firms with more female directors are relatively riskier than non-financial firms. Atkinson et al. (2003) also highlight the finding of self-selection with their finding that male- and female-managed funds do not differ significantly in terms of performance, risk, and other fund characteristics, but that gender does influence the decision-making of mutual fund investors. Studies of corporate finance decisions find that women play it safer (Huang and Kisgen, 2013; Adams and Ferreira, 2009), with a note that this is not necessarily good for preventing take-overs. Specifically in crowdfunding, there is evidence that women investors are more risk-averse (Hervé, 2017).

There is also a large literature that looks into the role of gender specifically in entrepreneurial finance, which the third chapter of the dissertation directly contributes to. Verheul and Thurik (2001) use a panel of Dutch entrepreneurs and find that after correcting for the female profile, female entrepreneurs have a smaller amount of start-up capital, a smaller proportion of equity and a higher proportion of bank loans. Similar conclusions are present in the data from the Kauffman Firm Survey—that women start businesses with a significantly smaller amount of capital and use external debt more (Coleman and Robb, 2009; Robb and Coleman, 2010). Fairlie and

Robb (2009) in turn investigate why female-owned businesses are less successful than male-owned businesses, finding that it is due to less startup capital and less prior business work experience. The gender of investors seems to matter in VC industry as well: in addition to the diversity-related facts documented by Gompers and Wang (2017b) regarding VC funding, there is evidence that female startups financed by VCs that include a female general partner do not experience the negative gender gap in performance (Raina, 2017), while the phenomenon of VC partners parenting daughters is linked to improved gender diversity and improved deal and fund performance (Gompers and Wang, 2017a).

Numerous analyses of gender thus find a negative correlation between female entrepreneurs and investment amounts. A limitation of this prior research is the difficulty in distinguishing demand and supply side forces. The equity crowdfunding context provides a new opportunity to revisit this important question, leveraging the unique features of equity crowdfunding platforms. The discussion of gender in the general crowdfunding literature is more nuanced and focuses on the gender-based matching of entrepreneurs and investors, as well as crowds' perception of women founders. For instance, Marom et al. (2016) find that men seek significantly higher levels of capital than women on Kickstarter, and also raise more funds, yet women enjoy higher rates of success in funding their projects, even after controlling for category and goal amount. This is driven by the fact that female-led projects are predominantly financed by women, and there is a large presence of women investors on the Kickstarter platform. Gender-based matching not driven by startup quality characteristics is also confirmed by Ewens and Townsend (2018) when looking at the AngelList platform. Looking at investor stereotypes and implicit bias in crowdfunding decisions, Johnson et al. (2018) find that investors tend to have an implicit positive bias towards women on Kickstarter, which translates into more positive funding outcomes. The analysis of gender factors in equity crowdfunding is limited, with notable exceptions

by Hervé (2017) and Mohammadi and Shafi (2018). Curiously, Mohammadi and Shafi (2018) find the opposite of matching on the Swedish equity crowdfunding platform FundedByMe—they find that female investors are more likely to invest in companies with a higher proportion of males.

Our analysis in Chapter 5 contributes to these strands of literature in entrepreneurial finance that examine how founder characteristics are related to fundraising outcomes. Moreover, the analysis also employs well-established methods of identification, such as instrumental variables, together with the fact that founder choices and investor demand are at times sequential. This allows us to tease out the extent to which fundraising outcomes are due to founder choices or investor preferences, or both. Therefore, this analysis illustrates the microeconomic dimension of the allocation of capital to startups.

Taken together, the three papers of this dissertation contribute to our understanding of how financial securities and intermediaries facilitate the allocation of resources between firms, consumers, and the government, in different financial settings. The contribution of my research spans many different fields, such as empirical asset pricing, entrepreneurial finance, financial intermediation and statistics literature.

Chapter 3

Can Investors Profit from Information Diversity? The Wisdom of Crowds in Security Analyst Recommendations

3.1 Introduction

There is substantial observed heterogeneity in individual forecasts and estimates for many variables that are used by decision-makers. As illustrated in Figure 3.1, there are differences in expert forecasts of GDP, inflation, interest rates, and currencies. There is also substantial disagreement in recommendations of sell-side security analysts. Among 47 analysts that cover Amazon, 13 analysts strongly recommend to buy it, while two analysts recommend to sell it (Figure 3.2). What can one do with such opposing views? The wisdom of crowds idea, which goes as far back as Galton (1907), is that an aggregate of these recommendations will be better than any one individual prediction. So the question is how to combine them.

The most commonly used statistic to aggregate individual predictions is the consensus view, i.e., a simple average. However, a simple average is only optimal when agents observe equally informative i.i.d. signals of the truth. If individuals each observe signals with different degrees of informativeness, a simple average is generally not optimal. If Jack knows everything that John knows, and some more, then averaging their predictions overweighs their common knowledge and underweighs Jack's unique information. In a setting like this, another type of aggregation method is needed to adjust for diversity and overlap of information. This paper extends the wisdom of crowds idea to such a setting in finance.

Is it plausible that individuals have different and overlapping information sets? Yes, for example in the context of limited information capacity and rational inattention (Sims, 2003). In fact, one's lifetime is not enough to read all the English-language books in print or to see all existing movies. The limiting factor is not just the quantity of information, but also cognitive costs, which have been documented in science[1] and in financial settings (Cohen and Lou, 2012; Cohen and Frazzini, 2008; Barber et al., 2008; Dellavigna et al., 2008). Besides rational inattention, models with sticky information (Mankiw and Reis, 2002) or noisy information (Woodford, 2009; Mackowiak and Wiederholt, 2009) are the alternative theoretical frameworks that incorporate heterogeneity in information and agents' predictions. Partial information sets can also arise if individuals have different sources of information, which they exchange through word-of-mouth communication, as modeled by the information percolation theory (Duffie and Manso, 2007). Heterogeneity in agents' informativeness also arises if individuals interpret the same information differently (Hong and Page, 2009).

[1]See Marois and Ivanoff (2005) and Moscoso del Prado Martín (2011). Although the human brain has a hundred billion neurons and several hundred trillion synaptic connections that can process and exchange information in a matter of milliseconds, we cannot easily multi-task or understand more than two people talking at the same time.

Figure 3.1: Heterogeneity in one quarter ahead forecasts of the Blue Chip Financial Forecasters. All data is in percent, except for the Fed's Major Currency Index. Forecasts made in July 2015, for Q3 2015.

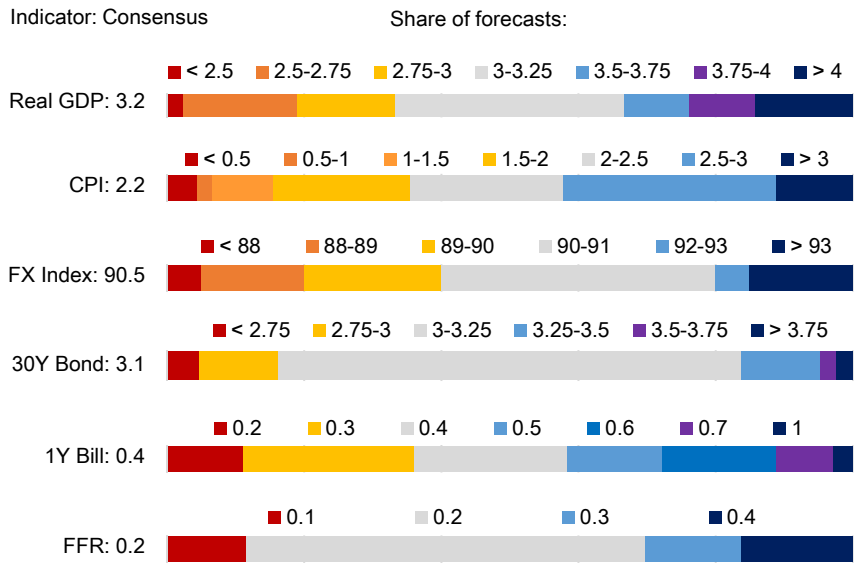
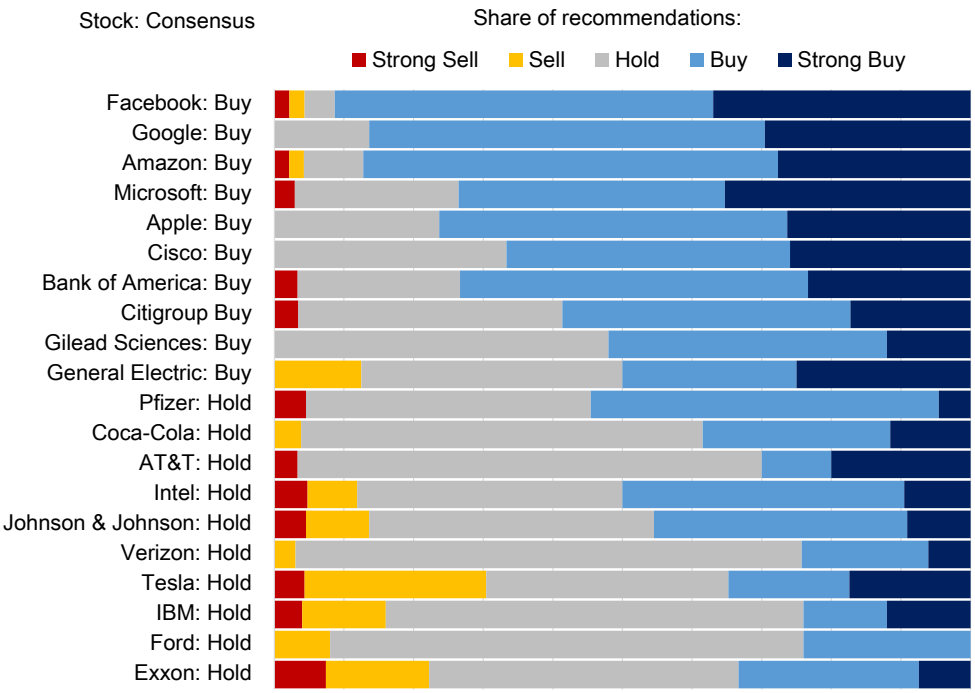


Figure 3.2: Heterogeneity in ratings of sell-side security analysts for top 20 most followed stocks on Seeking Alpha. As of September 22, 2017.



The revealed aggregator by Satopää, Jensen, Pemantle, and Ungar (2016) was developed specifically for this kind of problem—individuals have different, but relevant, information, and some individuals know more than others—such as for the theoretical reasons described above. In the example where Jack knows as much as John, and more, when we consider their predictions to be random variables, Jack will make more extreme predictions with greater probability. In other words, Jack's forecasts will be more bullish or bearish relative to John's forecasts. The revealed aggregator method thus maps the information sets of individuals based on the cross-sectional covariance of their estimates of several different variables. Agents receive greater weight in the final aggregate if they are more bullish or more bearish in the cross-section of forecasts, and simultaneously less strongly correlated with other agents in this cross-section. Once the information overlap between agents, if any, is stripped out, the final estimate is based on the union of everyone's information. The revealed aggregator thus finds many needles in the haystack and efficiently combines them together. In the special case when all individuals are equally well-informed, it nests the simple average.

I apply this statistical methodology to an asset-pricing context of security analyst recommendations and use it to highlight the economic significance of information diversity as a phenomenon that is prevalent, measurable, and tradable. I use stock recommendations of sell-side security analysts because this is a setting where this statistical technique is suitable, and because the information content of analyst stock recommendations is particularly interesting to study. The literature on sell-side security analysts is fairly conclusive that trading on the level of consensus recommendation scores of sell-side analysts is not profitable. Using portfolio analysis with daily rebalancing, Barber and Odean (2001) show that net abnormal returns are not possible. Jegadeesh et al. (2004) show that after controlling for return predictability of other signals, the marginal predictive ability of the level of analyst recommendation is not

significant. It is therefore interesting to re-examine this standard result in the context of information diversity, with a recently developed forecast aggregator that infers differences in information across agents and combines them more efficiently. The applicability of this statistical technique is not limited to forecasts and estimates. It can be extended to any decision that is based on agents' expectations of future outcomes.

Analysts and brokers have many potential sources of information to predict future stock returns. These include models of cash flow forecasts, industry trends, the state of the economy, social media analysis, private stock tips, and so forth. If one broker has five different models, while the second broker only uses cashflow forecasts, the first broker is better informed and will be able to issue more extreme recommendations with higher probability. The first broker's stock forecasts will be more bullish or bearish overall. The intuition of overweighing more bullish or bearish agents is also in line with the Active Share measure for mutual fund performance developed by Cremers and Petajisto (2009). Moreover, there is empirical evidence that that analyst accuracy improves with firm-specific experience (Mikhail et al., 1997), that stock-picking ability is persistent (Mikhail et al., 2004), and that analysts with school ties to senior corporate officers have an information advantage (Cohen et al., 2010). If this or any other kind of information diversity is prevalent, then trading on consensus scores may be suboptimal, trading on the revealed aggregator developed by Satopää et al. (2016) could outperform trading on consensus, and might deliver abnormal returns, without any input about the models used, forecasting skill or school ties.

I use recommendations of sell-side analysts from the Institutional Brokers' Estimate System (I/B/E/S), from large brokerage houses, covering popular stocks listed on the NYSE and NASDAQ. This particular data coverage ensures that analysts in the sample forecast the same stocks, which means that it's possible to calculate a cross-sectional covariance of their recommendations and estimate their information structure. In the final data sample, I use recommendations of analysts from 28-33

brokerage houses that cover 351-595 stocks on average. Every week on a Friday, I calculate the consensus score for each stock, and also use the cross section of recommendations for other stocks to calculate the revealed score for each stock. I form portfolios on the basis of consensus scores, and a different set of portfolios on the basis of revealed recommendation scores. I rebalance these portfolios weekly.

Comparing the performance of the two portfolios over the last ten years, Jan 2007–Dec 2016, I find that trading on revealed aggregate beats trading on consensus. A portfolio formed on stocks with the most favorable revealed aggregate recommendation scores could have earned at least 3% more in annualized terms than a portfolio formed on stocks with the most favorable consensus scores. This finding implies the rejection of the null hypothesis—that all information is common knowledge—and suggests that some analysts do in fact know more than others.

I also find that a portfolio formed on the best revealed aggregate scores delivers an abnormal gross return that is equivalent to about 4% in annualized terms, using standard risk adjustments. After accounting for transaction costs, abnormal returns are still present and economically significant, at 2-3%. This stands in contrast to a zero abnormal net return when trading on the level of consensus recommendation, and provides new insights about positive private information analysts might have.

These findings are robust to alternative ways of trimming the data set for estimation of the revealed aggregator, a longest possible time horizon based on data availability, 1994-2016, and alternative ways of constructing portfolios. I also consider whether analyst skill could account for these findings. I find that there is no relationship between analysts' *Institutional Investor* rating and weights implied by the statistical model. It is unlikely that my results are driven by a measure of analyst skill, and instead capture a range of information content.

It is important to note that besides the notion that some agents know more than others, there are several additional identifying assumptions behind this revealed ag-

gregation technique. Namely, partial information sets should be agent-specific, rather than agent-stock-specific, and agents should make predictions that are unbiased. The statistical framework also abstracts away from the amount of noise in the information that agents receive. In order to check that the setting of analyst recommendations is indeed appropriate for this statistical methodology, I also examine each of the identifying assumptions with a calibrated simulation, which allows me to quantify the forecasting setting. I model the spread of noisy information following Duffie and Manso (2007), generate differentially informed agents, let them predict a cross-section of outcomes, and compare the performance of the two aggregators. First, I confirm that when agents have different but relevant information, and some know more than others, the revealed aggregator indeed beats consensus. Next, I introduce biased forecasts, agent-stock specific information, and vary the amount of noise in the signals in order to test the sensitivity of the revealed aggregator method to these assumptions.

I find that the setting with sell-side analysts holds up to additional identifying assumptions regarding noise and information structure. The calibrated simulation shows that when signal noise is high, the benefit of using the revealed aggregator is less pronounced, but it does not vanish. Since analysts are professional forecasters, they likely have limited to moderate noise in their information signals. However, they may specialize in certain subsets of stocks or may have additional unique information only for select stocks among all the stocks that they cover, which would make their information structure agent-stock-specific rather than agent-specific. The calibrated simulation shows that the revealed aggregator is applicable even in this case.

Yet when there is bias, such as when all agents are optimistic across the board or treat negative signals as neutral, the revealed aggregator fails to beat the consensus. The question of bias is therefore critical for this technique, and prior literature suggests that analyst recommendations are asymmetrically biased due to their career concerns (Hong and Kubik, 2003). It is plausible that analysts reveal positive infor-

mation truthfully, through a favorable recommendation, but not always the negative information. My empirical results suggest that this asymmetric bias, when present, could be addressed and exploited by means of portfolio design. I find that it is a long-only trading strategy where revealed beats consensus and delivers abnormal returns, which suggests that this statistical technique and this strategy design potentially enables trading on analysts' private positive information, akin to finding a needle in a haystack among very many buy and strong buy recommendations that analysts issue.

By showing how information diversity can be empirically tested, and by highlighting its economic significance with a profitable trading strategy, this paper complements the literature that finds pervasive evidence in favor of information frictions and information segmentation (Coibion and Gorodnichenko, 2015; Veldkamp, 2006; Van Nieuwerburgh and Veldkamp, 2009; Andrei and Cujean, 2017; Agarwal et al., 2013; Cohen et al., 2016; Hong et al., 2005; Huang et al., 2016; Cohen et al., 2010). The findings of this paper also build on the literature which documents disagreement in beliefs and explores asset-pricing implications (Basak, 2005; Banerjee et al., 2009; Banerjee and Kremer, 2010; Banerjee, 2011; Sadka and Scherbina, 2007; Buraschi et al., 2017). I illustrate how disagreement among analysts can be used to infer diversity in their information, by looking at individual forecasts across a cross-section of outcomes and combining heterogeneous information more efficiently. This sheds new light on the information value of analysts recommendations, usefulness of disaggregated individual predictions, and the wisdom of crowds more generally.

3.2 Statistical aggregation methods

Consider a setting with $j = 1, ..., J$ agents. An agent j provides an estimate X_j, which is an estimator of some unknown quantity Y, or a forecast of some outcome Y. In everything that follows, forecasts and estimates are interchangeable. Existing

statistical methodologies to optimally combine multiple X_j into a single estimator of Y range from the static measurement error setting to more nuanced methods for time-series forecasts. In this section, I first review these established methodologies so as to clearly demonstrate the contrast with the subsequent discussion of the partial information framework and the revealed aggregate method, which is a novel approach developed by Satopää, Jensen, Pemantle, and Ungar (2016). After a brief explanation of this approach and its identifying assumptions, I describe the model used to estimate the revealed aggregate. In addition, I show how it can be represented as a weighted average of individual predictions, and show that it nests the simple average (equal weights) in one special case.

3.2.1 Measurement error framework

In a standard statistical model, the reason different agents provide different forecasts is only due to noise. An agent observes true outcome Y with an idiosyncratic error. While signals can vary in noise, each signal is equally informative in the sense that each agent has all the information necessary to observe true Y, but can only do so imprecisely, with small mistakes.

$$X_j = Y + e_j$$
$$e_j \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \quad \forall \quad j$$

When agents observe such equally-informative, unbiased, independent, and identically distributed noisy signals of true returns, the optimal combination of all the X_j forecasts is the simple average, $\frac{1}{J}\sum_{j=1}^{J} X_j$. It is optimal because it minimizes the loss that arises from aggregating versus knowing true Y.

3.2.2 Optimal combination of a time-series of forecasts

If noisy signals are not independent, and agents observe correlated noisy signals of true returns, there is a portfolio diversification argument to be made in favor of combining their forecasts (Bates and Granger, 1969). The informal premise is that information sets underlying individual forecasts might sometimes include private information. Hence a super model cannot be readily constructed, and forecast combinations add value through the desirable diversification (or hedging) property. Although this justification informally relies on different information sets, this is not considered explicitly in forecast combination theory.

Optimal forecast combination method relies on observing the true outcome Y over time, alongside forecasts X_j and corresponding forecast errors. The first two moments of the joint distribution of $(Y, X_1, ..., X_J)$ are key. Note that the conditional covariance matrix of individual forecasts (and equivalently the covariance matrix of individual forecast errors) is based on time-series pairwise covariance between two forecasters.

$$\begin{bmatrix} Y \\ X \end{bmatrix} \sim \left(\begin{bmatrix} \mu_Y \\ \mu_X \end{bmatrix}, \begin{bmatrix} Var(Y) & Cov(Y,X) \\ Cov(Y,X) & \Sigma_X \end{bmatrix} \right) \tag{3.2.2.1}$$

The first two moments in 3.2.2.1 determine optimal forecast combination weights W^* on forecasts X (Timmermann, 2006) as:

$$W^* = \frac{\Sigma_{e_j}^{-1} 1_J}{1_J' \Sigma_{e_j}^{-1} 1_J}$$

where Σ_{e_j} is the variance-covariance matrix of J forecast error vectors $(e_1, ..., e_J)$, each a time series for some $t = 1, ..., T$, and 1_J as a vector of ones of length J. The optimal combination of all X_j forecasts is therefore $W^{*'}X$. The weights in W^* have an intuitive interpretation—weights are larger for historically more accurate forecasts

that are less strongly correlated with other forecasts. This leads to improvements from diversification, in the context of forecasting precision, just like gains from diversification with Markowitz portfolios. Note that a forecast is more accurate when its $Cov(Y, X_j)$ is greater, because that means the time-series forecast errors are smaller. Furthermore, these weights are time-invariant, unless they are estimated to be time-varying by design. The underlying assumption is that differences in information among forecasters are constant over time.

Timmermann (2006, p.149-150) shows that equally-weighted forecasts are in fact a special case of these optimal weights. That is the case when forecast errors have the same variance, σ^2, and same correlation, ρ, with an arbitrary number of forecasters. In other words, if noisy signals of true returns are equally correlated among all forecasters, and all forecaster's errors have the same variance, the optimal forecast combination reduces to a simple average of the individual forecasts.

Several alternative empirical methodologies are also possible: trimming, or clustering followed by pooling[2]. There is also a variety of other methodologies that rely on shrinkage in the style of machine learning. These are outside the scope of this paper since they have no conceptual similarities with the Satopää et al. (2016) methodology. The Satopää et al. (2016) methodology is a statistical model of information diversity and does not have any machine learning features in a sense that it does not use past data as a training set.

[2]Trimming is appropriate if a fraction of forecasters have informative signals, while the rest observe signals that are pure noise. A trimming scheme removes poorly performing forecasts, based on historical forecast errors. As a second step, the remaining forecasts can be used to estimate optimal weights. Another possibility is a clustering scheme. It is motivated by the assumption of a common factor structure, and can be implemented based on correlation patterns among forecast errors. The forecasts within clusters are then pooled, and optimal weights are estimated as a second step.

3.2.3 Novel approach to combine a cross-section of forecasts

The approach developed by Satopää, Jensen, Pemantle, and Ungar (2016) shares many conceptual similarities with forecast combination methodologies described by Timmermann (2006). In fact, the revealed aggregator method has the same intuition about diversification gains when correlated forecasts are combined optimally, just like assets in a Markowitz portfolio. Nonetheless, it has distinct differences from the measurement error and forecast combination. On a conceptual level, the framework is based on an explicit statistical model of partial and overlapping information, contemporaneously. Signals observed by different forecasters only differ in their informativeness. There is no noise. The practical implication of a contemporaneous setting is that there is no need for true Y and for past forecast errors, because the framework relies on cross-sectional covariance. Agents need to be making a cross-section of forecasts instead. Therefore, this methodology can be easily applied to data where a long historical time series of forecasts is not available, but multiple contemporaneous forecasts are available instead. This is not a machine learning technique in the sense that there is no training on historical data. It also does not require each forecaster to separately predict the distribution of other forecasters' beliefs, which is one other novel way to extract the wisdom of crowds by finding minority experts (Prelec et al., 2017). The revealed aggregator infers differences in information across forecasters from their contemporaneous event predictions alone.

Partial information framework

The setting of the partial information framework is the following. Treat the forecast X_j and the true outcome Y as generic random variables, which are then measurable in probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is a set of all possible outcomes of Y, sigma-field $\sigma(\mathcal{F})$ is a collection of all possible information about Y from information partitions $\mathcal{F}$, and $\mathbb{P}$ is a probability measure like the common prior that reflects information

known to all agents.

Crucially, each agent j has a personal partial information set $\mathcal{F}_j \subseteq \mathcal{F}$. Information sets of agents can overlap. In other words, there is information diversity. High overlap is equivalent to low information diversity, and vice versa. Additionally, agents' information is only revealed through their forecasts, and information that agents disregard when making recommendation X_j is assumed not to exist, hence $\mathcal{F}_j = \sigma(X_j)$.

Assume that each agent is assumed and makes a reliable forecast X_j of the outcome Y. Being reliable (or calibrated) is a statistical property, defined as predicting $X_j = \mathbb{E}(Y \mid X_j)$ such that $X_j = \mathbb{E}(Y \mid \mathcal{F}_j)$. If agents can be considered experts, this is a reasonable assumption because forecasts of experts like meteorologists, bridge players, and bookmakers are shown to be reliable, particularly in case of repetitive cognitive processes (Murphy and Winkler (2010) and Keren (1987)), as also argued by Satopää et al (2016, p.8-9). However, if agents are strategic and their forecasts are biased, this assumption is likely violated.

In this setting, under this assumption, Satopää et al (2016, p.10) prove a useful proposition below. It states that (i) prior mean of true outcome Y can be estimated as the average of individual forecasts X_j, and that (ii) variance of X_j (which is conceptually a random variable) quantifies the information set of agent j. Part (iii) of the proposition quantifies overlap between information sets of any two agents. Proof of the Proposition is in the Appendix.

Proposition. *If $\mathcal{F}_j = \sigma(X_j)$ such that $\mathbb{E}(Y \mid \mathcal{F}_j) = \mathbb{E}(Y \mid X_j) = X_j \forall j = 1, ..., N$:*

i Forecasts are marginally consistent: $\mathbb{E}(Y) = \mathbb{E}(X_j)$.

ii More information implies more variance: $Var(X_i) \leq Var(X_j)$ if $\mathcal{F}_i \subseteq \mathcal{F}_j$. Given that $Y = \mathbb{E}(Y \mid \mathcal{F})$, the variances of the forecasts are upper bounded as $Var(X_j) \leq Var(Y) \forall j$.

iii Information overlap between any two agents: $Cov(X_j, X_i) = Var(X_i)$ if $\mathcal{F}_i \subseteq \mathcal{F}_j$. With $Y = \mathbb{E}(Y \mid \mathcal{F})$, it implies that $Cov(X_j, Y) = Var(X_j) \forall j$.

Part (ii) of the Proposition says that agent j's recommendation has higher variance than agent i's recommendation if agent j has a larger information set. This may seem counterintuitive, but it makes sense in this framework. Recall that signals observed by different agents only differ in their informativeness. An agent receiving a signal that encompasses a greater variety of information will be able to make more bullish or bearish forecasts. That agent will have a higher variance of the underlying random variable X_j, which represents agent's recommendation—a forecast and therefore an estimator of future stock returns Y.

Part (iii) is more intuitive—that information overlap between any two agents is determined by the agent with a smaller information set. It follows that the relationship between true outcome and what is forecasted by any agent j is similarly determined by the smaller information set. Agent j has a partial and therefore a smaller information set, so $Cov(X_j, Y) = Var(X_j) \forall j$. Covariance between true outcome and outcome forecasted by agent j is therefore the variance of agent j's forecast. This characterization of information overlaps extends to the relationship between Σ_X, the $J \times J$ covariance matrix of all forecasts X_js, and the $(J+1) \times (J+1)$ covariance matrix of a vector $(Y, X_1, ..., X_J)$ which includes unknown true outcome Y. It then follows that the joint distribution of true outcome and its forecasts simplifies to (Satopää et al 2016, eq.1):

$$Cov((Y, X_1, ..., X_J)') = \begin{bmatrix} Var(Y) & diag(\Sigma_X)' \\ diag(\Sigma_X) & \Sigma_X \end{bmatrix} \tag{3.2.3.1}$$

With Proposition (iii) and 3.2.3.1, Satopää et al. (2016) are in fact generalizing the forecast combination setting in 3.2.2.1. They are able to do so because they rely on partial information sets in the model. This leads to an application with cross-sectional, rather than time-series, covariance across agents. This also eliminates the need to know true return Y, or its historical time series.

This mapping of information sets of agents based on the correlation of their cross-sectional forecasts leads to the revealed aggregator. The revealed aggregator uses the union of everyone's information, i.e., using all information that is revealed through the forecasts.

$$X'' := \mathbb{E}(Y \mid X_1, ..., X_N) = \mathbb{E}(Y \mid \mathcal{F}''), \text{ where } \mathcal{F}'' = \sigma(X_1, ..., X_J) \tag{3.2.3.2}$$

I follow (Satopää et al 2016, p.12-13) with the choice of a multivariate Gaussian distribution for $(Y, X_1, ..., X_J)$.

Recall that forecast X_j is a random variable, while Σ_X is an $J \times J$ covariance matrix of all recommendations $(X_1, ..., X_J)$. However, in the real world we observe only one realization per forecast X_j. It is convenient that agents make recommendations for many outcomes, $K \geq 2$. These multiple forecasts per agent can be treated as realizations of random variables that are all based on an agent-specific information set. As such, having forecasts for a cross-section of outcomes makes it feasible to estimate the information structure of agents, Σ, and to obtain the revealed aggregator, X''.

Model-based estimation of a revealed aggregate

Take an $J \times K$ agent-event matrix of estimates X_{jk}, and apply the following:

> **Model.** Suppose that Y_k and X_{jk} are real-valued. Assuming a multivariate Gaussian distribution, the non-informative prior distribution of Y_k is $\mathcal{N}(\mu_{0k}, \sigma_0^2)$. With a link function $g()$, which is a location-scale normalization of the real-valued model variables Y_k and X_{jk}, the random variables in terms of Z_k are: $Y_k = g(Z_{0k}) = Z_{0k}\sigma_0 + \mu_{0k}$ and $X_{jk} = \mathbb{E}(Y \mid Z_{jk}) = Z_{jk}\sigma_0 + \mu_{jk} \forall j = 1, ..., J$. Hence $X_j \sim \mathcal{N}(\mu_{0k}, \sigma_j^2)$ for some $\sigma_j^2 \leq \sigma_0^2$. If $Z_k = (Z_{1k}, ..., Z_{Jk})'$, then the revealed aggregate estimate for the kth event is: (Satopää et al 2016, p.34, eq.9):
>
> $$X_k'' = \mathbb{E}(Y_k \mid Z_k) = diag(\Sigma)'\Sigma^{-1}Z_k\sigma_0 + \mu_{0k} \tag{3.2.3.3}$$

Since the prior mean of Y can be estimated as the average of the forecasts, use $\hat{\mu}_{0k} = \frac{1}{J}\sum_{j=1}^{J} X_{jk}$ as an initial estimate of μ_{0k}. Similarly, estimate the prior variance of Y, denoted σ_0^2, as $\hat{\sigma}_0^2 = \frac{J+1}{J} max(s_j)$, where $s_j = \frac{1}{K-1}\sum_{j=1}^{K}(X_{jk} - \hat{\mu}_{0k})^2$ is the sample variance[3]. Using these estimates $\hat{\mu}_{0k}$ and $\hat{\sigma}_0^2$, transform X_{jk}s into Z_{jk}s as $\frac{1}{\hat{\sigma}_0}(X_{jk} - \hat{\mu}_{0k}) = Z_{jk}$. The sample $J \times J$ covariance matrix of Z_{jk}s, denoted as S_Z, provides the unconstrained estimator of Σ. Missing values in Z_{jk}, if any, are handled by using pairwise covariance based on only the events for which both agents provide an estimate. Follow Satopää et al (2016, p.18-23) estimation procedure to calculate a corrected covariance matrix Σ_{cov} from the unconstrained sample covariance matrix S_Z. Update the prior means with the precision-weighted averages:

$$\mu_{0k} = \frac{X_k' \Sigma_{cov}^{-1} 1_J}{1_J' \Sigma_{cov}^{-1} 1_J} \tag{3.2.3.4}$$

where 1_J is a vector of ones of length J. Finally, plug these precision-weighted averages, transformed recommendations Z_{jk}, estimated covariance matrix Σ_{cov}, and the estimate of variance $\hat{\sigma}_0^2$, in to 3.2.3.3. This is the revealed aggregator X_k'' for each event $k = 1, ..., K$.

Weights implied by the revealed aggregator

The revealed aggregator X'' can be represented with implied weights for each individual estimate X_{jk}. Since the revealed aggregator estimate $X'' = \mathbb{E}(Y \mid \mathcal{F}'')$ for particular outcome k can be expressed as:

[3]Following Satopaa, I let σ_{0k}^2 be i.i.d. on $[0, \sigma_0^2]$, and therefore use the maximum likelihood estimator $max(\sigma_j^2)$, correcting for a downward bias.

$$X_k'' = \sigma_0 diag(\Sigma_{cov})' \Sigma_{cov}^{-1} Z_k + \frac{X_k' \Sigma_{cov}^{-1} 1_J}{1_J' \Sigma_{cov}^{-1} 1_J}$$

where:

$$Z_k = [Z_{1k}, ..., Z_{Jk}]' \text{ and } Z_{jk} = \frac{1}{\hat{\sigma}_0}(X_{jk} - \hat{\mu}_{0k}) = \frac{1}{\hat{\sigma}_0}(X_{jk} - \frac{1}{J}\sum_{j=1}^{J} X_{jk}) \tag{3.2.3.5}$$

the weights representation is $X_k'' = W_k' X_k$, where:

$$\begin{aligned} W_k' &= [w_k + \phi \tilde{w}_k - S_k]' \\ w_k' &= diag(\Sigma_{cov})' \Sigma_{cov}^{-1} = [w_{1k}, ..., w_{Jk}], \ \phi = \frac{1}{1_J' \Sigma_{cov}^{-1} 1_J} \\ \tilde{w}_k &= \Sigma_{cov}^{-1} 1_J = [\tilde{w}_{1k}, ..., \tilde{w}_{Jk}]', \ S_k = [s_k, ..., s_k]' \text{ and } s_k = \frac{1}{J}\sum_{j=1}^{J} w_{jk} \end{aligned} \tag{3.2.3.6}$$

For balanced data where every agent provides an estimate for every event, Σ_{cov} is the same for each event, and therefore the implied weights W_k are also the same. Conceptually, these implied weights correspond to the contemporaneous information set the agent has. Recall that an agent j has a personal partial information set $\mathcal{F}_j \subseteq \mathcal{F}$ and makes a reliable forecast $X_j = \mathbb{E}(Y \mid X_j)$ such that $X_j = \mathbb{E}(Y \mid \mathcal{F}_j)$. Information sets $\mathcal{F}_j$ in this framework are agent-specific, rather than agent-stock-specific. However, for unbalanced data where there is missing data in the $J \times K$ agent-event matrix of recommendations, Σ_{cov} is different for each event because it is calculated based on the pairwise covariances for overlapping sets of events that any two agents both estimate. As a result, the implied weights W_k are not the same for every event. Lastly, note that if analysts make recommendations over and over, and this revealed aggregate is applied at each point in time, the obtained implied weights are time-varying.

Nesting the simple average

Furthermore, the revealed aggregator X_k'' nests the simple average in one special case. When there are many information sets with identical overlap, i.e., symmetric information diversity, and the information sets are all the same size, then a simple average turns out to be the optimal aggregator. In this special case, in Σ_{cov}, all agents have the same variance across the events, as σ^2, and all pairwise covariances of agents, across events they provide estimates for, are also the same, as ρ. A proof of this special case with a simplified 2×2 setting is in the Appendix. If either the agents' variances are different, or any of the pairwise covariances are different, then the revealed aggregator will be different from the consensus aggregator. Note that this is the same special case as when optimal weights in the forecast combination literature nest the simple average, but in a cross-sectional rather than time-series context that is used to calculate the variance-covariance matrix. This similarity is reassuring.

This concludes the theory behind partial information framework and how a model-based revealed aggregate can be estimated in practice. If the assumptions about information diversity hold, the intuition behind the revealed aggregator is that it maps the information sets of individuals based on the cross-sectional covariance of their forecasts across several cases, and distinguishes the needle in the haystack from popular information that people are herding on. As a second step, the revealed aggregate estimate takes the union of everyone's information.

3.3 Revealed aggregate recommendation score of security analysts

Now consider a setting with J analysts, who recommend stocks to buy, hold, or sell. A recommendation of 1 corresponds to a "strong buy", 3 corresponds to "hold", and

5 corresponds to "strong sell"[4]. This recommendation is linked to analyst's forecast of future stock price changes. Therefore it is a direct application of the statistical setting described above. An analyst j makes a recommendation X_j, which is a forecast and therefore an estimator of a future stock return Y.[5] Recommendations of sell-side security analysts provide an interesting test with respect to the prevalence of information diversity using the novel partial information framework.

Besides recommendations, analysts also issue corporate earnings forecasts every quarter. However, that is not my application of choice. In fact, forecast combination method could be well suited to earnings forecasts because past forecast errors are straightforward to calculate based on realized corporate earnings, and these forecasts are also issued frequently, so that there is typically a relatively long historical time series of earnings forecasts per analyst. In contrast, the contemporaneous partial information framework is well suited for recommendations data, for several reasons. Recommendations are not revised very frequently and we don't typically have a long historical time series of recommendations per analyst to be able to apply time-series techniques. Since analysts' recommendations X_j are $[1-5]$, with subsequent return $Y \in \mathbb{R}$, it is not clear how to calculate past forecast errors. Analyst recommendations represent tradable advice—to buy, sell, or hold—so portfolio analysis is directly applicable. Additionally, without the use of historical data, everything is estimated out-of-sample.

There are a number of further practical issues I have to consider when taking this partial information theory to the data, and estimating the revealed recommendation score X'' on the basis of individual recommendations X_{jk} of J analysts forecasting K stocks. This section describes how I estimate the revealed recommendation in practice.

[4]This kind of recommendation is rare.

[5]In the analysis that follows, I am assuming X_j and Y are continuous variables, while in fact, stock recommendations [1,2,3,4,5] are bounded and more accurately treated as ordinal variables. This will be incorporated in the forthcoming version of the paper.

3.3.1 Desired data properties and descriptive statistics

An analyst does not issue a recommendation for every stock in the world. So the data is unbalanced, and any pair of analysts taken at random may not have any stocks they both cover. It is methodologically important to have relatively few missing observations. Intuitively, if two analysts forecast different sets of stocks, it will not be possible to map their information sets using the cross-sectional correlation of their recommendations. The problem is that analysts typically issue recommendations for 10-20 stocks and overlap in stock coverage is very limited. In fact, the overlap is not big enough to be able to estimate the information structure of analysts, Σ, for a sufficiently large number of stocks that can later be split into portfolios. This is also the case even when grouping analysts by the same industry.

Therefore, when taking this methodology to the data, I consider broker-level analysis rather than analyst-level analysis. Brokerage houses typically issue recommendations for 50-600 stocks and coverage is less of a problem. There are several justifications for using brokerage-specific information sets rather than analyst-specific information sets. One story is that same forecasting models and methodologies are used in a brokerage house. Another story is that analysts derive institutional-level information or tips from their managers who communicate with clients. Regardless, using brokerage-specific information sets underestimates—rather than overestimates—the significance of diverse information at an individual level.

Therefore this is a setting with J brokers, who recommend stocks to buy, hold, or sell, rated $[1-5]$ as before. The unknown return on stock k is Y_k and jth broker's forecast of this return is the recommendation X_{jk}. Brokerage-level information sets produce data with more desirable properties for the revealed aggregate estimator. Yet the share of missing data in the $J \times K$ broker-stock matrix of recommendations still remains pretty high. To address this, I apply a recursive data trimming algorithm that targets 50% share of missing observations and only includes stocks that

have a pairwise overlap of at least five brokers. The details are described in the Appendix. The 50% share of missing observation is the same fraction as in the Satopää et al (2016) application with unbalanced data. Using brokerage-level recommendations only for the popular stocks listed on the NYSE and NASDAQ, and only for the most prolific brokerage houses, means that I use recommendations about companies with greater market capitalization and from large brokerage houses that don't specialize in any one industry. Lack of specialist or hard-to-come-by information also underestimates—rather than overestimates—the significance of information diversity in my results.

Analyst-level data on stock recommendations comes from the Institutional Brokers' Estimate System (I/B/E/S) database. I use data only for the popular stocks (which have at least 10 brokerage houses covering them) listed on the NYSE and NASDAQ. This results in 700-1300 stocks that are included in the sample each week. The names of analysts behind the recommendation are sometimes listed as "Permission Denied" in I/B/E/S. However, each brokerage house has a unique anonymous identifier, for example, "Permission Denied 32". Thus I have an advantage in using brokerage-level data because I don't have to discard unidentified analysts, which represent about a third of the sample. The returns data comes from Thompson Reuters and CRSP, hence inclusive of dividends and split adjustments.

Table 3.1: Data dimensions after the trimming algorithm is applied

	Average		
Year	Number of brokers	Number of stocks	Missing share
2007	28	351	0.482
2008	25	372	0.483
2009	22	385	0.484
2010	27	416	0.486
2011	29	463	0.488
2012	31	494	0.488
2013	32	529	0.486
2014	32	562	0.49
2015	33	579	0.488
2016	33	595	0.486

Average number of brokers, stocks, and the corresponding missing share of observations in the data used for main results. Averages are taken per year.

As seen in Table 3.1, the final data used in the analysis spans 28-33 brokers that cover 351-595 stocks on average. The results presented in the paper are not sensitive to this trimming algorithm, and the same conclusions hold when I only use the most prolific brokerage houses, such that each brokerage house covers at least 100 stocks. This alternative corresponds to top 80% of brokers and a 32-61% of missing observations in the broker-stock matrix of recommendations. The $J \times K$ matrix dimensions alternatively range from 23×435 to 41×807. The advantage of the recursive algorithm is that it can better handle data from earlier years, such as the 1990s and 2000s, when brokers were less prolific, and fewer stocks were listed.

The main results are based on ten years of weekly data for 2007-2016, i.e., Jan 6, 2007–Dec 23, 2016, sampled on Fridays unless a holiday. I also consider a longer time horizon as a robustness check—as far back as the data is accessible, starting from Jan 7, 1994. Qualitative conclusions of portfolio analysis do not change with the longer time horizon, but they are weaker. The longer time period includes 2001-2002, the years when sell-side analysts were notably wrong in that the poorly-rated stocks substantially outperformed well-rated dot.com bubble stocks. Results of the portfolio

analysis pick up evidence of this phenomenon.

3.3.2 Deducing the information sets: an example

The formal discussion so far has presented the revealed aggregator as a black box. Here I illustrate how it works, in a more intuitive and transparent manner. I use the data from Friday, April 15, 2016 as an example. There are $J = 33$ brokers with recommendations for $K = 575$ stocks, and the share of missing data in the $J \times K$ matrix of recommendations is 48%. Consider two stocks, Ally Financial Inc (NYSE:ALLY) and Teradata Corporation (NYSE:TDC). Their consensus and revealed recommendations as of April 15th 2016 are reported in Table 3.2 and are very different. Consensus recommendation rates ALLY very favorably, and TDC least favorably, while the revealed information method ranks them the other way around.

Table 3.2: Consensus and revealed aggregates for stocks ALLY and TDC as of April 15, 2016

		Consensus			Revealed		
Stock	No. brokers	Score	Percentile	Favourable	Score	Percentile	Favourable
ALLY	12	1.75	0.06	yes	2.31	0.58	no
TDC	19	3.05	0.93	no	1.83	0.39	yes

An example that illustrates how the revealed aggregate score can substantially differ from the consensus score.

Why does the revealed aggregate differ so much from the simple average, if they are based solely on the same underlying data of broker's recommendations as of April 15th 2016? By considering variances and covariances of brokers across all their recommendations for this date, the brokers can be grouped into teams of like-minded thinkers, which correspond to their underlying information. This yields roughly two large, one medium, and one small distinct information sets. The revealed aggregate is then based on the union of these distinct information sets. This is further explained

below.

Of the 19 brokers covering TDC, the majority vote is unfavorable: 17 brokers recommend a 3 (hold) or a 4 (sell); only anonymous brokers 32 and 22760 recommend a 1 (strong buy). It's the opposite for ALLY, where 11 brokers recommend a 1 (strong buy) or a 2 (buy), while only one anonymous broker 32 recommends a 3. Looking exclusively at brokers that covered both stocks, only broker 32 recommends 1 (strong buy) for TDC and 3 (hold) for ALLY, whereas four brokers (let's call them Team ALLY) have a recommendation of 2 (buy) for ALLY and 4(sell) for TDC[6]. So brokers 32 and 22760 seem to be both taking a stand against the crowd. We don't know whether this taking-a-stand outcome is due to additional information or this is actually a mistake. If it's due to additional information, their bet on TDC as opposed to ALLY will win, yet if it's due to a mistake, their bet will lose, and the majority vote will win.

Let's focus further on TDC recommendations. The revealed aggregator gives brokers 32 and 22760 very high weights, 0.63 and 0.61 respectively, and downweighs recommendations made by four brokers in Team ALLY and all other brokers that only covered TDC and recommended to sell or hold it, as they have an average weight of 0.15.

This is because under partial information framework, a higher variance is equivalent to having a larger information set, as per proposition (ii). Since both brokers 32 and 22760 have high variances (0.81 and 0.91 respectively, as seen in Table 3.3), their recommendations are treated as based on a lot of information. Furthermore, there is not a lot of information overlap between broker 32 and broker 22760 since the covariance of their recommendations is negligible and negative, at -0.04. This means that brokers 32 and 22760 are treated as having large and distinct information sets by the revealed aggregator.

[6]These brokers are Credit Suisse North America, and 3 anonymous brokers with codes 23816, 398136,73704

Table 3.3: Splitting brokers into teams

			Big Two	Team ALLY
Average pairwise ingroup covariance:			0.040	0.001
Variance-covariance matrix:				
	Anonymous 32	Anonymous 22760	Big Two	Team ALLY
Anonymous 32	0.813	-0.036	-0.069	-0.065
Anonymous 22760	-0.036	0.907	-0.019	-0.088
Big Two	-0.069	-0.019	0.679	-0.042
Team ALLY	-0.065	-0.088	-0.042	0.249
Recommendation:	1(strong buy)	1(strong buy)	3(hold)	4(sell)

An approximate illustration of how variances and pairwise covariances can be used to group individual recommendations into like-minded, or orthogonal, thinkers.

Referring to Table 3.3 again, among all the brokers outside of Team ALLY that recommended to sell or hold TDC, only two have high variance (let's call them Big Two)[7], with an average variance of 0.68. Since their pairwise covariance is positive, at 0.04, this is treated as some information overlap. Therefore these two brokers together contribute some additional information to the separate information sets of broker 32 and broker 22760. The four brokers in Team ALLY all have low variance (with an average of 0.25) and therefore smaller information sets. At 0.001, their average pairwise covariance is small, but for convenience of interpretation, let's keep Team ALLY brokers grouped together. Since the covariance of Team ALLY with brokers 32, 22760, and Big Two respectively is negative, Team ALLY does not have an information overlap with them. Therefore, the revealed aggregator treats Team ALLY and Big Two as bringing some additional information to what are separate information sets of broker 32, broker 22760, though with less weight.

As such, the data is reduced to roughly two large distinct information sets of broker 32, of broker 22760, medium information set of Big Two, and one smaller distinct information set of Team ALLY. This is obviously a highly stylized simplification, but it is nonetheless useful to understand how the revealed aggregator works. In conclusion,

[7]These two brokers are Stifel Nicolaus & Co Inc and William Blair & Co

Brokers 32 and 22760 both recommend a strong buy for TDC, while the other two both recommend holding the stock, and Team ALLY brokers recommend to sell it. Based on the union of four information sets and their respective recommendations, the revealed aggregate for TDC is 1.83. It corresponds to the top 40th percentile of all revealed recommendations, and is much more favorable than the consensus score.

A bet on April 15th in line with brokers 32 and 22760 taking a stand on TDC against the consensus would have paid off in the medium term, which is illustrated in Figures 3.3-3.4. Following the release of quarterly earnings and news announcement of CEO replacement, TDC stock outperformed the market and ALLY until mid-July. Coincidentally, the revealed and consensus aggregator roughly converged by July 15, 2016. It is possible that brokers 32 and 22760 did have some information in advance of earnings release and news announcement. However, the alternative that this is by chance cannot be ruled out, and this is just an illustrative example.

Figure 3.3: Recommendations for ALLY and TDC stocks, Apr 16 - Sept 15, 2016

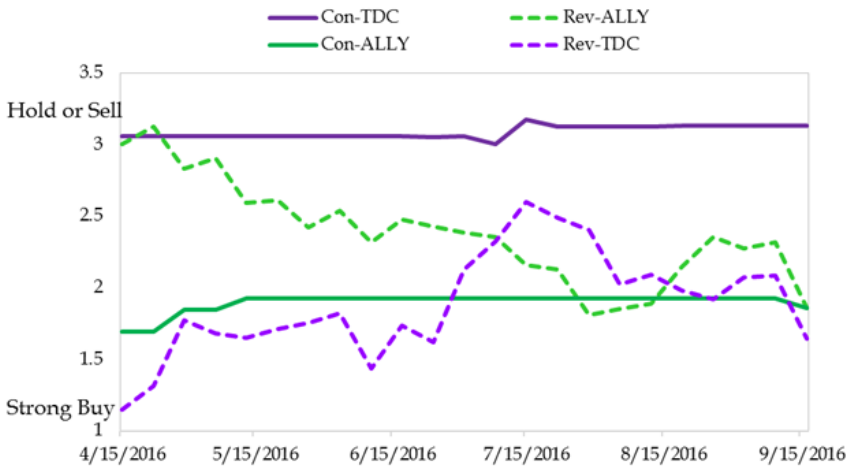
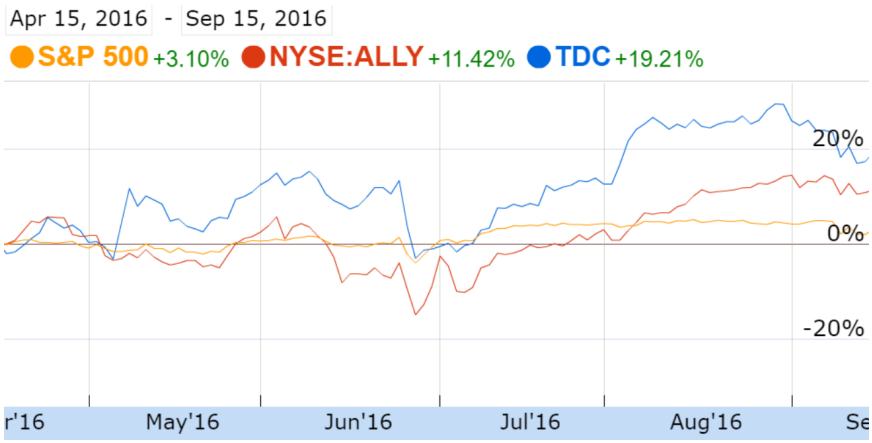


Figure 3.4: Performance of ALLY and TDC stocks, Apr 16 - Sept 15, 2016



More illustrative examples, such as the impact of an upgrade and some further observations, are in the Appendix.

3.4 Portfolios formed on consensus and revealed aggregate recommendation scores

If differences in analyst recommendations stem from information diversity, then trading on consensus score is suboptimal. Trading on the revealed aggregator should outperform trading on consensus, and might deliver abnormal returns. Therefore, I can first measure information diversity. The null hypothesis is that everyone is equally well-informed and differences in recommendations mostly reflect noise. Then averaging is optimal and cannot be beaten. The alternative is that differences in recommendations reflect information diversity, revealed aggregate score differs from the simple average, and trading on the revealed aggregate score is more profitable. This test provides a simple measure of information diversity.

If the first null hypothesis is rejected in favor of information diversity, this has further testable implications concerning market efficiency. Semi-strong market efficiency hypothesis states that investors should not be able to trade profitably on

publicly available information such as recommendations of sell-side security analysts, but could profit from private information. Therefore the second null hypothesis is that diversity of information concerns public information, which is efficiently aggregated by the market, and trading on the revealed aggregate does not deliver abnormal returns, just like trading on the consensus. One alternative is that differences in recommendations are due to some private information, which leads to non-zero abnormal returns. In the context of weekly rebalancing of portfolios, this concerns private information among analysts that is persistent and is not fully reflected in the price drift subsequently following an upgrade or a downgrade by an informed analyst, and this leads to non-zero abnormal returns.

3.4.1 Portfolio construction

I follow Barber and Odean (2001) for this portfolio analysis, who use the level of consensus recommendation. Therefore, the only methodological innovation in portfolio analysis here is aggregating individual recommendations with an alternative statistical technique that can leverage differences in information among agents and combine them more efficiently.

These main results are based on ten years of weekly data for 2007-2016, sampled on Fridays. If Friday falls on a holiday and is a non-trading day, I use the data from Thursday, the day before. Each week, for each stock, I calculate the revealed recommendation and the consensus (simple average) recommendation. Formally, the simple average recommendation $Avg_{k,\tau-1}$ for stock k on date $\tau - 1$ is below, for the $j = 1, ..., N_{k,\tau-1}$ brokers who have recommendations for the stock on that Friday is $Avg_{k,\tau-1} = \frac{1}{N_{k,\tau-1}} \sum_{j=1}^{N_{k,\tau-1}} X_{jk,\tau-1}$. The revealed recommendation, as $Rev_{k,\tau-1}$, is based on 3.2.3.3 and using the same brokers.

I repeat this for a total of 521 weeks and then construct two calendar-time portfolios for the same stocks. One portfolio is based on the revealed recommendation

scores, and another portfolio is based on consensus scores. These portfolios are formed in the standard way, which is described below.

Each week, I sort stocks using the score metric, $Score_{k,\tau-1} = (Avg_{k,\tau-1}, Rev_{k,\tau-1})$. I therefore place each stock into one of five quintiles as of the close of trading on the basis of $Score_{k,\tau-1}$. The first portfolio consists of the most highly recommended stocks, defined by the top quintile, i.e., 0-20th percentile. Similarly, the second, third, and fourth portfolios have cutoffs equal to 40th, 60th, and 80th percentiles. An alternative for portfolio construction is to consider hard cut-offs, such that $1 \leq Score_{k,\tau-1} \leq 1.5$ for the first portfolio, $1.5 \leq Score_{k,\tau-1} \leq 2$ for the second portfolio, $2 \leq Score_{k,\tau-1} \leq 2.5$ for the third portfolio, $2.5 \leq Score_{k,\tau-1} \leq 3$ for the fourth portfolio, and $Score_{k,\tau-1} > 3$ for the fifth portfolio, as the least favorably rated stocks. The results are not qualitatively different and are presented in the Appendix.

After determining the composition of each portfolio p as the stocks that are in the portfolio, $k_p = 1, ..., K_p$, at the close of trading on date $\tau - 1$, I calculate the value-weighted return for subsequent week, denoted by $R_{p,\tau}$ for portfolio p, using market capitalization as weights:

$$R_{p,\tau} = \sum_{k_p=1}^{K_{p,\tau-1}} x_{k,\tau-1} R_{k,\tau}$$

where

$x_{k,\tau-1}$ = market value of equity for stock k divided by the total market capitalization of all stocks in the portfolio p as of the close of trading on date $\tau - 1$

$R_{k,\tau}$ = the return on the common stock k on date τ

$K_{p,\tau-1}$ = number of stocks in portfolio p at the close of trading on date $\tau - 1$

The resulting average annualized return $R_{p,\tau}^{avg,ann}$ of the five value-weighted portfolios are calculated as an arithmetic average[8]. I also report geometric average returns,

[8] $\bar{R}_{p,\tau}^{avg,weekly} = \frac{1}{521} \sum_{\tau=1}^{521} R_{p,\tau}$ and $R_{p,\tau}^{avg,ann} = (1 + \bar{R}_{p,\tau}^{avg,weekly})^{52} - 1$

i.e., annualized returns $R_{p,\tau}^{cum,ann}$ for the five value-weighted portfolios that are compounded over all the weeks in the sample[9]. These cumulative returns are interpreted as returns from continuously following this strategy.

3.4.2 Test for information diversity

Trading on the revealed aggregate is more profitable than trading on the consensus recommendation. Consider a continuous trading strategy and annualized returns on the basis of compounding. For the most favorably rated stocks, buying stocks on the basis of the revealed aggregate would have earned an 11.57% annualized return, compared to 8.32% annualized return if buying stock on the basis of the consensus recommendation, with a difference of 3.25%. This is a 3.63% return in excess of the market (Table 3.4). The performance of this first portfolio over time is reported in Figure 3.5, in contrast to a similar portfolio using stock consensus scores and in contrast to the market.

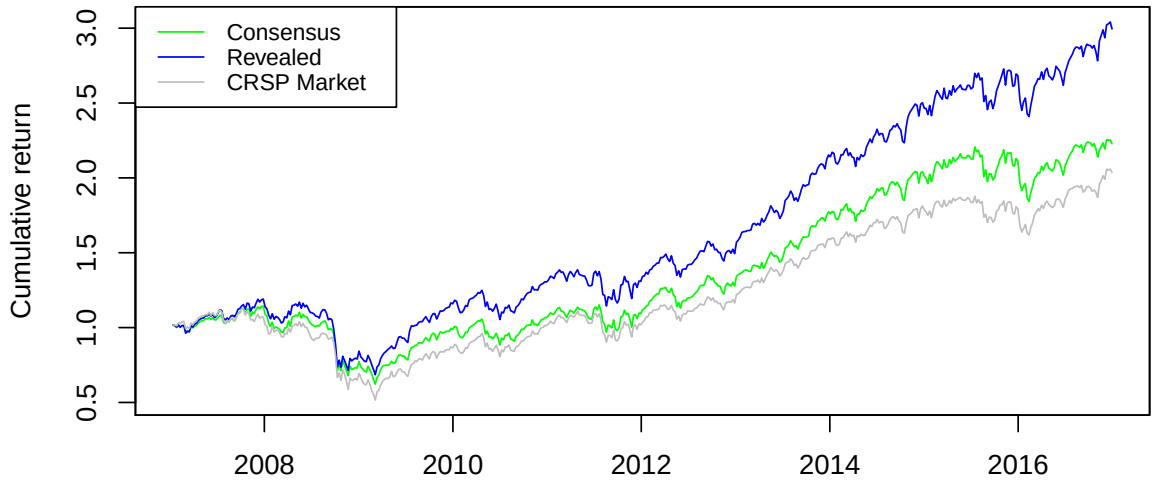
Table 3.4: Geometric average returns earned by portfolios formed on the basis of consensus and revealed recommendations.

	Portfolios				
Formed on the basis of:	1st (best)	2nd	3rd	4th	5th (worst)
	Annualized cumulative gross returns:				
Consensus	8.32	6.44	8.06	8.85	6.76
Revealed	11.57	5.50	5.81	8.29	7.40
CRSP Market			7.35		
	Annualized cumulative market-adjusted returns:				
Consensus	0.56	-1.09	0.59	1.61	0.02
Revealed	3.63	-2.01	-1.69	0.98	0.61

Compound annualized returns of value-weighted portfolios, with weekly rebalancing, from Jan 6, 2007 to Dec 23, 2016. The 1st portfolio is one with the most favorably recommended stocks (to buy), while 5th portfolio is one with the least favorably recommended stocks (to sell). Returns are expressed in percent. There are no associated significance tests.

[9] $R_{p,\tau}^{cum,total} = \prod_{\tau=1}^{521}(1+R_{p,\tau}) - 1$ and $R_{p,\tau}^{cum,ann} = (1+R_{p,\tau}^{cum,total})^{\frac{1}{10}} - 1$

Figure 3.5: Geometric returns earned by the first (best) portfolio.



Compound returns of value-weighted portfolios, with weekly rebalancing, for $1 invested in Jan 6, 2007 to Dec 23, 2016. These are portfolio returns for the 1st portfolio, with the most favorably recommended stocks (to buy).

In terms of average returns, which are reported in Table 3.5, trading on the best revealed recommendations beats trading on the best consensus recommendations by 3.36% in annualized terms. I also apply a paired t-test to compare these two time series of non-overlapping weekly returns. The paired t-test indicates that this difference is statistically significant at a 1% level, and only in the case of portfolios formed on most favorably rated stocks. Next I compare these individually with the CRSP Market weekly return. Only the portfolio that comprised of most favorably rated stocks formed on the basis of the revealed aggregator beats CRSP Market in the context of a paired t-test, with an annualized average excess return of 3.78%.

Table 3.5: Average returns and Sharpe Ratios earned by portfolios formed on the basis of consensus and revealed recommendations.

	Portfolios				
Formed on the basis of:	1st (best)	2nd	3rd	4th	5th (worst)
	Annualized average gross returns:				
Consensus	10.09***	8.24	10.06	11.34	9.85
Revealed	13.45***	7.26	7.58	10.57	10.51
CRSP Market			9.33		
	Annualized average market-adjusted returns:				
Consensus	0.69	-0.99	0.67	1.85	0.48
Revealed	3.78**	-1.9	-1.6	1.13	1.08
	Annualized Sharpe Ratio:				
Consensus	0.50	0.40	0.46	0.47	0.37
Revealed	0.65	0.35	0.36	0.46	0.39
CRSP Market			0.43		

Average annualized returns of value-weighted portfolios, with weekly rebalancing. The 1st portfolio is one with the most favorably recommended stocks (to buy), while 5th portfolio is one with the least favorably recommended stocks (to sell). Sample data is weekly, Jan 6, 2007 - Dec 23, 2016, taking into account non-trading days. Returns are expressed in percent. The stars correspond to 0.1, 0.05, and 0.01 significance levels for a paired t-test. In case of gross returns, the null is that portfolio returns formed on consensus and revealed aggregates are the same. In case of excess returns, the null is that portfolio returns formed on consensus aggregate equals zero, and similarly for the revealed aggregate.

These results imply the rejection of the first null hypothesis, and in favor of the alternative—that information diversity matters for explaining the differences in stock recommendations by brokerage houses. This is because revealed aggregate scores differ from the simple average, and trading on it is more profitable than trading on the average.

Note that there is no monotonic relationship between portfolio and average returns. Given that this is weekly portfolio rebalancing (rather than daily portfolio rebalancing) and none of the portfolios formed on consensus earn excess returns, it is not that surprising. This is overall in line with the literature that finds trading on the level of consensus recommendation is unprofitable.

However, it is noteworthy that trading on the least favorable revealed aggregate

scores does not deliver lower returns than trading on the least favorable consensus scores. Simply put, the results are not symmetric. This is likely due to the known bias of analysts not to issue sell recommendations. Hong and Kubik (2003), for example, find that optimistic analysts have better career outcomes. Michaely and Womack (1999), among others, provide evidence that analysts tend to be more bullish if their brokerage house has an underwriting relationship with a stock, compared to analysts from non-affiliated brokerage houses. In line with this, it seems likely that analyst bias is asymmetric. If an analyst has negative information about a stock, s(he) may be reluctant to issue a sell recommendation and the analyst might issue a neutral (hold) recommendation instead of a negative recommendation. In this case, the assumption that forecast is reliable (calibrated) is violated. In fact, in the next section I present calibrated simulations and show that in this setting, the revealed aggregator will not necessarily perform better than consensus. For positive information, however, self-censorship and bias are a lot less likely. An analyst with positive information is likely to issue a positive recommendation regardless of whether the firm is a client or not. Therefore, the revealed aggregator can be trusted to leverage difference in information diversity for favorably rated stocks, but not necessarily for moderately or unfavorably rated stocks. This could explain why the results are not symmetric.

3.4.3 Portfolio performance evaluation

I employ the intercept test to check for abnormal performance of each portfolio. In one model specification, I use the CAPM model and in another specification, I use the three-factor model by Fama and French (1993). I therefore estimate the following two sets of weekly time series regressions:

$$R_{p\tau} - R_{f\tau} = \alpha_p + \beta_p \left(R_{m\tau} - R_{f\tau}\right) + \epsilon_{p\tau}$$
$$R_{p\tau} - R_{f\tau} = \alpha_p + \beta_p \left(R_{m\tau} - R_{f\tau}\right) + s_p SMB_\tau + h_p HML_\tau + \epsilon_{p\tau}$$

where $R_{p\tau}$ is the week τ return on portfolio p, $R_{f\tau}$ is the week τ return on treasury bills, $R_{m\tau}$ is the week τ return on CRSP value-weighted market index, α_p is the estimated CAPM intercept, β_p is the estimated market beta, $\epsilon_{p\tau}$ is the regression error term, SMB_τ is the difference between the week τ returns of a value-weighted portfolio of small stocks and that of large stocks, HML_τ is the difference between the week τ returns of a value-weighted portfolio of high book-to-market stocks to that of low book-to-market stocks.

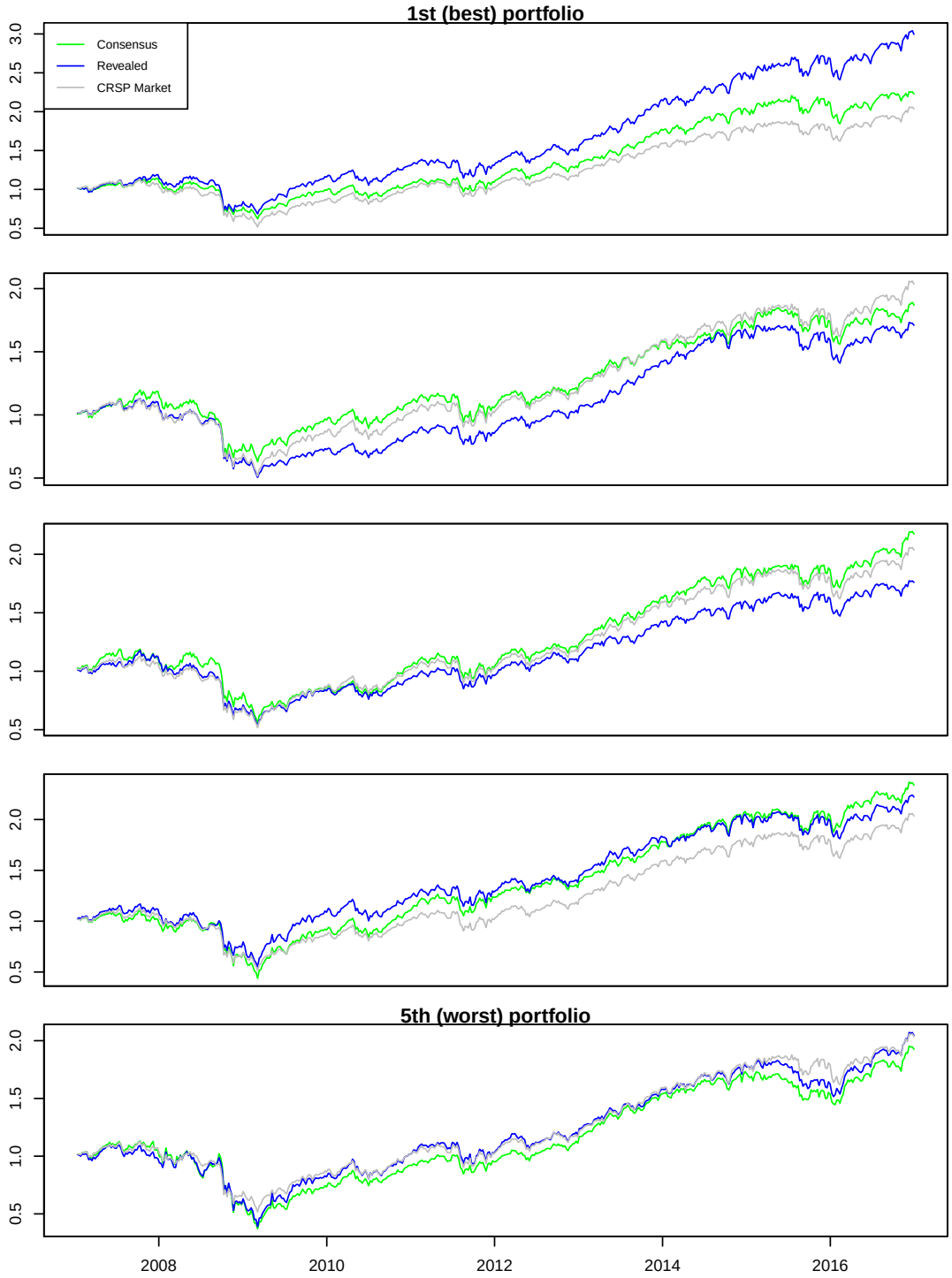
As reported in Table 3.6, a portfolio formed on the basis of revealed aggregate recommendation scores, with most favorably rated stocks, delivers a weekly abnormal return of 7-8bp, which is equivalent to 3.7-4.5% in annualized terms. This portfolio loads on value stocks, and is just about as risky as the market. There is no evidence of abnormal gross returns for a portfolio with most favorably rated stocks on the basis of consensus. This suggests that some of the information diversity could reflect private information. Note that all portfolios considered load on big stocks (with large market capitalization). Loading on big stocks is due to the data trimming algorithm that included only the most prolific brokers covering only the most popular stocks, with a target share of missing observations set to 50%.

Table 3.6: Abnormal performance of portfolios formed on the basis of consensus and revealed recommendations.

		Portfolios formed on the basis of									
		Consensus					Revealed				
		1st (best)	2nd	3rd	4th	5th (worst)	1st (best)	2nd	3rd	4th	5th (worst)
CAPM:	Intercept	0.028	-0.009	0.016	0.025	-0.014	0.084***	-0.024	-0.02	0.017	-0.003
		(0.029)	(0.026)	(0.025)	(0.041)	(0.056)	(0.031)	(0.028)	(0.024)	(0.034)	(0.056)
	Rm-Rf	0.906***	0.932***	0.98***	1.064***	1.145***	0.92***	0.917	0.933***	1.031***	1.151***
		(0.011)	(0.01)	(0.01)	(0.016)	(0.021)	(0.012)	(0.011)	(0.009)	(0.013)	(0.021)
Fama-French:	Intercept	0.016	-0.015	0.018	0.041	0.013	0.071***	-0.03	-0.023	0.026	0.023
		(0.023)	(0.022)	(0.025)	(0.035)	(0.043)	(0.025)	(0.026)	(0.023)	(0.031)	(0.043)
	Rm-Rf	0.991***	0.996***	0.982	0.989***	1.014***	1.006	0.968***	0.967***	1.002***	1.018***
		(0.01)	(0.01)	(0.011)	(0.015)	(0.019)	(0.011)	(0.012)	(0.01)	(0.014)	(0.019)
	SMB	-0.103***	-0.146***	-0.069***	-0.088***	-0.09**	-0.068***	-0.098***	-0.129***	-0.125***	-0.09**
		(0.021)	(0.02)	(0.022)	(0.031)	(0.038)	(0.022)	(0.023)	(0.02)	(0.027)	(0.038)
	HML	-0.287***	-0.184***	0.025	0.333***	0.558***	-0.304***	-0.156***	-0.078***	0.171***	0.561***
		(0.017)	(0.016)	(0.018)	(0.025)	(0.031)	(0.018)	(0.019)	(0.017)	(0.022)	(0.032)

Time series regressions of weekly portfolio returns. The 1st portfolio is one with the most favorably recommended stocks (to buy), while 5th portfolio is one with the least favorably recommended stocks (to sell). Sample data is weekly, Jan 6, 2007 - Dec 23, 2016, taking into account non-trading days. Returns are expressed in percent. Standard errors are in parentheses and the stars correspond to 0.1, 0.05, and 0.01 significance levels.

Figure 3.6: Geometric returns earned by all portfolios.



Compound returns of value-weighted portfolios, with weekly rebalancing, for $1 invested on Jan 6, 2007 to Dec 23, 2016.

As before, the finding of abnormal returns is asymmetric, in the sense that the portfolio formed on least favorably rated stocks does not deliver negative abnormal returns, while the one formed on most favorably rated stocks does deliver positive abnormal returns. The performance of these portfolios over time is also presented in Figure 3.6. These results suggest that a long-only trading strategy design may be able to exploit the issue of asymmetric bias in security analyst recommendations. A long-only strategy potentially enables trading on analysts' private positive information, which is like finding a needle in a haystack among very many buy and strong buy recommendations that analysts issue.

3.4.4 With transaction costs

The returns presented so far have all been gross returns. They do not take into account transaction costs that arise from the bid-ask spread, brokerage commissions, and the market impact of trading. It's important to incorporate these frictions when considering market efficiency hypotheses. I follow Barber and Odean (2001) and measure turnover for portfolio p during trading day τ as the percentage of portfolio's holdings as of the close of trading on the previous day $\tau - 1$ that has been sold off by the close of trading on date τ.

For each stock k in portfolio p as of close of trading on $\tau - 1$, I first calculate the fraction it would have comprised of the portfolio at the end of trading on date τ without portfolio rebalancing. This fraction is given by

$$G_{k\tau} = \frac{x_{k,\tau-1}(1 + R_{k,\tau})}{\sum_{k_p=1}^{K_{p,\tau-1}} x_{k,\tau-1}(1 + R_{k,\tau})}$$

where $x_{k,\tau-1}$ is as before the market value of equity for stock k divided by the total market capitalization of all stocks in portfolio p as of the close of trading on date $\tau - 1$, and $(1 + R_{k,\tau})$ is the gross return on the common stock k on date τ.

Next I calculate the actual fraction firm k makes up of portfolio p at the end of trading on date τ, as $F_{k\tau}$, after portfolio rebalancing as a result of changes in scores.

$$F_{k\tau} = \frac{x_{k,\tau}}{\sum_{k_p=1}^{K_{p,\tau}} x_{k,\tau}}$$

Comparing the two as percentages, $G_{k\tau} - F_{k,\tau}$ is the decrease in the percentage holding of each of the date $\tau - 1$ stocks. I sum this decrease (if any) across the stocks in the portfolio, yielding the week's portfolio turnover, as $U_\tau = \sum_{k_p=1}^{K_{p,\tau}} \max(G_{k\tau} - F_{k,\tau}, 0)$.

When trading on the revealed aggregator, portfolio turnover is about 1.5-2 times higher. This is in line with the fact that revealed scores are more volatile. Yet the magnitude of turnover is not very big in either case.

Table 3.7: Turnover in portfolios formed on the basis of consensus and revealed recommendations.

	Average weekly turnover					Average annual turnover				
Trading on	1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
Consensus	1.99	3.8	3.62	3.5	3.29	103.3	197.7	188.3	181.9	170.9
Revealed	3.66	3.35	3.23	3.56	4.4	190.4	174	168	185	228.6

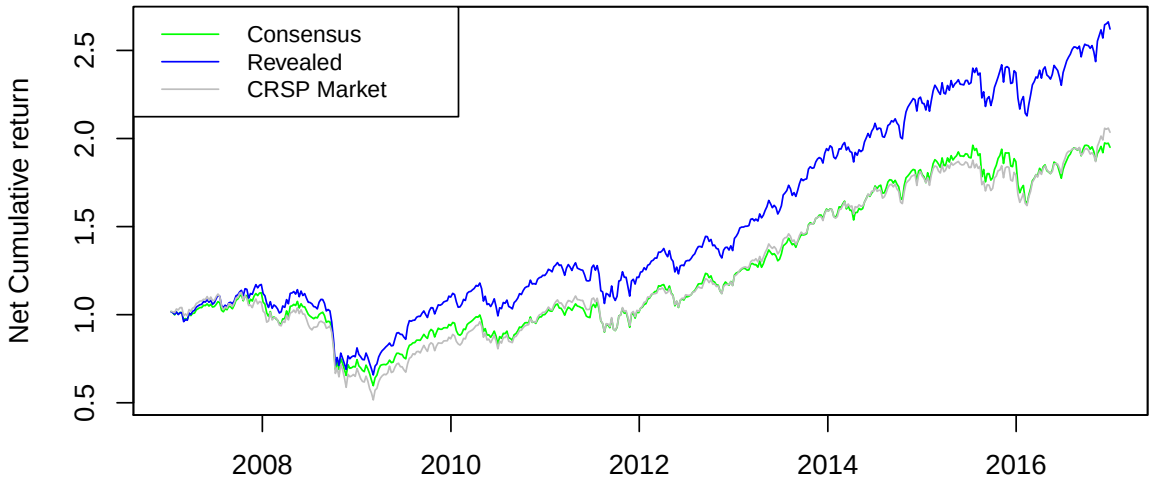
Portfolio turnover is reported in percent. Value-weighted portfolios, with weekly rebalancing. Sample data is weekly, Jan 6, 2007 - Dec 23, 2016, taking into account non-trading days.

As a very conservative estimate of transaction costs, I use 0.7 percent of share value traded for largest stocks. This estimate comes from Keim and Madhavan (1998), also used by Barber and Odean (2001), and represents the round-trip transaction costs that financial institutions incur when trading exchange-listed and Nasdaq stocks. Therefore I use 0.7% of turnover as the measure of how much investment returns are lowered by transaction costs. It should be noted that this represents a conservative estimate. As Frazzini et al. (2012) show, a large liquidity providing financial institution like AQR incurs transaction costs that are a magnitude lower, particularly for large and liquid stocks such as those included in these portfolios. This is due to direct

exchange access and automatic trade execution algorithms. Therefore my estimate of round-trip transaction costs provides an upper bound, and is a useful measure of returns possible for other types of financial institutions.

Net portfolio returns that have been adjusted for transaction costs are reported in Table A.8 and Table A.9 in the Appendix. Trading on revealed aggregator still beats trading on the consensus. Positive and statistically significant abnormal returns are still possible for going long on the portfolio formed on the basis of best revealed recommendation scores. The net abnormal returns are equivalent to 2.3-3% in annualized terms. Performance of this portfolio over time is reported in Figure 3.7 below.

Figure 3.7: Geometric returns earned by the first (best) portfolio, net of transaction costs.



Compound net returns of value-weighted portfolios, with weekly rebalancing, for 1$ invested on Jan 6, 2007 to Dec 23, 2016. These are portfolio returns for the 1st portfolio, with the most favorably recommended stocks (to buy), and net of transaction costs.

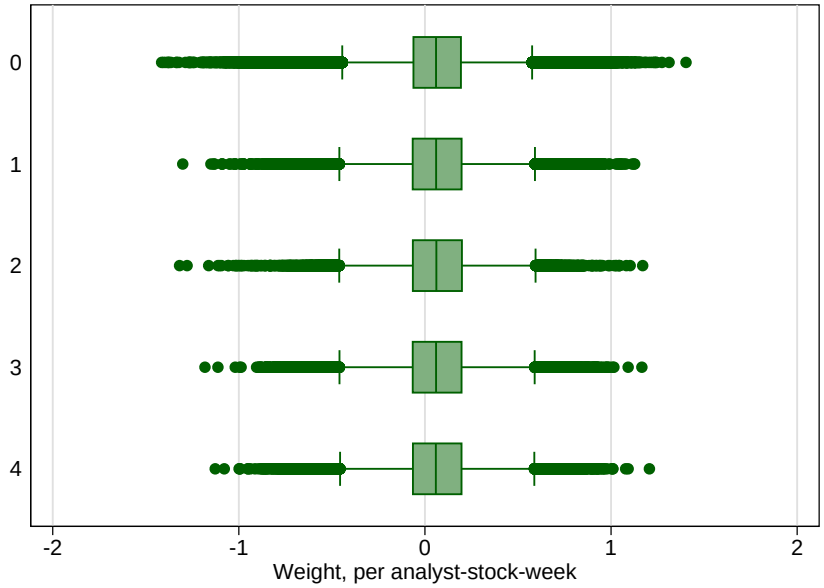
3.4.5 Information and skill

The findings of portfolio analysis support the argument that there is implicit information diversity among agents, which could arise because of different sources of

information. However, there is no direct way to test for the exact source of information and the analysis does not say what is driving the results. One possibility is that the revealed aggregator has an advantage over the consensus aggregator, in case of most favorably rated stocks, due to forecasting skill and presence of super-forecasters, and their depth of specific industry knowledge. The idea of agent-specific information sets could be interpreted as skill. Yet I do not find evidence in support of this interpretation.

If analyst skill is driving the results, then broker-stock-specific weights will not vary much over time during analysts' tenure, and will be related to the quality of analyst forecasting skill as measured by the All Star Analyst rankings by *Institutional Investor*. I use the data from *Institutional Investor* going back to 2001, and revealed aggregate scores data from the longer sample. Note that the weights range from -1.4 to +1.4, and are not bounded [0,1]. I find that the average time series standard derivation of weights is 0.2; the average difference between highest weight and lowest weight within the panel is also very high. This means that there are significant changes over time in weights, even though they are associated with the same stock and analyst. Additionally, I link the weights with All Star Analyst rankings by *Institutional Investor* and find no support for a relationship between the two. Lack of a relationship is visually illustrated in Figure 3.8 below.

Figure 3.8: Lack of a relationship between implied weights and All Star Rank by *Institutional Investor*



Data on broker-stock-specific weights used to construct the revealed aggregator is combined with All Star Analyst ranking. The total dataset includes 70 brokers, 3,625 analysts that are identified by name (analysts classified as "Permission Denied" have to be excluded from the analysis), and cover 1,031 stocks, over the course of 834 weeks in 2001-2016. A rank of 0 means not part of the All Star list, while the rank of 4 corresponds to First Team and best ranking on the list. A rank of 3 corresponds to Second Team, 2 corresponds to Third Team, 1 corresponds to being a Runner Up.

3.5 Simulated model with information diversity and noise

Any optimal estimator is context-specific. Formally, every estimation technique relies on identifying assumptions. In the case of this recently-developed revealed aggregation technique, the main assumption is that some agents know more than others. Yet there are several additional assumptions as well—partial information sets are agent-specific, and agents make predictions that are unbiased. The statistical framework also does not explicitly feature any noise when agents are making predictions based on their respective information sets.

How sensitive is this method to violations of these assumptions? It is important to consider that analysts might have noisy information, their predictions are biased, and their information set could be greater for some stocks, either due to expertise, experience, alumni connections, and so forth. In this section, I explore the sensitivity of this aggregation technique to violations of these assumptions, as a way to check that the setting of analyst recommendations is indeed appropriate for this statistical methodology. I do this with a simulated model where agents are differentially informed due to how information spreads from different sources, following information percolation theory (Duffie and Manso, 2007). The model relies on individual private signals, which agents then exchange in random bilateral meetings, representing word-of-mouth communication. The model thus features both information diversity and noise. This allows me to compare the performance of the revealed and consensus aggregators, depending on the extent of information diversity, the noisiness of the signals, and the information structure of agents. I also explore several scenarios with bias.

3.5.1 Information percolation

Let's consider an explicit mechanism that causes partial and overlapping information sets, and also allows for some noise. As before, we are in a setting with $j = 1, ..., J$ agents that issue forecasts X_{jk} across K outcomes. At $t = 0$, each agent j obtains a private noisy signal $\tilde{\theta}_{jk}$ about true outcome Y_k:

$$\tilde{\theta}_{jk} = \frac{1}{J} Y_k + \tilde{e}_{jk}$$

where e_j is drawn from a Normal distribution with mean zero, independent of all other $\tilde{e}_{-jk}$. The noisiness of the signal is the same across all agents, $\tilde{\sigma}^2 \forall j, k$. In expectation, true outcome Y_k is the sum of all the signals received by J agents,

$Y_k = E\left(\sum_{j=1}^{J} \tilde{\theta}_{jk}\right)$. This closely corresponds to the idea that sigma-field $\sigma(\mathcal{F})$ is a collection of all possible information about Y from information partitions $\mathcal{F}$. For convenience, information partitions are assumed to be additive, and so are the original signals.

At date $t = 1$, agents meet each other randomly and share information. Meetings take place continuously at Poisson arrival times with intensity λ. In these bilateral meetings, agents exchange their initial signal $\tilde{\theta}_{jk}$ and any other signals they received during previous meetings, if any[10]. Therefore, agents start off with one signal at $t = 0$ and end up with different numbers of signals at $t = 1$. I model the additional number of signals to gather, S, by the Poisson distribution, where the probability of gathering S signals is $Pr(X = S \mid \lambda) = \frac{\lambda^S e^{-\lambda}}{S!}$. The Poisson process determines the number of additional signals $S \in [0, J-1]$ that agent j gathers. An agent j could gather no additional signals, with $S = 0$, some additional signals, or even become fully informed and collect $S = J - 1$ additional signals.

It is random who the other S agents are that j collects signals from, for each event K. All information is exchanged in bilateral meetings. The number of agents is finite. This means that if an agent i has to meet agent j because j is one of the S agents that i has to gather a signal from, but j doesn't have to meet i because i is not one of the S agents that j has to gather a signal from, then j still receives the signal from i. In this example, agent j ends ups with $S + 1$ signals. Let's call these additional signals that an agent passively receives, rather than gathers, as R.

Then the agents provide and estimate X_{jk} by adding up their one original signal, and their additional signals S and R.

$$X_{jk} = \tilde{\theta}_{jk} + \sum_{s=1}^{S} \tilde{\theta}_{sk} + \sum_{r=1}^{R} \tilde{\theta}_{rk}$$

[10]Following Andrei and Cujean (2017), I assume that they tell the truth. I similarly assume that meetings do not overlap and there is no persuasion bias in the sense of one agent's original information coming back to her.

As a result, there is both information diversity and noise in this setting. Each agent j has a personal partial information set $\mathcal{F}_j \subseteq \mathcal{F}$. Whether an agent has more or less information is determined by the total number of signals S that have been collected, which is drawn from a Poisson distribution. Yet even if two events have an identical true outcome Y_k, an agent j's forecast of these two outcomes will differ, because the signals have been collected from different meetings, and all signals are noisy. This allows me to simulate both a varying degree of information diversity, via the Poisson intensity parameter λ, and the amount of noise, via the variance of the original signals $\tilde{\sigma}^2$.

3.5.2 Calibrated simulation

True outcome Y_k is drawn from a uniform distribution such that $Y_k \in (-10, 10)$. Once the signals for each agent are drawn, then information spreads. I use the Poisson distribution with intensity parameter λ to determine the number of signals each agent will gather, regardless of the stock K. Each agent becomes differentially informed, to the same extent across the K stocks, and produces a forecast X_{jk}. These forecasts X_{jk} for K outcomes across J agents serve as inputs to calculate the estimates of true outcome Y_k with the revealed aggregator method, and by taking a simple average (the consensus method). Keeping the number of agents and the number of events fixed, at $J = K = 30$, and the amount of noise set at $\sigma = 0.25$, I vary the degree of information diversity by varying the Poisson intensity parameter λ. I run 100 Monte Carlo simulations, to keep computations manageable.

Table 3.8 first reports statistics measuring information overlap, in terms of the distribution of the total number of signals, $S + R$, that agents have for a given λ. Each agent starts off with one signal, and low $\lambda = 0.3$ means gathering 0.3 other signals on average, plus any additional signals received if someone randomly meets you. For $\lambda \in [6, 12]$, there is some information overlap, with some agents that have

gathered relatively few signals, and some agents that are nearly completely informed. At one extreme, with $\lambda = 0.3$, information is very dispersed, with little overlap, and at another extreme, with $\lambda = 22.5$, information is very concentrated, with a lot of overlap. This can be seen from the distribution of the total number of signals.

Next, Table 3.8 compares the absolute error, $|Y_k - X_{jk}|$, for the Consensus and Revealed aggregator, for all stocks K, across 100 simulations, in a one-sided paired t-test. The last two columns also report the overall root mean square error across 100 simulations and for all stocks, under the two aggregation methods. Since Y_k is drawn from a uniform distribution such that $Y_k \in (-10, 10)$, this range could be thought of as percent return. Therefore a root mean square error of 5 percent is high, and 0.5 percent is low.

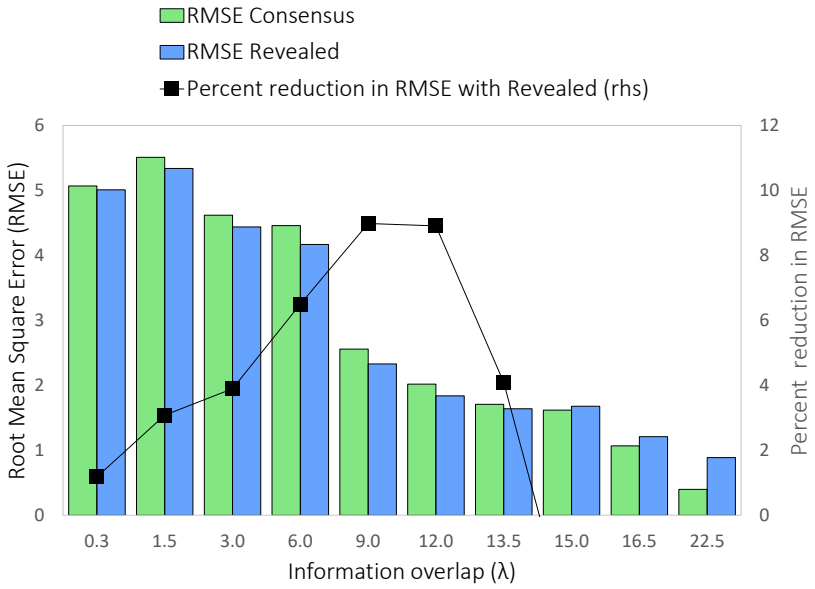
The results from this information percolation simulation show that when there is some information diversity, then estimates using the revealed aggregator are better (closer to the true value and therefore more precise) than the consensus estimates. Specifically, revealed beats consensus so long as agents on average receive less than 2/3 of total information. In case of low information diversity and high information overlap, the revealed aggregator is no longer as good as the consensus. This is the case when the majority of agents have obtained a very high proportion of total signals, information overlap is very high, and most of the information is common knowledge. Furthermore, there seems to be an optimal amount of information diversity and overlap where percent reduction in root mean square error—with the revealed aggregator—is greatest. This is visually illustrated in Figure 3.9, which reports the same information as Table 3.8. The peak improvement from using the revealed aggregator occurs at $\lambda \in (9, 12)$ in this calibrated simulation. That is when a median agent has a total of 14-19 signals out of possible 30.

Table 3.8: Performance of the two aggregators when varying information diversity.

	Total number of signals					Absolute Error	Root Mean Square Error	
λ	min	25th	median	75th	max	T-test Ha	Consensus	Revealed
0.3	1	1	1	2	5	Con>Rev***	5.07	5.01
1.5	1	3	3	4	10	Con>Rev***	5.51	5.34
3.0	1	4	6	7	15	Con>Rev***	4.62	4.44
6.0	2	8	10	12	23	Con>Rev***	4.46	4.17
9.0	5	12	14	17	30	Con>Rev***	2.56	2.33
12.0	8	16	19	23	30	Con>Rev***	2.02	1.84
13.5	9	18	21	25	30	Con>Rev***	1.71	1.64
15.0	10	19	23	28	30	Rev>Con***	1.62	1.68
16.5	12	21	26	30	30	Rev>Con***	1.07	1.21
22.5	16	27	30	30	30	Rev>Con***	0.40	0.89

Simulation of the noisy information percolation model, where the number of agents and number of stocks is kept constant, at $J = K = 30$, the amount of noise is set at $\sigma = 0.25$, and true outcome Y_k is drawn from a uniform distribution such that $Y_k \in (-10, 10)$. The Poisson intensity of meeting arrivals λ increases down the column, which corresponds to increasing information overlap and concentration. Higher absolute or root mean square error means the estimates are less precise, so a lower error value indicates better performance.

Figure 3.9: Performance of the two aggregators when increasing information overlap.



Information overlap increases along the x axis. The number of agents and number of stocks is kept constant, at $J = K = 30$, the amount of noise is set at $\sigma = 0.25$.

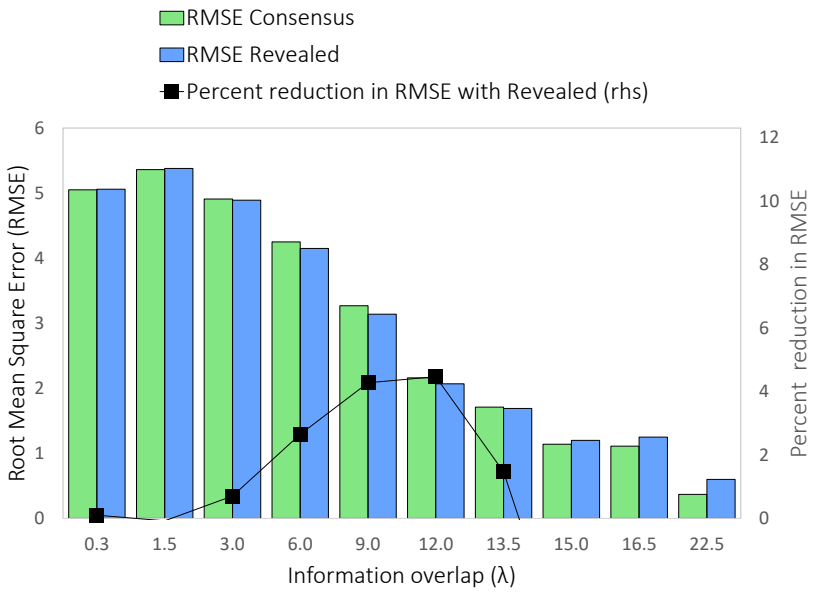
In another simulation, I vary the number of stocks K and find that these results

are robust. Varying the number of agents J corresponds to changing the extent of information diversity, since a higher number of agents means a greater number of information partitions and more dispersed information. It is mechanically equivalent to changing the Poisson intensity parameter λ which is already shown in Table 3.8.

3.5.3 Assumptions and sensitivity analysis

Next, I examine the sensitivity of these aggregators to the assumption of agent-specific information sets, rather than agent-event-specific information sets. As an alternative model, I make a new draw from the Poisson distribution for each stock K. Each agent becomes differentially informed and to a different extent across the K stocks. In this case, revealed aggregator still outperforms the consensus aggregator when there is a moderate amount of information overlap. However, overall the benefits of using the revealed aggregator are less pronounced. This is illustrated in Figure 3.10, which only differs from Figure 3.9 in the specification of information sets.

Figure 3.10: Performance of the two aggregators when information sets are agent-event-specific.

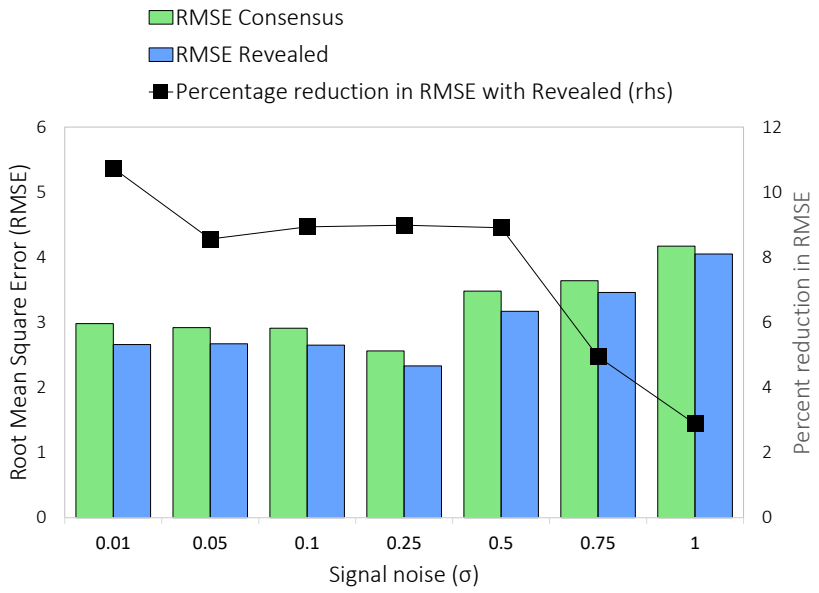


Information overlap increases along the x axis. The number of agents and number of stocks is kept constant, at $J = K = 30$, the amount of noise is set at $\sigma = 0.25$.

This noisy information percolation framework also allows me to examine the sensitivity of these aggregators to noise. Going back to agent-specific information sets, I vary the noisiness of the original signals that the agents receive, by varying σ. As seen from Figure 3.11, the findings are somewhat sensitive to noise. The benefits of using the revealed aggregator, relative to the consensus, decrease as more noise is added to the system. With increasing noise, the revealed aggregator incorrectly places higher weights on the more bullish and bearish forecasts, which are more extreme due to noise and not a higher number of signals for that agent. However, in the presence of some information diversity, for moderate levels of noise, the revealed aggregator still always works better than the consensus. The amount of noise corresponding to $\sigma = 1$, for example, is very high. With $Y_k \in (-10, 10)$, and $E[\theta_{jk}] \in \left(-\frac{1}{3}, \frac{1}{3}\right)$, the noise term $e_j \sim N(0, 1)$ with such high variance results in θ_{jk} that could range from -2 to 2 in the tails of the distribution. In contrast, the amount of noise correspond-

ing to $\sigma = 0.25$ is in line with a reasonable calibration and therefore represents the benchmark used. Table A.10 in the Appendix reports varying noise and information overlap simultaneously.

Figure 3.11: Performance of the two aggregators when increasing noise.



Noisiness of the signals, $\sigma = \sqrt{(V[\theta_{jk}])}$, increases along the x axis. The range of signals means is $E[\theta_{jk}] \in \left(-\frac{1}{3}, \frac{1}{3}\right)$. Information diversity is fixed at $\lambda = 9$. The number of agents and number of stocks is kept constant, at $J = K = 30$.

Additionally, I explore the sensitivity of the main results to the other assumption of the revealed aggregator model. I relax the assumption of unbiased forecasts and test the following four scenarios with bias:

i All agents are equally optimistically biased: the original signals include bias, $\tilde{\theta}_{jk} = \alpha + \frac{1}{J} Y_k + \tilde{e}_{jk}$ where $\alpha = 1 \forall j, k$.

ii Agents have an optimistic bias for stock predictions by their clients: the original signals include bias if the stock is issued by the agent's client, with $\tilde{\theta}_{jk} = \alpha I_{client} + \frac{1}{J} Y_k + \tilde{e}_{jk}$ where $\alpha = 1 \forall j, k$.

The indicator variable I_{client} is a draw from a Bernoulli(p) distribution and denotes

whether the stock is issued by agent's client and takes on 1 or 0 with probability $p = \frac{1}{2}$.

iii All agents are equally reluctant to make negative predictions: any negative signal becomes zero, with
$\tilde{\theta}_{jk} = I_{\frac{1}{J}Y_k + \tilde{e}_{jk} < 0,}(\frac{1}{J}Y_k + \tilde{e}_{jk})$
where the indicator variable denotes whether the stock carries a negative return signal.

iv Agents are reluctant to make negative predictions for their client's stocks: any negative signal becomes zero if this stock is issued by a client, with
$\tilde{\theta}_{jk} = I_{\frac{1}{J}Y_k + \tilde{e}_{jk} < 0, client}(\frac{1}{J}Y_k + \tilde{e}_{jk})$
where the indicator variable denotes whether the stock carries a negative return signal and at the same time is issued by a client, which occurs independently with probability $p = \frac{1}{2}$.

I find that when there is bias, such as when all agents are optimistic across the board or treat negative signals as neutral, the revealed aggregator fails to beat the consensus. The question of bias is therefore critical for this technique.

Table 3.9: Performance of the two aggregators when predictions are biased.

Scenario	Revealed better than Consensus?	Root Mean Square Error	
		Consensus	Revealed
Baseline	Yes	2.56	2.33
Bias: optimistic	No	14.91	17.65
Bias: optimistic for clients	No	8.13	8.91
Bias: no negative predictions	Unclear	4.47	4.47
Bias: no negative predictions for clients	Unclear	3.79	3.70

Simulation of the noisy information percolation model, where the number of agents and number of stocks is kept constant, at $J = K = 30$, the amount of noise is moderate at $\sigma = 0.25$, and the information overlap is also moderate at $\lambda = 0.9$. The true outcome Y_k is drawn from a uniform distribution such that $Y_k \in (-10, 10)$. In the baseline, there is no bias and information sets are agent-specific. Higher absolute or root mean square error means the estimates are less precise, so a lower error value indicates better performance.

Altogether, these robustness and sensitivity simulations confirm that revealed aggregator works better than the consensus under the following conditions: when information diversity is present, the noisiness of the signals is low or moderate, agents' predictions are unbiased, and partial information sets are agent-specific, rather than agent-stock-specific. These correspond to the assumptions of the model. However, when signal noise is high or when partial information sets are agent-stock-specific, the benefit of using the revealed aggregator is less pronounced. Lastly, the revealed aggregator fails to perform better than consensus when everything is common knowledge and there is no information diversity, or when signals are biased.

3.5.4 The context of security analysts

The context of security analysts holds up to main identifying assumptions regarding information diversity, noise and information structure. The calibrated simulation confirms that when agents have different but relevant information, and some know more than others, the revealed aggregator indeed beats consensus. As argued in the introduction, security analysts plausibly have different sources of information. Although their information might be noisy, analysts likely have low to moderate levels of noise since these are professional forecasters. The benefit of using the revealed aggregator should still be in place in the presence of some noise, as shown in the calibrated simulation. Furthermore, analysts may specialize in a certain subset of stocks or may have additional unique information only for select stocks among all the stocks that they cover, which would make their information structure agent-stock-specific rather than agent-specific. The calibrated simulation shows that the revealed aggregator is an improvement over consensus even in this case. Altogether, this confirms that the results of portfolio analysis, which relied on the revealed aggregate recommendation scores, should be robust and their conclusions sound.

Yet the calibrated simulation also clearly shows that bias is problematic. To

the extent that analyst bias is asymmetric, the long-only portfolio design strategy is able to exploit this bias. The fact that portfolios formed on the revealed aggregate recommendation scores do not perform worse than portfolios formed on consensus scores is in line with this intuition, since the revealed aggregator fails to effectively leverage differences in information when predictions are biased. Among numerous buy and strong buy recommendations that analysts issue, it may be difficult to infer which analysts base their favorable forecasts on common positive information and which recommendations reflect additional positive information which may be private. The long-only trading strategy formed on the basis of revealed aggregate recommendation scores delivers an abnormal return because it potentially enables trading on positive private information, akin to finding the needle in a haystack.

3.6 Conclusion

In this paper, I use a simulated noisy information percolation model to quantify and test a recently developed statistical aggregation technique. I show how this technique can infer differences in information across agents and combine agents' heterogeneous predictions more efficiently than a consensus estimate. I also provide a first finance application of this aggregation technique—to stock recommendations of security analysts—which are particularly well suited due to data properties and assumptions. Using analyst recommendations, I develop a trading strategy which not only outperforms a consensus-based strategy by 3% on an annualized basis, but also delivers an economically significant net abnormal return.

These finding thus show that information diversity matters when assessing the value of analyst recommendations. More generally, this paper highlights the prevalence of information diversity as a phenomenon that is significant and tradable, and why disaggregated and crowdsourced forecasts are useful. This paper shows how

information diversity can be empirically measured, and demonstrates the economic significance of this phenomenon by means of a profitable trading strategy.

Areas for future research include a closer look at the source of information that this statistical technique can leverage, and this can be addressed with a survey or a controlled experiment. Furthermore, future research calls for theoretical improvements that incorporate time series data into the revealed aggregator methodology. Mutual fund flows provide an interesting empirical application to examine information content of fund managers, using an improved version of the revealed aggregator that exploits both the historical fund flows as well as the cross-section of different asset classes and securities.

Chapter 4

The Impact of Demographic Factors on Sovereign Debt Maturity

4.1 Introduction

For any given level of debt that needs to be issued, a national Treasury makes several choices: currency to issue debt in, indexation (for example to inflation), and maturity of the instruments. Maturity is the length of time until repayment of debt, measured in months or years, and it is the focus of this paper. As shown in Figure 4.1, there is considerable time-series variation in observed maturity outcomes (Panel A). Yet even among the countries that appear to keep maturity stable over time (Panel B), there is cross-country variation. How do national treasuries manage government debt and decide about its maturity profile?

The question of government debt maturity structure is significant for several reasons. First, maturity choice determines debt repayment profile and the cost of debt service. A shorter maturity, for example, means a more heavily front-loaded debt

redemption profile for the borrower, and it brings with it a higher rollover risk. Any unexpected rise in interest rates will much more swiftly translate into higher cost of debt service for the borrower. Let's compare the maturity structure of the United States and the United Kingdom as an example. The debt profile of the United States has remained stable over time, with the weighted average remaining maturity of government debt ranging from 4 to 6 years since 1980. This is a standard metric: average maturity of all outstanding government bonds, weighted by the face value of each bond. In the United Kingdom, in many aspects similar to the United States, the maturity profile is much longer and ranges from 10 to 14 years. If the UK's debt maturity profile was like that of the US—much shorter—then in 2011 this would have translated into £9.2bn per year more in interest costs for UK Treasury (Alloway, 2011). The impact of maturity choice on the cost of debt service is economically significant.

The second reason why government debt maturity structure is important is that debt manager's maturity choice can have an effect on asset prices and the real economy[1]. For example, if a longer maturity structure of government debt translates into higher term spread, this increase in long-term borrowing rates will likely lead to lower corporate investment and consumption, which has a contractionary effect on the macroeconomy. The maturity structure of debt has in fact been used as an economic policy tool in several historical cases: post-WWII debt management in the UK, US Operation Twist in 1961, and Quantitative Easing (QE) in Japan in the

[1]As an aside, it is true that in an economy where agents are perfectly forward-looking, and where taxation is lump sum, the exact method of financing a government's budget deficit does not matter. For example, if the government finances its deficit with debt, perfectly forward-looking consumers internalize that taxes will increase in the future to repay this debt and save more in the present. This is the Ricardian equivalence result. It extends to the choice of maturity as well, not just debt and taxation level. Similarly, if there is perfect substitutability between instruments of different maturities, then the maturity structure does not matter for asset prices. However, the setting in which government debt management is practiced includes non-lumpsum taxation (which breaks Ricardian equivalence) and/or imperfect substitutability between instruments of different maturities. In this case, both the level of debt and its maturity choice can have an effect on the real economy and asset prices.

1990s. Maturity structure has also been used as a policy tool during recent QE in 2010-2012 when several Central Banks have shortened the maturity of publicly-held debt by purchasing long-term government bonds in exchange for short-term bonds or reserves. This was intended to have real effects on the economy. These effects have been found to be small but significant, in both empirical and theoretical work[2].

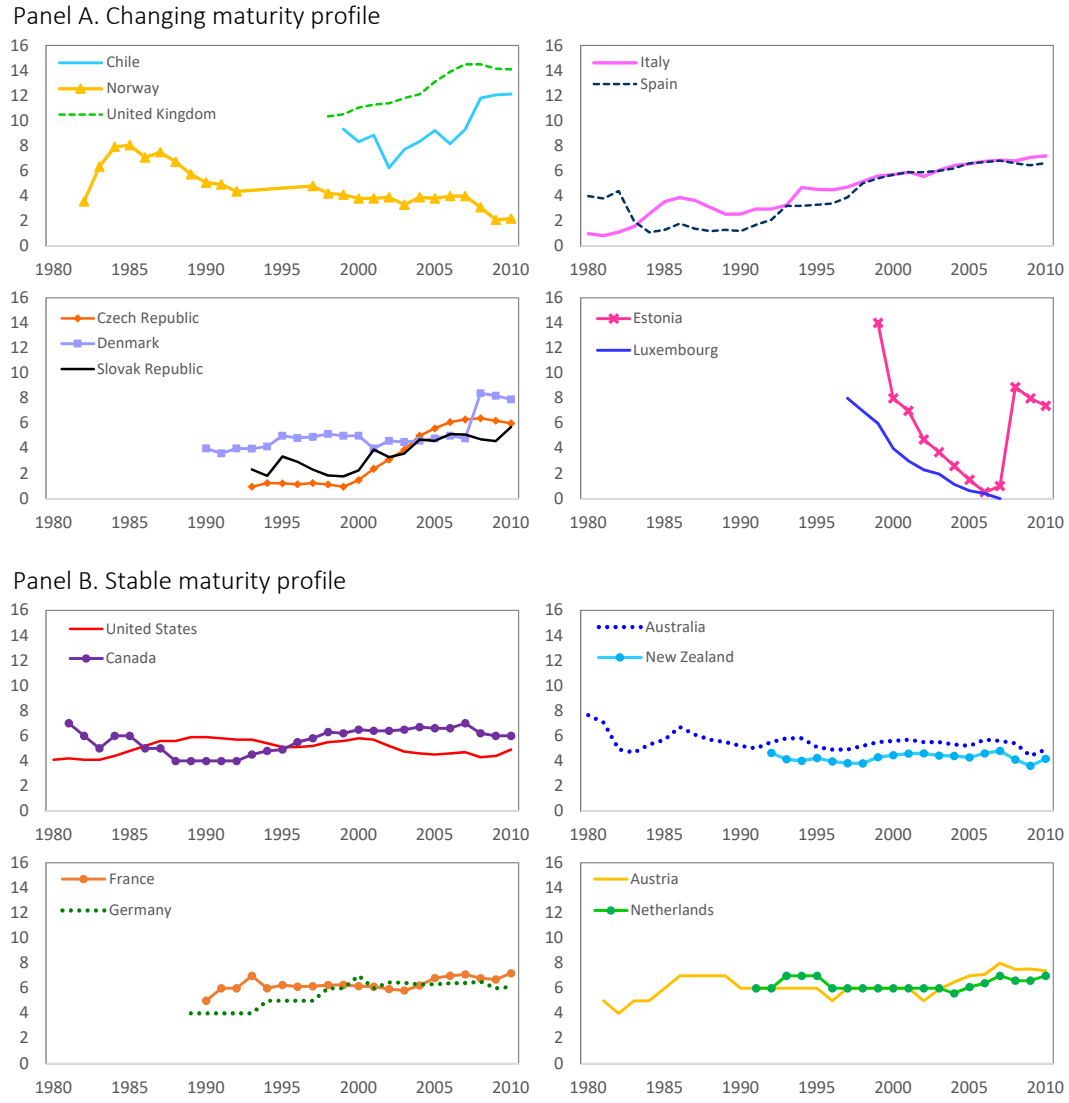
To understand and predict prices and yields of sovereign bonds, we need to know what debt managers respond to. In the U.S. alone, the Treasury market as of 2015 was $13tr, comparable to the $25tr equity market that is the focus of most finance research to date. Hence we care about the question of government debt and its maturity structure as investors, policymakers, taxpayers, and consumers.

The third reason why maturity structure is interesting to study is that empirical analysis of debt manager's choices is scarce. Existing empirical literature has included occasional stylized facts or has focused on individual determinants in line with specific theories[3]. The empirical papers closest to my analysis are Goudswaard (1990) and Gonzalez-Fernandez and Gonzalez-Velasco (2014). Goudswaard's analysis of public debt maturity in the Netherlands in 1960-1985 did not include demographics, while Gonzalez' analysis that focused on socio-demographic factors did not fully investigate whether debt managers cater perfectly or imperfectly to demographic clienteles. In contrast, this paper fills the gap in empirical research of how national treasuries manage sovereign debt by comprehensively testing many alternative debt management theories as competing hypotheses, by using reduced-form specifications from formal theories.

[2]See Krishnamurthy et al. (2011), Krishnamurthy and Vissing-Jorgensen (2012), D'Amico and King (2013), Joyce et al. (2011), Gagnon et al. (2011) and Kuttner (2006) for empirical evidence. Exiting theoretical approaches to model central bank intervention in the market for long-term government securities in a general equilibrium setting that use a mechanism in the spirit of preferred habitat include Ellison and Tischbirek (2014), Chen et al. (2012), while Chadha et al. (2013) rely on liquidity preference for short-term bonds.

[3]See Missale and Blanchard (1994) and Missale (1999), Broeck and Guscina (2011), Abbas et al. (2014), Guibaud et al. (2013), Eichengreen and Hausmann (1999).

Figure 4.1: Maturity structure of government debt in select advanced economies



Weighted average remaining maturity of marketable central government debt. In years. Annual data from 1980 to 2010. Source: OECD Central Government Debt Database.

Before turning to alternative debt management theories, it is helpful to consider what are the national mandates of debt management offices. The stated mandates usually entail some form of maximizing the expected return of a debt strategy while avoiding excessive risk (see, for example, the UK Report of Debt Management Review by Portes and Tucker (1995) and Canadian Debt Strategy Model by Bolder (2008)). According to survey results in Missale (1999, chap. 1), all representatives of OECD national treasuries share the objective of containing interest rate costs, the desire to insulate government portfolio from future interest rate shocks, and are concerned with refinancing risk—the risk of having to roll over a large amount of debt at unfavourable interest rates. The best practice in sovereign debt management is to determine maturity structure in cost-risk portfolio theory framework based on forward-looking simulations in a dynamic macro model (Missale, 2012 and Bolder, 2008). Consistency with monetary policy is also often listed as a desirable objective, in the context of not interfering with financial market conditions. Yet it is not clear what such different statements could mean in practice, which risks weigh more, or which strategies are actively pursued by government debt managers.

There are several theoretical models of optimal government debt management, and I use each of them in this paper. Each model provides different prescriptions for the choice of maturity structure in response to a particular variable. Take the classic partial equilibrium portfolio theory as the first example. The debt manager is risk averse and chooses maturity structure so as to minimize the cost of servicing the debt while avoiding excessive rollover risk. Dottori and Manna (2016) show that in case of a perpetual planner, this entails issuing more shorter-term debt the steeper the slope of the yield curve.[4] Over the short horizon, when term premium is positive, and therefore long-term debt is relatively more expensive compared to rolling over short-term debt, portfolio theory prescribes shortening the maturity structure.

[4]With an adjustment for debt manager's risk parameter, volatility of interest rates, and covariance between the short and the long-term rate

In a general equilibrium setting, on the other hand, there is scope for maturity structure to smooth taxation over states and over time, thereby smoothing the consumption profile of a representative agent. A debt manager that wishes to pursue tax smoothing will set maturity structure so as to hedge against sources of uncertainty in output and government spending. Such debt manager thereby minimizes fluctuations in tax rates (Bohn, 1990). Alternatively, an overlapping generations environment is also useful when thinking about government debt management. In this environment, the young generation has a long investment horizon. The share of savers in the population corresponds to this long-horizon clientele, while the share of retirees corresponds to the old generation. Shifting demographics and growing portfolios of pension funds are a known source of demand for long-dated government bonds. In fact, the notion that institutional investors—particularly pension funds and insurance corporations—are clienteles in the long-term government bond market goes back to Modigliani and Sutch (1966).

Guibaud, Nosbusch, and Vayanos (2013) use an overlapping generations model to show why the sovereign debt manager should cater to demographic clienteles, yet imperfectly. When facing an increase in the share of savers in the population and a corresponding downward pressure on the long-term interest rates, the optimal response is to increase the supply of long-term bonds. Yet it's optimal to cater imperfectly and let the long-term interest rate fall due to demand pressure. This is because a welfare-maximizing debt manager internalizes the fact that catering to a growing long-horizon clientele imposes more risk on future generations through a more uncertain tax burden. Such welfare-maximizing debt manager therefore allows the long-term interest rate to decrease. A cost-minimizing debt manager, on the other hand, would cater to increased demand fully, and there would be no change in the long-term interest rate.

Furthermore, any debt-related policy is also subject to the time-inconsistency

problem. Debt holders face the risk of surprise inflation and the possibility that debt will be inflated away. A welfare-maximizing debt manager who wants to show commitment to an anti-inflationary stance will pick an overall maturity structure that is a negative function of real debt service (Missale and Blanchard, 1994). Intuitively, the greater the debt level, the greater the temptation to inflate it away. A shorter maturity is required so as to prevent this outcome.

These are the main normative debt management prescriptions in a nutshell. In this paper, I review these formal models, provide reduced-form specifications of maturity determinants, and test them as competing hypotheses. Where theory specifies steady-state maturity structure, I test this in the panel setting with advanced economies using GMM models, for the period 2000-2010. Where theory specifies time-varying maturity structure, I test this in a time series setting for the United States, 1972-2013. Maturity structure has unit root properties, but rather than treating this as a hindrance, I exploit this with an equilibrium correction mechanism (ECM), and estimate theoretically-consistent long-run and short-run relationships.

I find widespread empirical support for the hypothesis that debt managers pursue cost minimization and also adjust to demographic clienteles. Results from panel analysis of OECD countries suggest that countries with higher expected interest rate term spread have a shorter maturity, after appropriately controlling for co-movement of interest rates. Additionally, countries with a higher share of savers in the population or higher growth in assets of long-horizon institutional investors have government debt of longer maturity. The conclusions are similar for the United States based on forty years of quarterly data. Short-run U.S. maturity structure can be predicted by past term premium, which is in line with the hypothesis that government debt managers chose maturity so as to minimize debt service. In the long-run, there is a positive equilibrium relationship between maturity structure and change in the share of savers in the population. There is also a negative relationship between change in the share of

savers and the term spread. This trivariate equilibrium relationship lends relevance to the clienteles model of debt management by Guibaud, Nosbusch, and Vayanos (2013). These findings confirm imperfect catering to demographic clienteles, as prescribed by theory.

One of the contributions of the analysis is to highlight the effect of demographic factors as determinants of government debt maturity choice, and the extent to which government debt managers cater to demographic clienteles, such as pension funds and insurance corporations. As such, my analysis contributes to the literature on habitats and clienteles in financial markets. A key recent theoretical contribution in the literature on habitats is a formal model by Vayanos and Vila (2009), where limits to arbitrage arise from risk-averse arbitrageurs that substitute across maturities, and each maturity has its own clientele. The work by Guibaud et al. (2013) builds on it with a normative analysis that underscores that it is optimal for the government to facilitate the sharing of risk across generations with the choice of maturity structure. The risk-sharing role of the government debt itself goes back to Fischer (1983) and Gale (1990).

In contrast, recent literature has explored asset-pricing implications of supply shocks in the market for government debt (Greenwood and Vayanos, 2010, 2014). My analysis provides further empirical evidence that debt manager's supply of government bonds of particular maturity is itself a function of the term structure. It is also a function of demographic factors and catering to clienteles such as pension funds and insurance corporations. More generally, this paper complements the literature that finds that company managers cater to investor sentiment and market clienteles in the corporate bond market (Baker et al., 2002; Greenwood et al., 2010). This catering argument was first theoretically proposed by Stein (1996). Last but not least, by testing competing hypothesis regarding optimal government debt management, this paper also contributes to the vast literature that addresses optimal public policy, tax

smoothing, and the time-inconsistency aspect of government debt (literature that sits at the intersection of macroeconomics and finance).

The paper proceeds as follows. Section 4.2 discusses the definitions, identities, and dynamics of maturity structure. Section 4.3 and reviews the theories of optimal debt management and derives reduced-form equations of maturity as a function of some variables that debt manager responds to according to each theory. I draw a distinction between steady-state theoretical predictions, which can be tested by looking at cross-country differences, and time-varying theoretical predictions, which can be examined through a time-series analysis. Section 4.4 presents an empirical model that tests alternative theories as competing hypotheses. Section 4.5 presents panel results while Section 4.6 presents results of the dynamic analysis. Section 4.7 concludes.

4.2 Definitions, identities, and dynamics

The first question in the analysis of maturity structure is how to measure it (see Table B.1 of the Appendix). The share of short-term debt, defined by maturity of less than one year, for example, is one such measure, but it does not tell us about the share of two-year bonds as opposed to twenty-year bonds in debt manager's strategy. The share of long-term debt, broadly defined, is the measure of maturity that is widely used in theoretical models that I discuss in this paper. Portfolio and issuance shares are more granular, but bond issuances at the high end of the maturity spectrum are infrequent. Weighted average remaining maturity, $WARM$, is the most conventional measure, in debt stock terms, that has been used in the literature and is the primary measure that I use in the econometric analysis. Its equivalent measure in debt flow terms is the weighted average original maturity, $WAOM$. Duration, on the other hand, conceptually measures an elasticity and that is not the focus of this paper.

There are several formulations of the maturity structure that illustrate that it

is by definition a persistent process. One formulation says that the current period weighted average maturity of outstanding debt is determined by last period maturity that is reduced by one, the debt that has matured, as well as the new debt and its maturity structure (Goudswaard, 1990):

$$WARM_t = (WARM_{t-1} - 1)\frac{D_{t-1}^{TOT}}{D_t^{TOT}} + WAOM_t\frac{D_t^{NEW}}{D_t^{TOT}}$$

It is also informative to illustrate persistence of maturity structure in a setting where there are only short-term securities with one period maturity, D_{ST}^{TOT}, and long-term securities of some maturity T, D_{LT}^{TOT}, with portfolio shares in total debt as $\beta = \frac{D_{LT}^{TOT}}{D^{TOT}}$ and $1-\beta$ respectively. In the steady-state, total debt stock remains constant[5] and the corresponding weighted average remaining maturity in this two-security steady-state is:

$$WARM = (1-\beta) + (\frac{1+T}{2})\beta$$

where β is the share of long-term debt and T is its term to maturity. This formulation further illustrates the extent of persistence in $WARM_t$ when debt manager can only determine β each period, while T is high.

In this paper, I focus exclusively on central government debt, i.e., obligations issued by national treasuries, and on the instruments that are in the typical toolset of treasuries. Based on available data, the focus is either on marketable or total debt, without a distinction as to debt holders.

[5]The total debt stock is $D^{TOT} = D_{ST}^{TOT} + D_{LT}^{TOT} = (1-\beta)D^{TOT} + \beta D^{TOT}$. All short-term debt and $\frac{1}{T}$ fraction of long-term debt is rolled over each period, such that Reimbursements = Gross Issues = $D_{ST}^{TOT} + \frac{1}{T}D_{LT}^{TOT} = D^{TOT}(\frac{\beta+(1-\beta)T}{T})$

4.3 Optimal maturity structure in debt management theories

There are two theoretical starting points when thinking about optimal public debt maturity structure: rational expectations theory and Barro's Ricardian equivalence. Within both frameworks, the maturity of public debt does not matter. So it is useful to understand whether or not the assumptions behind these frameworks hold, and if they do not, what friction leads to the failure of either rational expectations or Ricardian equivalence and makes maturity of public debt an important subject of analysis.

(i) Based on the expectations hypothesis, maturity structure of government debt does not matter because there is perfect substitutability between instruments of different maturity. The intuition is that in complete markets with forward-looking agents, no-arbitrage condition ensures that any long-term bond is equivalent to a sequence of short-term bonds over the same time period and the interest rate of a long-term bond should be equal to the expected path of interest rates of a sequence of short-term bonds over the same time horizon. This assumes market completeness and rationality of individuals.

(ii) The classical Ricardian equivalence result states that the choice of deficit financing, i.e., the timing of taxes, has no impact on the real economy. Maturity structure is irrelevant in the standard New Keynesian models as a consequence of infinitely-lived representative agents in a complete market setting, where Ricardian equivalence holds. When Barro's Ricardian equivalence is extended to the choice of public debt instruments, this leads to the irrelevance theorem of debt management (Barro, 1979; Wallace, 1981; Missale, 1999). The key assumptions for this irrelevance to hold are the absence of distortionary taxes, the absence of financial frictions, and intergenerational altruism.

However, the setting in which debt management is practiced is likely such that (i) rational expectations theory does not apply because of imperfect substitutability between instruments, and (ii) Ricardian equivalence does not hold because taxes are distortionary. A non-zero term premium, for example, is classic evidence in favour of imperfect asset substitutability (Kane, 1981; Froot, 1989; Longstaff, 2000). Governments also typically levy proportional taxes, rather than lump-sum ones. As such, maturity structure matters for the real economy and asset prices, and therefore we care about this as investors, policymakers, taxpayers, and consumers. This section reviews what the formal theory says debt managers should be doing when selecting the maturity structure.

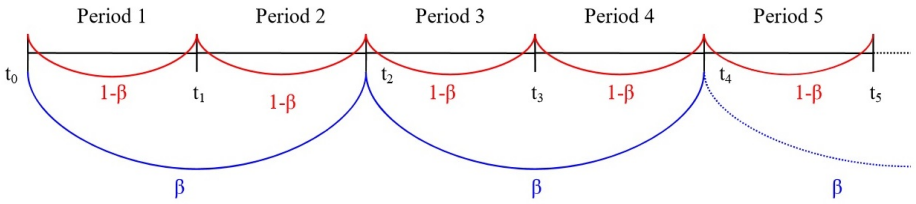
4.3.1 Portfolio theory

One potential prescription for debt management comes from portfolio theory. Also referred to as the "budgetary convenience" approach to debt management (Allen, 2013), this entails minimizing the cost of debt service subject to an acceptable level of rollover risk. This is the standard finance environment, where the agent (the debt manager) is risk averse and has mean-variance preferences. Following Dottori and Manna (2016), in case of a perpetually-optimizing debt manager, the prescription is to issue more shorter-term debt the steeper the slope of the yield curve. If debt manager optimizes only over the short horizon, when term premium is positive (and therefore long-term debt is relatively more expensive compared to rolling over short-term debt), portfolio theory prescribes shortening the maturity structure. A more formal discussion of this theory follows.

Debt manager's problem is to minimize the expected cost of servicing the debt, over the planning horizon, by setting an allocation between long-term and short-term debt, $\omega = [\beta, 1 - \beta]'$, where β is the share of long-term debt. The level of debt is assumed exogenous. Because the manager is risk averse and has mean-variance

preferences, optimal maturity structure minimizes both the interest paid on the debt as well as indirect costs arising from uncertainty about future interest rates. Cost of servicing the debt is a function of interest rates $z = [r, i]'$, where r is the long-term spot rate and i is the short-term spot rate. Risk-aversion parameter λ determines the costs arising from uncertainty. The optimal weight on long-term security (β^*) informs Propositions 1 and 2 below.

Figure 4.2: Issuance and maturity in a long-run model



Debt manager's problem is:

$$\min_{\beta} \sum_{t=0}^{N-1} \omega_t' z_t + \lambda(\omega_t' \Omega_t \omega_t)$$

$$\text{s.t.} \beta \in [0, 1],\ \lambda > 0$$

where

$$\begin{aligned}
\text{for all } t: &\quad \omega_t = [\beta, 1-\beta]' \\
\text{for } t = 0: &\quad z_t = [r, i]', \quad \Omega_t = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right) \\
\text{for } t = 1: &\quad z_t = [r, E(i)]', \quad \Omega_t = \left(\begin{smallmatrix} 0 & 0 \\ 0 & \sigma_i^2 \end{smallmatrix}\right) \\
\text{for } t > 2: &\quad z_t = [E(r), E(i)]', \quad \Omega_t = \left(\begin{smallmatrix} \sigma_r^2 & \sigma_{ir} \\ \sigma_{ir} & \sigma_i^2 \end{smallmatrix}\right)
\end{aligned}$$

When a debt manager only considers the current and the next period, the optimal weight on long-term security is the following (equation (6) in Dottori and Manna (2016)):

$$\beta^*_{T=2} = 1 - \frac{ts - \frac{1}{2}(E(i) - i)}{\lambda \sigma_i^2} \quad \text{where } ts = r - i \text{ is the slope of the yield curve}$$

Here cost minimization over a short planning horizon has two implications: (1) issue more long-term debt when short-term interest rates are expected to increase, so as to lock-in current interest rates; (2) issue more short-term debt the steeper the slope of the current yield curve. This is holding debt manager's risk parameter constant.

I extend this solution with the weak expectations hypothesis, which specifies the relationship between current interest rates and future interest rates by introducing the term premium[6]. A positive term premium means that long-term debt is relatively more expensive in expectation compared to rolling over short-term debt. I thus obtain a negative relationship between the term premium and share of long-term debt, which establishes Proposition 1.

$$\beta_2^* = 1 - \frac{\rho}{2\lambda\sigma_i^2} \quad \text{where } \rho \text{ is the term premium}$$

Proposition 1. For a mean-variance optimizing debt manager with a short planning horizon, when term premium is positive, cost minimization implies issuing less of the more expensive long-term security, holding debt manager's risk parameter and volatility of interest rates constant.

This proposition implies frequent rebalancing of the maturity structure. Hence it is most appropriate to test this in a short-run time-series setting. Assuming that the volatility of short-term interest rates is constant, Proposition 1 specifies $mat_t = 1 - c\rho_t + \epsilon_t$ where $c = \frac{1}{2\lambda\sigma_i^2} > 0$ is a constant term and where ϵ incorporates other factors omitted from this simplistic model. Equivalently, $mat_t = \mu + \gamma\rho_t + \epsilon_t$ can be

[6]The expectations hypothesis in two-period setting states that $(1+r)^2 = (1+i)(1+E(i))$, which following log transformation and up to first order approximation is equivalent to $2r = i + E(i)$. This illustrates that the return on long-term bonds should be equal to the return from rolling over short-term bonds. Without full equality, the weak expectations hypothesis here is that $2r = i + E(i) + \rho$, where ρ is the term premium.

tested with a reduced-form time series regression:

$$\Delta mat_t = \alpha_1^m \Delta mat_{t-1} + ... + \alpha_p^m \Delta mat_{t-p} + \alpha^{ecm} ecm_{t-1} + \mu + \gamma^\rho \rho_t + \gamma^d \left(\frac{D^{TOT}}{GDP} \right)_t + \gamma^\pi \pi_t + \epsilon_t$$

As shown earlier, maturity is by definition a persistent process. Therefore the time series specification regresses the changes in maturity structure on ecm, the disequilibrium in some relationship of maturity and other non-stationary determinants that explain it's unit root path. In line with the theory, volatility of short-term interest rates is assumed constant and positive. Under the null hypothesis, $\gamma^\rho = 0$, while under the alternative, $\gamma^\rho < 0$, and debt managers choose maturity structure based on short-term cost-risk optimization in line with portfolio theory. It is necessary to control for the level of debt and inflation because these are assumed constant in the theory.

Now consider an infinite planning horizon with $T = \infty$. Based on equation (5) in Dottori and Manna (2016), the optimal share of long-term debt for a perpetually cost-minimizing debt manager is the following:

$$\beta^*_{T=\infty} = \frac{\sigma_i^2 - \sigma_{ir}}{\Sigma} - \frac{E(ts)}{2\lambda\Sigma}, \qquad \text{where } \Sigma = \sigma_i^2 - 2\sigma_{ir} + \sigma_r^2 \text{ and}$$
$$E(ts) = E(r) - E(i) \text{ is the slope of}$$
$$\text{the long-run yield curve}$$

The economic interpretation of the two terms is the following. The first term is the net hedge against rollover risk, which is stated in relative terms since the co-movement of the two interest rates matters. The second term is the premium paid for this hedge. Equivalently, this is insurance against roll-over risk that debt manager implicitly purchases when issuing more long-term debt (Dottori and Manna, 2016).

Proposition 2. For a mean-variance perpetually optimizing debt manager, minimizing interest rate costs entails more shorter-term debt the

steeper the slope of the long-run yield curve, with an adjustment for debt manager's risk parameter, volatility of interest rates, and covariance between the short- and the long-term rate.

Proposition 2 is a steady-state statement that is directly testable with panel regressions. Assuming that $\Sigma, \sigma_i^2, \sigma_{ir}$ are constant over time but vary across countries, this specifies $mat_{j,t} = \tilde{c_j} - c_j E(ts)_{j,t} + \epsilon_{j,t}$. The constant term is $\tilde{c_j} = \frac{\sigma_{i,j}^2 - \sigma_{ir,j}}{\Sigma_j}$, the coefficient is $c_j = \frac{1}{2\lambda_j \Sigma_j}$, and $\epsilon_{j,t}$ incorporates other factors omitted from this simplistic model. Since $\tilde{c_j} = c_j \left(2\lambda_j(\sigma_{i,j}^2 - \sigma_{ir,j})\right)$, the specification is $mat_{j,t} = c_j \left(2\lambda_j(\sigma_{i,j}^2 - \sigma_{ir,j})\right) - c_j E(ts)_{j,t} + \epsilon_{j,t}$. This can be tested with a reduced-form panel regression of maturity on the time-invariant term c_j that captures risk preferences and interest rate variance and covariance, and on the interaction between long-run term spread and this term c_j. With j countries and t periods, the reduced-form panel regression is:

$$mat_{j,t} = \alpha mat_{j,t-1} + \gamma^c c_j + \gamma^{cts} c_j \bar{Ets}_{j,t} + \gamma^d \left(\frac{D^{TOT}}{GDP}\right)_{j,t} + \gamma^\pi \pi_{j,t} + \epsilon_{j,t}$$

where $c_j = \frac{1}{2\lambda_j \Sigma_j}$, $\Sigma_j = \sigma_{i,j}^2 - 2\sigma_{ir,j} + \sigma_{r,j}^2$, and $\bar{Ets}_{j,t}$ is the expected long-run term spread. The test of portfolio theory in this setting concerns the coefficient on the interaction term, γ^{cts}. Under the alternative, $\gamma^{cts} = -1$ i.e., debt managers choose maturity structure based on long-term cost-risk optimization in line with portfolio theory. The lag term is included to deal with persistence properties of the maturity variable. It is necessary to control for debt and inflation as before.

4.3.2 Tax smoothing and fiscal insurance theory

The optimization problem laid out in the previous section considered a partial equilibrium setting. An alternative strand of theories considers maturity structure in general equilibrium. When taxes are distortionary, and therefore when Ricardian equivalence does not hold, there is a theoretical welfare argument for policymakers to prevent high and variable tax rates. There is scope for the choice of maturity structure to

minimize the risk of future sharp changes in tax rates or government expenditures coming from uncertain future events. Hence, in this environment, the composition of debt is a tool for the government to minimize future budget volatility and to smooth taxes and consumption across states and over time. A debt manager that wishes to pursue tax smoothing will set maturity structure so as to hedge against sources of uncertainty in output and government spending. A review of the formal model follows.

The model is due to Bohn (1990), where identical infinitely-lived individuals maximize present discounted value of consumption subject to a budget constraint that features distortionary taxation on endowments Y and K assets:

$$\begin{aligned} \max_{c_t} \quad & E_t \sum_{j \geq 0} \rho^j c_{t+j} \\ \text{s.t.} \quad & c_t + \sum_k p_{t,k} A_{t,k} = Y[1 - \tau_t - h(\tau_t)] + \sum_k (p_{t,k} + f_{t,k}) A_{t-1,k} \end{aligned}$$

Hence the central planner, the government, maximizes the welfare of a representative agent subject to its own budget constraint:

$$\begin{aligned} \max_{\tau_t} \quad & E_t \sum_{j \geq 0} \rho^j Y_{t+j}[1 - h(\tau_{t+j})] \\ \text{s.t.} \quad & T_t = \tau_t Y_t = G + \sum_k (p_{t,k} + f_{t,k}) D_{t-1,k} - \sum_k p_{t,k} D_{t,k} \end{aligned}$$

As shown by Bohn (1990), the first order conditions of this problem combine into:

$$E_t(\Delta\tau_{t+1}, \hat{r}_{t+1,k}) = E_t(\hat{\tau}_t, \hat{r}_{t+1,k}) = Cov_t(\tau_{t+1}, r_{t+1,k}) = 0$$

where $\hat{\tau}_t = \tau_{t+1} - E_t(\tau_{t+1})$ is the tax rate innovation. For convenience, the first order condition is specified using excess returns $\hat{r}_{t+1,k}$ – over the risk-free rate corresponding to $k = 0$. This first order condition says that stabilizing taxes across possible states of

nature implies zero conditional covariance between taxes and returns on all available securities.

In combination with an additional link between innovations in debt and tax rates[7], Bohn (1990) shows that these first order conditions yield a system of K equations for government securities, $d_{t,k}$:

$$\sum_{l} Cov_t(\hat{r}_{t+1,l}, \hat{r}_{t+1,k}) d_{t,l} + Cov_t(\hat{r}_{t+1,k}, \sum_{j\geq 0} \rho^j \hat{g}_{t+1+j}) - w_t Cov_t(\hat{r}_{t+1,k}, \sum_{j\geq 0} \psi^j \hat{y}_{t+1+j}) = 0 \quad \forall \quad l$$

where w_t is a weighting factor, $\hat{r}$ is a real return in excess of a benchmark risk-free one-period security, $\hat{y}$ is the innovation in growth rate of output, $\hat{g}$ is the innovation in government spending relative to output, and d_k is the ratio of security k debt to output. Once d_t is represented as a vector of government securities, Σ_r as a variance-covariance matrix of excess returns, assumed to be non-singular, $\Sigma_{g,r}$ and $\Sigma_{y,r}$ as a vector of covariances between returns and the present-value expressions $\sum_{j\geq 0} \psi^j \hat{y}_{t+1+j}$ and $\sum_{j\geq 0} \rho^j \hat{g}_{t+1+j}$, respectively, the system of K equations simplifies to:

$$\Sigma_r d_t + \Sigma_{\hat{g},r} - w_t \Sigma_{\hat{y},r} = 0$$

Proposition 3. For a debt manager that maximizes welfare of a representative agent, the optimal debt structure is a function of covariances of returns with innovations in output and government spending. The optimal debt structure is chosen to hedge against sources of uncertainty in output and government spending and thereby minimizes fluctuations in tax rates. (Bohn 1990, p.1224).

Following Bohn (1990), I consider that optimal debt structure is quantitatively proportional only to covariances of returns with innovations in output and leave out

[7] This is $\hat{\tau}_{t+1} = \tau_{t+1} - E_t \tau_{t+1} = (1 - \psi) \exp(-\bar{y})(\sum_k \hat{r}_{t+1,k} d_{t,k} + \sum_{j\geq 0} \rho^j \hat{g}_{t+1+j}) - \tau_t \sum_{j\geq 0} \psi^j \hat{y}_{t+1+j}$ where $\bar{y}$ is mean output growth rate.

the covariances of returns with innovations in government spending.[8] Therefore,

$$d_t = w_t \Sigma_r^{-1} \Sigma_{\hat{y},r} = w_t s$$

where the vector $s = \Sigma_r^{-1} \Sigma_{\hat{y},r}$ indicates whether securities enter with positive or negative sign into the optimal portfolio and in which relative quantities (Bohn, 1990, p.1225).

If besides the risk-free $k = 0$ security there are only long-term and short-term government securities, then $d^{LT} = w_t s_t^{LT}$ and $d^{ST} = w_t s^{ST}$ and optimal share of long-term bonds is the following β^{TS}. This is a steady-state optimal share of long-term bonds for a welfare-maximizing and tax-smoothing debt manager, to the extent that covariances of returns with innovations in output are constant over time[9].

$$\beta^{TS} = \frac{s^{LT}}{s^{LT} + s^{ST}}$$

I compute s^{LT} and s^{ST} following Bohn's methodology[10] and accommodate the fact that optimal security s may be positive or negative, as described in the Appendix. I test Proposition 3 in panel regressions of maturity structure on the time-invariant β^{TS}, with j countries and t periods:

$$mat_{j,t} = \alpha mat_{j,t-1} + \gamma^{ts} \beta_j^{TS} + \gamma^d \left(\frac{D^{TOT}}{GDP} \right)_{j,t} + \gamma^{\pi} \pi_{j,t} + \epsilon_{j,t}$$

Under the null hypothesis, $\gamma^{ts} = 0$, while under the alternative, $\gamma^{ts} > 0$. A positive γ^s indicates that countries with greater optimal weight on long-term debt based on

[8] When extending this theory to the data, Bohn (1990) finds correlations of returns with output as empirically significantly interesting, but correlations with government spending as small and insignificant.

[9] This is also assumed in Bohn (1990).

[10] Bohn (1990) uses bi-variate VARs of GNP growth rate and excess returns to extract point estimates of the elements of $\Sigma_{\hat{y},r}$ and s, as functions of VAR coefficients and the residual variance-covariance matrix. See Appendix for further details.

tax smoothing considerations have in fact a longer maturity structure.

4.3.3 Clientele and liquidity preference theories

The theory of fiscal insurance, presented in the previous section, relies on distortionary taxation to provide scope for maturity structure to matter for tax and consumption profiles. An alternative approach relies on some financial friction that breaks the rational expectations theory. This leads to imperfect substitutability across the yield curve. Non-zero term premium is classic evidence against the expectations hypothesis and perfect asset substitutability.

Imperfect substitutability could be in place if short-term bonds are preferred due to their liquidity and cash-like properties, an assumption at the heart of liquidity preference theory. This theory prescribes that when market preferences shift to more liquid securities in times of high uncertainty or heightened sovereign risk, the debt manager should and will be constrained to issue shorter-maturity debt.

Imperfect substitutability can also arise with the existence of preferred habitats and clienteles (Modigliani and Sutch, 1967). An environment with multiple generations—the young that save, the old that are retirees—produces demographic clienteles in the long-term government bond market. We can think of institutional investors, particularly pension funds and insurance corporations, as clienteles in the long-term government bond market (Modigliani and Sutch, 1966). For example, shifting demographics and growing portfolios of pension funds are a source of demand for long-dated government bonds and have been attributed as the cause of lengthening maturity structure in advanced economies (Abbas et al., 2014).

An overlapping generations model by Guibaud, Nosbusch, and Vayanos (2013) prescribes that the sovereign debt manager should cater to such demographic clienteles, yet imperfectly. When facing an increase in the share of savers in the population and a corresponding downward pressure on the long-term interest rates, the opti-

mal response is to increase the supply of long-term bonds. Yet it's optimal to cater imperfectly and let the long-term interest rate fall due to demand pressure. This is because a welfare-maximizing debt manager internalizes the fact that catering to a growing long-horizon clientele imposes more risk on future generations through a more uncertain tax burden. Such welfare-maximizing debt manager therefore allows the long-term interest rate to decrease. A cost-minimizing debt manager, on the other hand, would cater to increased demand fully, and there would be no change in the long-term interest rate. This section discusses these clienteles and liquidity preference theories in greater detail and how to operationalize the theories into empirical analysis.

The clientele-based model of the yield curve proposed by Guibaud, Nosbusch, and Vayanos (2013) provides convenient reduced-form relationship between optimal maturity structure and clienteles that are based on demographic features of the economy.[11] In this three-period model, clienteles are represented with an overlapping-generations economy that features the young generation, who receive an endowment α and pay taxes on it, alongside middle-aged generation and old generation, who consume their endowment. Agents trade when young and middle-aged, and have CRRA utility with the same risk aversion parameter γ. The government issues non-contingent bonds and collects income taxes, with the following budget constraints:

$$\begin{aligned}
&\frac{b_1}{R} + \frac{B_1}{L_1^2} = 0 && \text{in period t=1} \\
&\alpha_2\tau_2 + \frac{b_2 - B_1}{R_2} = b_1 && \text{in period t=2} \\
&\alpha_t\tau_t + \frac{b_t}{R} = b_{t-1} && \text{in period t>2}
\end{aligned}$$

Here R_t is the gross one-period interest rate and L_t is the gross two-period interest

[11]Besides the model proposed by Guibaud et al. (2013), Vayanos and Vila (2009), Greenwood and Vayanos (2010),Greenwood and Vayanos (2014) have also proposed theoretical models featuring clienteles and preferred habitats within a micro-founded partial equilibrium setting.

rate, while b_t is a one-period bond and B_t is a two-period bond issued in period t.

Guibaud, Nosbusch, and Vayanos (2013) show that in equilibrium for the first period of the model, the following characterizations at the heart of classical clientele theory hold (propositions 2 and 3, p.1921-1922):

i An increase in the size of the young clientele (with a longer investment horizon) in period 1 lowers the two-period interest rate L_1, if $\gamma > 1$ (agents are more risk-averse than log).

ii More long-term bonds B_1 leads to higher two-period interest rate L_1. Intuitively, young and middle-aged generations absorb larger exposure to the risk that one-period interest rate will increase in the subsequent period and therefore demand a higher return.

Thus when debt manager maximizes welfare (consumption) in this overlapping generations environment, there is an optimal maturity structure that induces efficient risk-sharing across generations:

Proposition 4. A welfare-maximizing debt manager should raise the optimal supply B_1^* of two-period bonds in response to an increase in the long-horizon clientele. It is optimal to not fully cater to increased demand, and this will lead to a higher relative price and a lower equilibrium two-period interest rate that prevails when two-period bonds are in supply B_1^*. (Guibaud, Nosbusch and Vayanos, 2013, proposition 8, p.1930)

As noted by the authors, this is in stark contrast to the cost-minimizing debt manager problem examined in portfolio theory. A cost-minimizing debt manager in this model would face a lower term premium as a result of an increase in the long-horizon clientele. So the debt manager would increase the supply up to the point where the two-period interest rate remains unchanged, fully capitalizing on the term premium. Yet a welfare-maximizing government internalizes that catering to long-horizon clientele imposes more risk on future generations through a more uncertain tax burden, and therefore allows the long-term interest rate decrease.

An increase in the long-horizon clientele can naturally be represented by demographic changes, which are of low frequency. Therefore this proposition is more about constant optimal maturity rather than about time-varying maturity policy. In fact, Guibaud et al. (2013) confirm these predictions with the following two separate panel regressions (p. 1936):

$$TS_{j,t} = a + bDem_{j,t} + u_j + e_{j,t}$$
$$WARM_{j,t} = c + dDem_{j,t} + v_j + f_{j,t}$$

for j countries and t time periods. In this specification, TS is the slope of the yield curve, $WARM$ is the weighted average maturity of government debt. Dem is a demographic variable measuring the relative size of the long- and short-horizon clienteles.[12]

Similarly to Guibaud et al. (2013), I capture the relative size of the long- and short-horizon clienteles with the share of workers or prime savers in the population, ws and ps:

$$mat_{j,t} = \alpha mat_{j,t-1} + \gamma^{dem}(ws_{j,t} \text{ or } ps_{j,t}) + \epsilon_{j,t}$$

The alternative hypothesis of $\gamma^{dem} > 0$ is consistent with a welfare-maximizing or cost-minimizing debt manager that caters to demographic clienteles. Another question is whether debt manager fully caters to demographic clienteles or only partially. It is difficult to implement in a reduced-form panel setting. In the time series setting, however, I can exploit the unit root properties of the variables and consider whether maturity, the share of prime savers in the population and term spread are cointegrated and whether the equilibrium-correcting response to demographic changes is in line

[12]Guibaud et al. (2013) use median age to measure the relative size of the long-horizon clientele. Higher median age indicates fewer young savers and hence the relative decline of the long-horizon clientele. Therefore they tested for a positive b and negative d, which under the null hypothesis would support Proposition 5.

with theory. Consider the equilibrium correction mechanism:

$$ecm = mat + \gamma^{dem}(ws \text{ or } ps) + \gamma^{ts} ts = 0$$

An alternative hypothesis of $\gamma^{dem} < 0, \gamma^{ts} < 0$ would be in line with the clienteles theory of Guibaud et al. (2013). This alternative hypothesis implies both a negative relationship between young-horizon clientele and term spread, and a positive relationship between young-horizon clientele and maturity structure, if changes in the young-horizon clientele are weakly exogenous in the econometric sense.

The clientele demand can also be captured by looking at the assets of pension funds and insurance corporations. Pension funds and insurance corporations hold a significant fraction of their portfolio in long-term assets in order to match the long maturity structure of their liabilities. Therefore, growth in the assets of the contractual savings sector represented by pension funds and insurance corporations, $\Delta ln(csa)$ also proxies for increased demand for long-term bonds. This proxy is also used in Goudswaard (1990).

The notions of demand can also be extended to market preferences for liquidity. When market preferences shift to more liquid securities in times of high uncertainty or heightened sovereign risk, neutrality approach to debt management would dictate a shortening of maturities, as an application of liquidity preference theory (Goudswaard, 1990). Shifts in market preferences due to either demographics or uncertainties can be examined in both panel and time-series settings, with the following regressions:

$$mat_{j,t} = \alpha mat_{j,t-1} + \gamma^{dem}(ws_{j,t} \text{ or } ps_{j,t}) + \gamma^{csa}\Delta ln(csa)_{j,t} + \gamma^{p} pmat_{j,t} + \gamma^{r} risk_{j,t} + \epsilon_{j,t}$$

In this specification, the alternative hypothesis of $\gamma^{csa} > 0$ is consistent with welfare-maximizing or cost-minimizing debt manager that caters to demographic clienteles. On the other hand, an increase in yield relative to the risk-free bench-

mark, *risk*, is indicative of lower creditworthiness, which leads to a shift in market preferences towards more liquid securities (short-term bonds). The alternative hypothesis with $\gamma^r < 0$ is consistent with a debt manager that accommodates market preferences for liquidity.

Following Goudswaard (1990), I also include the maturity of private sector bond issuances in the regression, *pmat*. The maturity of private sector bond issuances serves as a control and proxy for market preferences concerning maturity in general, under the assumption that individual corporate bond issuers are small and therefore maturity-takers, responding to overall market demand.

4.3.4 Inflation credibility

Besides portfolio theory, fiscal insurance theory, liquidity preference theory and clienteles theory, last but not least this section reviews the role that inflation plays in optimal government debt management. Debt issuance is fraught with the time-inconsistency problem since debt holders face the risk of surprise inflation and the possibility that debt will be inflated away. Most of the debt issued is in fact nominal rather than inflation-indexed. Therefore, Missale and Blanchard (1994) consider the joint problem of debt management with respect to maturity and monetary policy that keeps zero inflation credible. They note that maturity structure should be set as a decreasing function of real debt service burden in order to keep the zero inflation stance credible.

Formally, first assume that taxes are set to keep the level of debt constant (in the absence of unexpected inflation), such that $T = r\,(1 - m(\pi - E\pi))$. Also, assume that government spending is exogenous. With these two assumptions, their simple model specifies real debt burden in terms of maturity and unexpected inflation:

$$D' = (1 - m(\pi - E\pi))D$$

where π is inflation and m, the maturity structure, is an index between zero and one, and represents a simplification of the average maturity concept, $WARM$. This maturity measure captures the negative effect that unexpected rate of inflation has on the real value of debt, D'. This is the well-known time-inconsistency problem of a debt manager who can inflate away debt with unexpected inflation.

Missale and Blanchard (1994) solve for conditions under which the government will prefer not to inflate away the debt, assuming that debt manager minimizes loss arising from unexpected inflation, $L = \frac{1}{2}\pi^2 - \theta_1(\pi - E\pi) + \theta_2 T$. For the value of the loss from not cheating (commitment to inflation) to be no greater than the value of cheating (inflating away), the maximum maturity that is consistent with zero inflation has to be m^*:

$$m^* = \frac{\theta_1(\sqrt{\delta} - \delta)}{\theta_2(1+\delta)rD}$$

where δ is the rate the government discounts the future by, θ_1 is the loss function parameter than reflects benefits of unexpected inflation, and θ_2 is the loss function parameter that reflects the cost of taxation, while r is the interest rate (Missale and Blanchard (1994), equation (9)). This established Proposition 5.[13]

Proposition 5. In order to keep zero inflation stance credible, the maturity of debt will be a decreasing function of the debt service burden. (Blanchard and Missale (1994), p.316)

This theory of inflation and Central Bank credibility specifies that $mat = \psi(rD)^{-1} + \epsilon$, where $\psi = \frac{\theta_1(\sqrt{\delta}-\delta)}{\theta_2(1+\delta)}$ is constant over time, and where r captures the effective interest rate paid on debt, and ϵ captures all the factors that are not represented in this simplistic model. Assuming that ψ, which captures preferences of debt manager as far as the discount rates and optimal weights in the loss function, is also the same

[13]In order to extend this from maximum maturity to actual maturity chosen, Missale and Blanchard (1994) informally argue that governments prefer long-term nominal debt to short-term nominal debt because of fiscal insurance properties. Therefore, the longest feasible maturity will be chosen over the zero maturity corner solution.

among advanced economies, then the reduced-form regression of Proposition 6 is:

$$mat_{j,t} = \alpha mat_{j,t-1} + \gamma^{rd}\big(r\frac{D^{TOT}}{GDP}\big)^{-1}_{j,t} + \gamma^{\pi}\pi_{j,t} + \epsilon_{j,t}$$

This can be examined in both cross-sectional and time-series settings. If data on the effective interest rate r is missing, then it can be calculated from the short- and long-term interest rates weighed by the respective short- and long-term debt shares. Under the alternative hypothesis of debt manager concerned with maintaining an anti-inflationary stance, $\gamma^{rd} > 0$.

4.4 Empirical model of maturity structure

The previous section described formal theories of optimal maturity structure and their reduced-form representations. A summary of testable theories is outlined in Table 4.1, while the variables are explained in Table 4.2. The reduced-form representations of theories present the determinants of maturity, *mat*. The functional form of the empirical analysis is therefore $mat_t = f(mat_{t-1}, x_t)$, where *mat* is the measure of maturity and x is a vector of determinants.

Most theoretical specifications use the share of long-term debt, broadly defined, as a simple modeling device to capture the maturity structure. In the empirical analysis that follows, I use the weighted average residual maturity measure, $WARM$, because it is a monotonic function of the share of long-term debt for a fixed T period long-term bond[14]. This weighted average measure is also more granular and allows to differentiate between countries that issue long-term debt of five-year maturity and twenty-year maturity. It is therefore an appropriate measure for the empirical analysis of cross-sectional differences and time-series dynamics of maturity structure.

[14]Recall that $WARM = (1-\beta) + (\frac{1+T}{2})\beta$ where β is the share of long-term debt and T is the number of periods to repay it.

I consider a single regression equation that includes all determinants that come from formal models. The determinants that are statistically significant and with the right sign tell us which debt management strategies might be followed by debt managers. This single-equation approach is preferred to the alternative of estimating each theory as a separate regression.[15]

4.4.1 Data and operationalized variables

For panel analysis, the data primarily comes from the OECD databases, which is the only source of cross-country data on weighted average remaining maturity. In fact, the OECD stopped providing the maturity data after 2010, hence the cross-sectional analysis in this study does not include the more recent years. For time series analysis of the United States data, I constructed the maturity structure from CRSP bond-level issuance data. The rest of the U.S. time series data primarily comes from FRED. See Table 4.2 for exact definitions, sources, and descriptive statistics of all variables.

[15] An alternative is to treat different theories as separate equations and compare the fraction of variation in the maturity structure that is explained by individual theories, by looking at the adjusted R^2 of the models. However, this method is only valid for OLS type estimation. As I argue later, for panel estimation, the most appropriate specification for these regressions is the generalized method of moments (GMM) estimation because of the lagged dependent variable. Yet in the GMM context, the adjusted R^2 is no longer appropriate for model comparison. Therefore, I present results from single-equation model.

Table 4.1: Predictions from debt management theories

Theory	Type	Action	in Response to	Controlling for	Coefficient sign
Portfolio theory with cost-risk portfolio optimization	Formal, with short horizon	$\downarrow mat$	$\uparrow$ term premium ρ	debt level d and inflation π	< 0
	Formal, with long horizon	$\downarrow mat$	$\uparrow$ interaction term capturing long-run slope of real yield curve and the covariance between short- and long-term interest rates ${}^{ts}/_{\Sigma}$	debt level d and inflation π	< 0
Welfare-maximization with tax smoothing and fiscal insurance	Formal	$\uparrow mat$	$\uparrow$ term capturing covariance of returns with innovations in output, β^{TS}		> 0
Welfare maximization in presence of clienteles	Formal	$\uparrow mat$, $\downarrow ts$	$\uparrow$ in long-horizon clientele, captured by share of savers in the population, ws, ps		$> 0, < 0$
	Informal	$\uparrow mat$	$\uparrow$ growth rate of contractual savings sector (total assets of pension funds and insurance corporations), $\Delta ln(csa)$		> 0
Neutrality approach to debt management by adjusting to market preferences	Informal	$\downarrow mat$	$\uparrow$ spread over the risk-free benchmark $risk$	original maturity of private sector bonds $pmat$	< 0
Minimizing loss from unexpected inflation with maintaining anti-inflationary stance	Formal	$\uparrow mat$	$\uparrow$ in inverse debt burden level $(rdebt)^{-1}$	inflation π	> 0

Table 4.2: Operationalized Variables

Variable	Label in Tables	Definition	Analysis	Notes	Source
$WARM$	GovtMaturity	Weighted average residual maturity of central government debt, in years	OECD panel	For most countries, coverage is at the marketable debt level, including debt held by the Central Bank. There is no data on original maturity basis, and therefore I use residual maturity basis.	OECD Central Government Debt Database
			U.S. time series	Marketable debt, including intragovernmental debt.	own calculation using CRSP (Centre for Research on Security Prices) data
π	Inflation	Inflation, in percent	OECD panel	Growth rate of the GDP Deflator, in percent	OECD National Accounts Database
			U.S. time series	Inflation (growth rate in seasonally adjusted GDP deflator), in percent, annualized	Federal Reserve Economic data (FRED)
i		Real 3-month interbank interest rate, in percent	OECD panel	Nominal interest rate less inflation	OECD National Accounts Database
			U.S. time series	3-Month Treasury Bill secondary market rate, in percent, annuazlized	Federal Reserve Economic data (FRED)
r		Real interest rate on a 10-year government bond, in percent	OECD panel	Nominal interest rate less inflation	OECD National Accounts Database
			U.S. time series	10-Year Treasury constant maturity rate, in percent, annualized	Federal Reserve Economic data (FRED)
$\bar{ts}$	LRTermSpread	Long-run real term spread, in percent, annual	OECD panel	Calculated as the difference between the interest rate on 10-year government bond (r) and the 3-month interbank interest rate (i), averaged over the last 5 years	own calculation using OECD National Accounts Database
ts	TermSpread	Real term spread, in percent, quarterly	U.S. time series	Calculated as the difference between the interest rate on 10-year government bond (r) and the 3-month interbank interest rate (i)	own calculation using FRED
σ_i^2		Variance of the 3-month interbank interest rate	OECD panel		own calculation using OECD National Accounts Database
σ_r^2		Variance of the interest rate on a 10-year government bond	OECD panel		own calculation using OECD National Accounts Database
$\sigma_{i,r}$		Covariance between real short and long rates	OECD panel		own calculation using OECD National Accounts Database
Σ		Interest comovement term	OECD panel	$\sigma_i^2 - 2\sigma_{i,r} + \sigma_r^2$	own calculation
c	c	Interest comovement interaction term	OECD panel	$\frac{1}{2\Sigma}$	own calculation using OECD National Accounts Database
$E(i_{t+1\|t})$		Expected 2-year ahead 3-month interbank interest rate, in percent, annualized	OECD panel	Forecast published in June, made in the beginning of the year	own calculation using forecasts in OECD Economic Outlook
$E(r_{t+1\|t})$		Expected 2-year ahead interest rate on a 10-year government bond, in percent, annualized	OECD panel	Forecast published in June, made in the beginning of the year	own calculation using forecasts in OECD Economic Outlook

$E\bar{t}s$	ETermSpread	Expected long-run term spread, in percent, annualized	OECD panel	Calculated as the difference between the expected interest rate on 10-year government bond $E(r_{t+1})$ and the 3-month interbank interest rate $E(i_{t+1})$	own calculation using forecasts in OECD Economic Outlook
ρ	TermPremium	Return on the long-term bond in excess of rolling over short-term bonds over the same horizon, in percent	U.S. time series		Adrian et al., 2013
β^{TS}	TaxSmoothingLT	Optimal share of long-term debt under the tax smoothing approach	OECD panel	See separate note on calculation methodology	own calculation following Bohn (1990)
ws	WorkingSavers	Share of working-age population, in percent	OECD panel	Ages 15-64.	World Development Indicators
ps	PrimeSavers	Share of prime savers in the population, in percent.	U.S. time series	Ages 35-54.	own calculation using U.S. Census Bureau data
d	Debt	Share of marketable debt, in percent of GDP	OECD panel		OECD Central Government Debt Database
			U.S. time series	Share of marketable (publicly-held) Federal debt, in percent of GDP	Federal Reserve Economic data (FRED)
rd	DebtService	Effective debt service, in percent of GDP	OECD panel	Debt multiplied by real effective interest rate. Real effective interest rate is the weighted average of long- and short-term rates r and i, with shares of long- and short-term debt as the weights	own calculation
			U.S. time series	Debt multiplied by real five-year Treasury constant maturity rate	Federal Reserve Economic data (FRED)
rd^{-1}	InflationCredibility	Inverse of the real debt service	OECD panel		own calculation
$risk$	RiskSpread	Long-term interest rate spread over the risk-free benchmark.	OECD panel	German long-term interest rate is used as a benchmark for EU countries as well as Turkey and Israel. U.S. long-term interest rate is used as a benchmark for the remaining countries: Australia, Canada, Chile, Japan, Mexico, New Zealand, South Korea. In nominal terms.	OECD National Accounts Database
csa	ContractSaveAssets	Growth rate of total assets of contractual saving sector, in percent of GDP	OECD panel, U.S. time series	Contractual savings sector covers insurance corporations and pension funds, based on data availability. Assets include: currency and deposits, shares and other equity, security other than shares (with the exception of financial derivatives), as well as loans and other assets.	own calculation using OECD Institutional Investor Database
$pmat$	CorpMaturity	Weighted average original maturity of private sector debt issued in the domestic market by AAA-rated issuers, in years.	OECD panel, U.S. time series	Aggregated from at least 10 bond issues in a year. Reported as missing for 9 or fewer bond issuances. Maturity of bonds issues in the Euromarket or similar category used for select countries where domestic bond market is too small[16]	own calculation using Thomson Reuters Bonds Database

[16]Czech Republic, Slovak Republic, Estonia, Turkey, Iceland, Hungary, Greece, Israel, Poland, Slovenia use the Euromarket data

Term premium ρ data for U.S. is from Adrian, Crump, and Moench (2013), using the five-factor, no-arbitrage term structure model. This captures interest rate expectations based on market measures. In contrast to market expectations for the term premium, expected long-term spread $E\bar{ts}$ is calculated from national forecasts. It is defined as the difference between the interest rate on 10-year government bond and the 3-month interbank interest rate. I use interest rate forecasts published in the June edition of the OECD Economic Outlook each year, for $Er_{t+1|t}$ and $Ei_{t+1|t}$. This represents the data available in the decision-makers information set and conceptually amounts to a two-year ahead forecast horizon because it captures the expected change over the current year and the next year. As an alternative measure, I also consider historical spread, an average over the last 5 years. Given the cyclicality of interest rates with business cycles in a setting with an independent central bank, the historical term spread should be negatively correlated with the expected future term spread.

The calculation of optimal share of long-term debt determinant under the tax smoothing approach is described in the Appendix. I also provide the summary statistics for each country in the sample in Table B.2.

To capture demographic aspects that represent long-horizon clientele as in the formal model, I use the share of prime savers in the population (defined as age 35-54 years old) in the U.S. time series setting, and the share of working age population in panel setting[17]. I also consider the assets of the contractual savings sector, which includes pension funds and insurance corporations, in percent of GDP, which is a proxy for demand for long-term maturities stemming from domestic institutional investors. Based on the time series properties of the contractual savings sector series, I apply the growth rate transformation, since it is the only variable smoothly trending upwards in all countries in the sample, and this is the only variable for which this

[17]The share of prime savers is the preferred measure, but it is not readily available for all countries in the panel.

transformation is applicable.

I use the weighted average original maturity of private sector bond issuances as a proxy for maturity demanded by the financial sector. This variable is constructed from micro-level data of private sector bond issuances from the Thomson Reuters Bonds Database.

These are not the only variables that could capture demand for specific or all maturity segments of government bonds. Sovereign wealth funds of China and the Gulf countries likely play a role as buyers of government bonds, particularly for advanced economies like the U.S. The data on sovereign wealth funds is scarce, and was not available both in time series perspective and appropriately disaggregated by country. Additionally, home bias due to a national savings bank might be a significant determinant of demand for certain maturity of government bonds, particularly in Japan. This data was also not readily available as a time series. Therefore, these are the variables that might cause omitted variables bias if these factors are time-variant.

4.5 Cross-sectional differences in advanced economies

This section presents the results of empirical analysis that considers cross-sectional differences in the weighted average residual maturity among advanced economies. The set of explanatory variables corresponds to the determinants derived from reduced-form expressions of optimal maturity that arise in different theoretical settings that describe the steady-state of an economy.

4.5.1 Model and methodology

Based on structural equations presented earlier in Section 2, maturity measured by weighted average remaining maturity $WARM_{j,t}$ is expected to be persistent and hence it is very important to include a lagged dependent variable $WARM_{j,t-1}$. It is also

appropriate to consider unobserved time-invariant individual effects η_j, because this is country-level data, and these could be unobserved country-specific effects that are significant determinants of the maturity structure. Therefore this is a dynamic panel setting, and the model specification is as follows:

$$WARM_{j,t} = \alpha WARM_{j,t-1} + \gamma x_{j,t} + \mu + (\eta_j + v_{j,t})$$

for $t = [1, ..., 11]$ period, j countries, with $\gamma = [\gamma_1, ..., \gamma_9]$ a 1×9 vector and $x_t = [c_j; c_j E\bar{t}s_t; \beta_t^{TS}; ws_t; csa_t; pmat_t; risk_t; (r_t d_t)^{-1}; \pi_t;]'$ a 9×1 vector. As previously defined, variables $c_j = \frac{1}{2\lambda\Sigma_j}$ and $\Sigma_j = \sigma_{i,j}^2 - 2\sigma_{ir,j} + \sigma_{r,j}^2$ capture the determinants of maturity structure based on portfolio theory, and naturally are a function of the covariance of asset returns, the rates on the short-term and long-term bonds.

First, note that in this setting with a persistent dependent variable and time-invariant individual effects, the standard estimators like OLS and FE (Fixed Effects) are not consistent. One way to get around this issues is by using the GMM (Generalized Method of Moments) approach, which is what I do in this paper. Second, note that the failure of OLS and FE in a dynamic panel setting leads to a useful property, which states that the consistent estimate of the coefficient on the lagged dependent variable is bounded by it's FE and OLS estimates: $\hat{\alpha}_{FE} < \alpha < \hat{\alpha}_{OLS}$ for $\hat{\alpha}$. I use this property later for model selection.

With this data, the dimensions of the panel play a critical role. From the original unbalanced data set $N = 34, T = 31$, I use the subset $N = 31, T = 11 = [2000 : 2010]$ where the data is available for all variables and all countries, and the time dimension is not too large relative to the small number of countries. With the GMM approach, the number of instruments is quadratic in T, and therefore the problem of too many instruments would be pervasive if the data for all years were used, for 1980-2010. By contrast, in the time series analysis that follows, the time series dimension of the U.S.

data spans more than forty years of quarterly data, for the period 1971:Q2-2013:Q3.

Taking into account possible year effects, OLS and FE models are presented in Table 4.3, models (1)-(4). Based on the property described earlier, a consistent estimate for the lagged dependent variable should be $0.6 < \alpha < 0.9$. This suggests that true $\alpha \to 1$, i.e., maturity structure may follow a random walk on average in the sample of countries. Therefore I use both the First Difference Generalized Method of Moments (GMM) model, and the System GMM model, which is useful when true $\alpha \to 1$. Following a first-difference transformation that eliminates the individual effects, these methods employ a set of moment conditions and find parameter estimates that come as close as possible to achieving these orthogonality properties in the sample. The details are described in the Appendix.

In very large samples, the GMM approach and testable conformity to set assumptions used would imply causality. Given the low sample size and expected low power of tests, however, the interpretation of causality does not fully apply in this model. GMM estimators are also consistent and asymptotically Normal for an appropriate choice of a weight matrix. I present both two-step and one-step GMM estimators for robustness, which differ in the choice of the weight matrix. Since the sample is very small and far from the asymptotic dimension, the results should be interpreted with some caution.

In an econometric environment where I have a low number of countries, I need to avoid the problem of too many instruments in order to obtain a well-specified GMM model. One approach towards a well-specified GMM model is to use a subsample with a shorter time dimension, which is what I do by considering the period 2000-2010. Additionally, I also use as few instruments as possible, but ensure that the few instruments used are also as restrictive and therefore as powerful as possible. For the First Difference GMM estimation, I use one lag depth of the dependent variable in the first-differenced equation, in levels. As far as the explanatory variables are concerned,

I use their differences as standard instruments in the first-differenced equation, which therefore gives only one instrument per variable. For the System GMM estimation, I use additional instruments in the levels equation: the difference of the dependent variable as the instrument in the levels equation, and the levels of all the explanatory variables as the instruments in the levels equation, which is as restrictive as it can be.

Some of these explanatory variables are suspected to be endogenous based on economic theory. However, as a first approach, I treat only the lagged dependent variable as endogenous and all variables in x_t as strictly exogenous, and then test this model specification. If the model is identified with the more powerful restriction of exogeneity required by standard instruments and passes the specification tests of overidentifying restrictions, then to the extent that tests have adequate power to detect misspecification, this is indicative of a well-specified model.

4.5.2 Estimation results

Main results are presented in Table 4.3 below. On the basis of several model selection tests (Table B.5 in the Appendix), the First Difference GMM model (6) is preferred. This model (6) is preferred both over model (5) and over the alternative specifications in case of System GMM models. These details are discussed in the Appendix. In this First Difference GMM model (6), which is preferred for estimating the determinants of maturity structure, three variables have coefficients that are statistically significantly different from zero: cETermSpread, WorkingSavers and ContractSaveAssets. These correspond to portfolio theory and clienteles theory, with the right sign. The results support the conjecture that debt managers set maturity structure so as to optimize their portfolios based on interest rates costs. The results also suggest that debt managers adjust to changes in the size of long-horizon clienteles.

Table 4.3: Cross-country determinants of maturity structure

	(1) OLS	(2) OLS	(3) FE	(4) FE	(5) FDGMM	(6) FDGMM
L.GovtMaturity	0.988***	0.989***	0.631***	0.642***	0.779***	0.773***
	(0.037)	(0.031)	(0.077)	(0.065)	(0.168)	(0.213)
c	0.062	0.057				
	(0.068)	(0.047)				
cETermSpread	-0.073	-0.064*	-0.096	-0.102	-0.191**	-0.318***
	(0.043)	(0.033)	(0.084)	(0.072)	(0.095)	(0.078)
TaxSmoothingLT	0.005***	0.004**				
	(0.001)	(0.001)				
WorkingSavers	-0.022	-0.009	0.202**	0.252**	0.212***	0.208**
	(0.013)	(0.016)	(0.065)	(0.071)	(0.061)	(0.091)
ContractSaveAssets	1.222	1.167**	1.648*	1.100**	1.222**	1.601**
	(0.784)	(0.411)	(0.954)	(0.453)	(0.525)	(0.649)
CorpMaturity	-0.004	-0.000	-0.006	-0.004	-0.011	-0.007
	(0.006)	(0.005)	(0.013)	(0.011)	(0.017)	(0.020)
RiskSpread	-0.062**	-0.049**	-0.061**	-0.061**	0.010	0.015
	(0.020)	(0.018)	(0.021)	(0.018)	(0.020)	(0.034)
InflationCredibility	-0.003**	-0.003**	-0.003**	-0.003**	0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Inflation	0.044**	0.043**	0.061**	0.063**	0.081**	0.101
	(0.019)	(0.014)	(0.027)	(0.025)	(0.040)	(0.067)
Constant	1.244	0.641	-11.298**	-14.621**		
	(0.836)	(1.078)	(4.072)	(4.483)		
Observations	241	241	241	241	213	213
Adjusted R^2	0.929	0.927	0.604	0.604		
No. countries			27	27	27	27
No. instruments					16	16
AB:AR(1)					0.048	0.032
AB:AR(2)					0.499	0.593
Hansen J					0.167	0.167

Notes: ***p<0.01, **p<0.005, *p<0.1. Year effects included in models (1) and (3), excluded from models (2) and (4). One-step weight matrix for model (5) and two-step weight matrix for models (6). Robust standard errors in parentheses. Windmeijer (2005) corrected errors are reported for two-step estimation; p-values reported for AB and Hansen J tests. The system GMM results are reported in Table B.6 in the Appendix.

The coefficient on the interaction between interest rate comovement and expected long-run term spread is negative and statistically significant in the preferred model. This means that countries with steeper expected slope of the yield curve have shorter maturity structure on average, controlling for the volatility of interest rates and covariance between the short and the long-term rate of that country. This is in line with Proposition 2 and therefore in line with debt managers that minimize costs in line with portfolio theory and a long planning horizon. This is economically significant: a one-standard deviation increase in expected long-run slope of the yield curve is correlated with weighted average remaining maturity that is shorter by 3.8 months.[18] This is also robust to an alternative measure of long-run term spread.[19]

The coefficient on the share of working-age adults in the population is positive and statistically significant determinant of the maturity structure in the preferred model. This is even more economically significant: a one-standard deviation increase in the share of savers in the population is correlated with weighted average remaining maturity that is longer by 6.1 months.[20] This positive relationship between maturity structure and the relative size of the long-horizon clientele in the population is consistent with the theory that government responds to demand-side preferences for longer maturity, in line with the clientele theory developed by Guibaud, Nosbusch, and Vayanos (2013). The growth of the assets of the contractual savings sector is also statistically and economically significant[21], controlling for general demographic changes. This further highlights the role that demographic clienteles play, and supports the hypothesis that debt managers cater to demand from pension funds and insurance corporations, which have preference for long-term assets in order to match their long-term liabilities.

[18] $1.01 \times -0.318 = 0.32$ which is 3.85 months.

[19] One alternative measure of long-run term spread is a five-year historical average term spread. This alternative measure is negatively correlated with the expected term spread. It is also statistically significant in this regression, but with an opposite sign, as expected.

[20] $2.45 \times -0.208 = 0.51$ which is 6.12 months.

[21] $0.13 \times 1.601 = 0.21$ which is 2.5 months.

One potential econometric concern is that persistency of the dependent and explanatory variables could be making the inference less valid. This is because the closer each variable is to a unit root, the lower the correlation between the first difference and lagged levels, and lagged levels are therefore less informative as instruments for the endogenous variables. However, such a concern for the power of the instruments is not justified. When testing for the power of instruments (see Table B.8 in the Appendix), the hypothesis of zero partial correlation between the instrumented variable and its instrument is rejected at 5% significance level, while the fit is good and the corresponding F-statistics are in favor of the joint significance of the variables. Therefore, the chosen instruments have explanatory power for the endogenous lagged dependent variable.

Another econometric concern is that some of the explanatory variables that were treated as strictly exogenous might be endogenous. In particular, clienteles theory suggests that term spread may be endogenous. The difference-in-Hansen J tests do not indicate that strictly exogenous specification is invalid (see Table B.9 in the Appendix). For further robustness, I also perform the Hausman tests of endogeneity, reported in Table B.10. I find that all but one of the additional moment conditions present in the First Difference GMM model (6)—beyond the moment conditions when the expected long-run term spread is treated as endogenous, one at a time—are valid. As such, I conclude against endogeneity of the expected long-run term spread. The concern for model mis-specification on the basis of endogeneity is not justified.

Overall, these specification tests and robustness checks speak to the representativeness of the main finding—that cross-sectional differences in maturity structure among advanced economies can be explained by cross-sectional variation in interest rates and demographic factors. Moreover, since these variables are included based on reduced-form equations derived from portfolio theory and clienteles theory, these results suggest that debt managers set maturity structure so as to optimize their

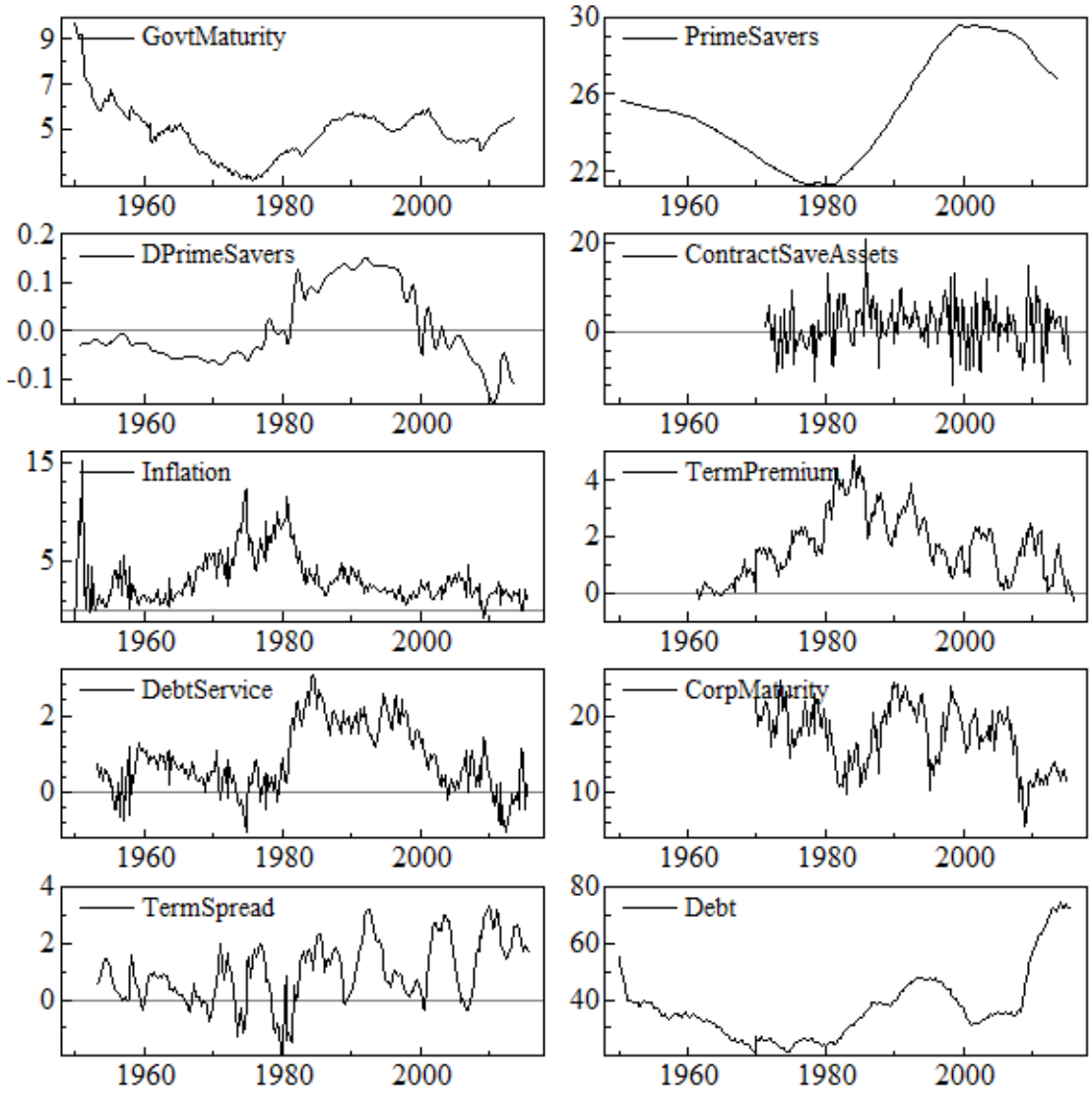
portfolios based on interest rates costs and also adjust to demographic changes by catering to demand pressure that is driven by the savers in the population.

4.6 U.S. maturity structure dynamics over time

This section presents the results of the time-series analysis that considers dynamic differences in the weighted average residual maturity in the United States between 1972 and 2013. The set of explanatory variables corresponds to the determinants derived from reduced-form expressions of optimal maturity that arise in different theoretical settings that describe the dynamic aspects of the economy.

Based on a visual inspection of the data in Figure 4.3, government debt maturity ($WARM$) of the U.S. appears non-stationary. Same is true for numerous other variables. Lack of stationarity poses serious econometric challenges because it could lead to spurious correlation between the variables in time series analysis and incorrect inference. So it is important to formally test for it. I confirm that numerous variables are non-stationary by applying the Augmented Dickey Fuller (ADF) test to detect presence of unit roots in the variables (see Table B.11 and Tables B.12-B.15 in the Appendix). I find that most variables are integrated to the fist order as $I(1)$: government debt maturity; 10-year term premium; corporate debt maturity; real government debt service; and inflation. The share of prime savers aged 35-54 is integrated as $I(2)$. The growth rate in assets of pension funds and insurance corporations and term spread are both stationary as $I(0)$. Lastly, I find that the data is balanced starting from 1971 and that will be the main sample period.

Figure 4.3: Graphical representation of the variables



4.6.1 Model and methodology

Since maturity structure and many other variables are non-stationary, it is important to apply cointegration analysis and subsequent modeling of the error correction mechanism (ECM). In the presence of unit roots, it is not appropriate to perform standard VAR analysis. Simply differencing the variables until they are stationary of

integration order $I(0)$ will not be informative about the stochastic trend of maturity structure. Short-run variation of maturity structure around the stochastic trend is only part of the story. Instead, all of maturity dynamics are relevant to the research question. Furthermore, many of the theories revolve around an equilibrium relationship between optimal maturity and other variables. Appropriately, ECM modeling allows for both a long-run relationship and short-run dynamics of deviation from that equilibrium relationship. Therefore unit root properties are not a hindrance. They are used to appropriately test the research question of maturity structure determinants.

First, I test whether maturity and other $I(1)$ explanatory variables follow a common stochastic trend in a bivariate setting. I test for presence of bivariate cointegration using the Johansen (1991) trace test. For the subset of variables that have a bivariate cointegrating relationship, I then apply the same in multivariate setting, in search of a stationary and well-specified error cointegration mechanism, $ecm_{t-1} = \beta' Y_{t-1}$. I also consider parameter restriction $\alpha_j = 0$ that refers to the equation modeling variable j : $\Delta y_{j,t} = \alpha_j[ecm_{t-1}] + ...$ If valid, this restriction of weak exogeneity implies that variable j does not respond to disequilibrium specified by the ECM, is a weakly exogenous variable, and can be considered as a causal variable to the long-run relationship, in an econometric sense (Juselius, 2006).

Next, I include the cointegrating relation in a single equation ECM model, where vector Δz contains $I(0)$ variables.

$$\begin{aligned}\Delta WARM_t =& \mu + \alpha_1^m \Delta WARM_{t-1} + ... + \alpha_p^m \Delta WARM_{t-p} + \alpha^{ecm} ecm_{t-1} + \\ &+ \gamma_1 \Delta z_{t-1} + ... + \gamma_p \Delta z_{t-p} + \epsilon_t\end{aligned}$$

I only consider lagged explanatory variables since the data is quarterly and information from previous quarters more accurately reflects what is in debt manager's current information set at time t. I start with $p = 4$ for all explanatory variables, or a higher lag order if specified by univariate lag selection. Next I select a more

parsimonious model by keeping only the most statistically significant lags for each determinants.

If parameter α^{ecm} is statistically significantly different from zero in a well-specified model with no residual autocorrelation and with an ECM that passes the vector normality and heteroskedasticity tests, this indicates that maturity structure responds to disequilibrium in the cointegrating relation (Juselius, 2006). The statistically significant variable in the ECM, such that $\beta \neq 0$, is therefore a determinant of government debt maturity structure in the long-run sense. Any such variable could also be responsible for short-run changes in the maturity structure if it is also part of the Δz vector of statistically significant determinants. This illustrates the distinction between short-run and long-run determinants that is possible with ECM modeling and is used to an advantage in this analysis of maturity structure which is a non-stationary variable.

4.6.2 Error Correction Mechanism results

Based on the Johansen (1991) trace test (see Table B.16 in the Appendix, and Tables B.17-B.20 for the corresponding lag selection), in a multivariate setting, I find that there is a single robust cointegrating relation between government maturity, inflation, and change in the share of savers. The test suggests that the three series are cointegrated with a single cointegrating vector, regardless whether lag length q is 8 or 12. This single cointegrating vector is also robust to sample period and specification (see Tables B.21-B.22 in the Appendix). A set of alternative cointegrating relations and tests are discussed and presented in the Appendix. These relations were less robust or lacking altogether.

Table 4.4: Johansen trace test for cointegration of maturity, inflation and change in the share of prime savers

Lags	Ho: rank = 0, Ha: rank ≥ 1		Ho: rank ≤ 1, Ha: rank ≥ 2		Ho: rank ≤ 2, Ha: rank=3		Single cointegrating
q	Statistic	p-value	Statistic	p-value	Statistic	p-value	vector
8	30.306	0.044*	7.5125	0.526	0.26445	0.607	yes
12	33.490	0.017*	6.9816	0.586	0.51212	0.474	yes

The sample period is 1971:Q2-2013:Q3. The trace test rejects the hypothesis of rank=0 in favour of the alternative, that rank=1, but cannot reject it in favour of the alternative that rank=2, or of the alternative that rank=3.

The error correction mechanism (ECM) $ecm_{t-1} = \beta' Y_{t-1}$ for $VAR(8)$ is the following, and standard errors are in parentheses below the coefficient estimates:

$$ecm_t = WARM_t - \underset{(1.381)}{3.356}\Delta ps_t + \underset{(0.054)}{0.389}\pi_t$$

This ECM represents a long-run relationship between these variables in equilibrium, which is when $ecm = 0$. Both the share of prime savers, ps, and inflation, π, are statistically significant for the long-run equilibrium relationship since we can reject the null hypothesis that their $\beta_j = 0$. In terms of specification, the parameter estimates are similar in case of alternative specifications and sample periods. Although this cointegrating vector does not formally pass the standard test of distributional assumptions for normality and heteroskedasticity, it is acceptable based on a visual inspection (see Figure B.1 in the Appendix). Therefore, in the long-run equilibrium, government debt maturity ($WARM$) has a:

(1) positive relationship with change in the share of prime savers in the population, Δps.

(2) negative relationship with inflation, π.

As discussed earlier, I also test parameter restrictions $\alpha_j = 0$, which are informative about weak exogeneity. The results of these tests are consistent with the notion

of a weakly exogenous demographic variable, while maturity structure and inflation do endogenously respond to disequilibrium. One caveat is that the exogeneity of the demographic variable is in an econometric sense. It is considered weakly exogenous because it does not respond to disequilibrium in this particular trivariate relationship. This does not suggest exogeneity in an economic sense, or econometric exogeneity to relationships with other macroeconomic variables. Becker (1992), for example, shows that demographic changes are endogenous to macroeconomic conditions. In this empirical application with the share of savers in the population, demographic changes potentially induced by changing inflation or maturity structure would translate into a changing share of savers 20-30 years later. I do not find evidence for this effect in the data, and instead conclude that demographics are weakly exogenous to the long-run equilibrium relationship between maturity, inflation, and the share of savers.

Table 4.5: Interpretation of equilibrium and adjustment

$\Delta \hat{WARM}_t = \underset{(0.013)}{-0.039} ecm_{t-1} + ...$	$\alpha_{WARM} < 0$: maturity structure responds to disequilibrium
$\Delta \hat{ps}_t = \underset{(0.000)}{-0.000} ecm_{t-1} + ...$	$\alpha_{ps} = 0$: change in the share of prime savers is weakly exogenous and does not respond to disequilibrium
$\Delta \hat{\pi}_t = \underset{(0.169)}{-0.446} ecm_{t-1} + ...$	$\alpha_{\pi} < 0$: inflation responds to disequilibrium

The sample period is 1971:Q2-2013:Q3.

The interpretation of the trivariate relationship is the following. In response to a disequilibrium resulting from a greater share of prime savers aged 35-54, inflation quickly increases with a relatively small magnitude until the equilibrium is restored, in econometric terms. In response to the same disequilibrium, maturity structure also increases but very slowly, and with a relatively large magnitude. In an economic sense, change in the share of prime savers that is positively and causally linked to inflation can be thought of as greater share of workers in the population influencing total factor productivity. This theoretical mechanism is not part of the formal theories

behind debt management examined earlier. To the best of my knowledge there is no formal theory yet that links together demographics, inflation and government debt maturity. Although there is an informal theoretical link between all three variables, with these particular signs and causal relationships, the error correction mechanism is above all an econometric technique.

There is one other alternative equilibrium correction mechanism that is better grounded in theory. The clienteles theory specifies that maturity, demographics, and term spread should be all jointly cointegrated. Using the $I(0)$ term spread variable indeed confirms a trivariate cointegrating relationship, as seen from Table 4.6, which is also robust to using a longer sample. The demographics variable is also weakly exogenous in the econometric sense. In the long-run equilibrium, government debt maturity ($WARM$) has a positive relationship with change in the share of prime savers in the population, Δps, while Δps has a negative relationship with term spread ts. These two relationships together exactly support the reduced-form predictions from the clienteles theory of debt management by Guibaud et al. (2013).

Table 4.6: Johansen trace test for cointegration of maturity, term spread and change in the share of prime savers

Lags	Ho: rank= 0, Ha: rank ≥ 1		Ho: rank≤ 1, Ha: rank ≥ 2		Ho: rank ≤ 2, Ha: rank $= 3$		Single cointegrating
q	Statistic	p-value	Statistic	p-value	Statistic	p-value	vector
8	32.374	0.024*	6.296	0.665	0.107	0.744	yes
12	31.975	0.027*	5.395	0.766	0.059	0.808	yes

The corresponding error correction mechanisms are:

$\tilde{ecm}_t = WARM_t - \underset{(1.846)}{5.209}\Delta ps_t - \underset{(0.167)}{0.924} ts_t$ for VAR(8), with $\alpha^{ps} = 0$ and
$\tilde{ecm}_t = WARM_t - \underset{(1.334)}{5.779}\Delta ps_t - \underset{(0.146)}{0.831} ts_t$ for VAR(12), with $\alpha^{ps} = 0$

The sample period is 1971:Q2-2013:Q3.The trace test rejects the hypothesis of rank=0 in favour of the alternative, that rank=1, but cannot reject it in favour of the alternative that rank=2, or of the alternative that rank=3.

Unsurprisingly, the two equilibrium correction mechanisms are highly correlated

(see Figure B.3 of the Appendix). Therefore, I include both $e\tilde{cm}_t$ and ecm_t equilibrium correction mechanisms from the respective $VAR(8)$ in the estimation.

4.6.3 Estimation results

Starting from a single predictive equation with $p = 4$ lags for each explanatory variable in vector Δz, or more as based on lag selection (see Appendix), I keep one most statistically significant lag for each explanatory variable. I also include three quarter dummies to capture seasonality effects. Between ecm and $e\tilde{cm}$, ecm is more statistically significant, so the other one was excluded. The resulting single equation model is:

$$\begin{aligned}
\Delta WARM \;=\; & \underset{(0.052)}{0.0632} \;+\; \underset{(0.077)}{0.241^{***}}\;\Delta WARM_{t-1} \;+\; \underset{(0.078)}{0.211^{***}}\;\Delta WARM_{t-2} \\
& + \underset{(0.081)}{0.146^{*}}\;\Delta WARM_{t-3} \;+\; \underset{(0.079)}{0.0627}\;\Delta WARM_{t-4} \;-\; \underset{(0.0085)}{0.0211^{**}}\;ecm_{t-1} \\
& - \underset{(0.014)}{0.0284^{**}}\;\Delta\rho_{t-4} \;+\; \underset{(0.027)}{0.0644^{**}}\;\Delta rd_{t-2} \;+\; \underset{(0.0094)}{0.0168^{*}}\;\Delta\pi_{t-2} \\
& + \underset{(0.001)}{0.000693}\;csa_{t-3} \;+\; \underset{(0.0025)}{0.00369}\;\Delta pmat_{t-1} \\
& + \underset{(0.019)}{0.0609^{***}}\;Q_t \;+\; \underset{(0.021)}{0.128^{***}}\;Q_{t-1} \;+\; \underset{(0.019)}{0.0479^{**}}\;Q_{t-2}
\end{aligned}$$

Table 4.7: Model specifications

N =	167	Adjusted R^2 =	0.373
$F_{AR1-5}(5, 148)$ =	1.335 (0.253)	$N_{Normal}(2)$ =	10.41^{**} (0.005)
$F_{ARCH1-4}(4, 159)$ =	3.1913^{*} (0.015)	$F_{Hetero}(23, 142)$ =	1.60 (0.051)

These diagnostic tests do not detect evidence of misspecification in terms of serial correlation of residuals or heteroscedasticity. Although the residuals do not pass the test for Normality, the deviation is not substantial based on visual inspection (see Figure B.4 of the Appendix).

It follows that maturity structure responds to disequilibrium in its trivariate relationship with demographics and inflation because α^{ecm} is statistically significantly different from zero. I can therefore confirm that in a long-run equilibrium, long-horizon

clienteles drive the positive relationship with maturity structure, in an econometric sense. This is a confirmation that debt managers respond to demographic clienteles. Furthermore, the fact that alternative $e\tilde{c}m$ is so robustly specified, as per Table 4.6, suggests that debt managers do not fully cater to the clienteles, allowing the term spread to vary. This lends support to the Guibaud, Nosbusch, and Vayanos (2013) model.

There is also support for portfolio theory among short-run determinants of government debt maturity, which entails minimizing costs of debt service. When long-term bonds become relatively more expensive as compared to rolling over short-term bonds, i.e., when the term premium increases, this is associated with future shortening of the maturity structure. In the short run, maturity is also positively associated with inflation and real debt service. The fact that the relationship is positive is contrary to what debt managers should be doing to maintain an anti-inflationary stance as per the model of Missale and Blanchard (1994). Other determinants that correspond to competing theories are not statistically significant.

4.7 Conclusion

Following a review of structural models of optimal maturity, this paper tests reduced-form specifications of maturity determinants as competing hypotheses. Panel setting is used to test steady-state differences in maturity structure, which correspond to cross-country determinants. I use GMM due to persistence properties of maturity structure. To test for time-varying determinants of maturity structure, I focus on the United States during 1972-2013. This sample period driven by data availability. I exploit the unit root properties of maturity structure and other variables with an equilibrium correction mechanism (ECM), and estimate theoretically-consistent long-run and short-run relationships.

I confirm that debt managers choose maturity structure so as to minimize expected cost of debt service in line with cost-risk portfolio optimization theory. This entails responding to term premium in the short-run, as in the example of the United States. In cross-country terms, countries with steeper long-run slope of the yield curve have shorter maturity structure, controlling for the volatility of interest rates and covariance between the short- and the long-term rates. Inflation also seems to play a role, though not exactly as specified in the reduced-form model of optimal maturity structure consistent with maintaining an anti-inflationary stance.

Most importantly, I find that the role of demographic clienteles is statistically and economically significant. For the U.S., this is based on the finding of a positive long-run equilibrium relationship between maturity structure and the share of prime savers in the population. In all developed economies, there is a positive association between the size of the working age population and maturity structure. Moreover, debt managers respond to growing assets of institutional investors (pension funds and insurance corporations), even after controlling for demographic changes. The estimated role of demographic clienteles is also economically significant.

Furthermore, the clienteles theory prescribes some catering to demographic clienteles, but not enough to fully offset demand pressures (Guibaud et al., 2013). I am able to test for this in time series setting, with evidence to support that there is a positive equilibrium relationship between demographic clienteles and maturity structure, and a negative equilibrium relationship with term spread.

Nowadays, many pension funds face funding pressures. One of the contributing factors is a prolonged period of low interest rates. Discounting promised pension fund payments at lower interest rates translates into higher present value of pension fund liabilities. This in turn requires a higher present value of assets. Therefore, falling interest rates have translated to widening pension fund deficits. This is particularly acute for defined benefit pension funds. Furthermore, in the aftermath of the financial

crisis, pension funds seeking safe returns that have good liability-matching properties have been fuelling the demand for long-term government debt (Bird, 2016). As a result, sovereign debt mangers have been facing downward pressure on long-term interest rates. Without full catering to increased pension fund demand, downward pressure on long-term interest rates persists and long-term interest rates remain low. This further exacerbates the problem of pension fund solvency.

As such, there is a negative feedback loop in place. This negative feedback loop is due to a combination of three factors: the zero lower bound, high demand for long-term government bonds by pensions funds, and a debt management policy that caters to such clienteles demand imperfectly. The imperfect catering is indeed prescribed by the clienteles theory. However, what may be optimal from the perspective of intergenerational risk-sharing may no longer be optimal once the expected welfare cost of potential pension fund bankruptcy is taken into account in the formal model. Future theoretical research should address the way in which demographic clienteles and stability of the institutional investor sector should influence debt manager's choice of maturity structure more generally.

Chapter 5

Be Careful What You Ask For: Fundraising Strategies in Equity Crowdfunding

5.1 Introduction

How much money different types of entrepreneurial companies raise is one of the central questions in entrepreneurial finance. For example, large literature examines the role of prior entrepreneurial experience for raising funds. Another important question is whether gender matters for fundraising. Much of the prior literature is limited by the fundamental difficulty of disentangling what factors can be attributed to the investors versus to the entrepreneurs. If we observe that a particular founder characteristic is associated with lower investment amounts, is that because the investors wanted to invest less, or because the entrepreneurs asked for less?

Equity crowdfunding is a recent financial innovation that constitutes an alternative method for raising funds from "the crowd", meaning from private individuals. In this paper, we exploit proprietary data from SEEDRS, one of the leading UK

platforms. The data records what actions are taken by entrepreneurs and investors, which allows us to disentangle how fundraising outcomes can be attributed to the investors' preferences versus the entrepreneurs' strategies.

The central question in this paper concerns the relationship between *company characteristics*, *fundraising strategies*, and *campaign outcomes*. Let us expand on each of these three core components. First, in terms of *company characteristics* we leverage the recent work of Bernstein et al. (2017), which shows that founder team characteristics are the most important determinant of investor interest. Specifically, we focus on prior entrepreneurial experience and gender as two key founder team characteristics. Second, in terms of *fundraising strategies*, we leverage the unique properties of the equity crowdfunding environment. In our context, entrepreneurs can determine an investment goal, as well as the price of equity, at the beginning of the campaign. If after 60 days investors' pledges fall short of the investment goal, the campaign is declared a failure and no investments take place at all. If, however, the goal is reached, the campaign stays open for an obligatory 7-day cooling off period. After that, the entrepreneur can determine when to close the campaign. This structure naturally breaks down the fundraising process into decisions made by the entrepreneur (namely the initial goal and price, as well as the final closing decision) versus decisions made by the investors (pledges made over the course of the campaign). Third, in terms of *fundraising outcomes*, we can observe not only whether campaigns succeed, but also exactly how much money is raised, and how long the entrepreneur holds out for more pledges before closing the campaign.

Our data encompasses campaigns on SEEDRS for the period July 2012 (when SEEDRS started) to June 2017. To obtain further information on founder characteristics we augment the data with information available online on Linked-In. We have complete information on 823 campaigns, of which 339 (41%) were successful. These campaigns involve 622 distinct companies and comprise 35,576 investors making a to-

tal of 113,632 pledges. Among the successful campaigns, the average goal is £205K, and the average amount eventually raised is £296K. Across all campaigns, 25% of all founder teams have at least one founder with prior entrepreneurial experience. 9% of all campaigns only have female founders, and another 16% have founders of mixed gender.

Our empirical analysis is divided into four steps. The first step is to understand the cross-sectional determinants of how entrepreneurs set up their campaigns. For this, we look at what determines campaign goals and company valuations. We find that experienced teams set higher campaign goals and higher pre-money valuations, and we control for industry and calendar time fixed effects throughout the analysis. Female teams set lower goals and lower valuations (although not all coefficients are significantly different from zero). These findings suggest that experienced teams pursue relatively more ambitious strategies, while female teams pursue safer ones.

The second step considers campaign success, defined as reaching the campaign goal and therefore receiving some funds. For this, we estimate the probability of success in the cross-section of campaigns. There is an important conceptual distinction between direct versus total effects. A company characteristic, such as experience or gender, can have a direct effect on campaign success, as well as an indirect effect through the campaign choices, namely the goal and the valuation. The total effect is then the sum of the direct and indirect effect (Hayes, 2014). For example, we find that experience has a strong direct effect on success, suggesting that investors prefer experienced founder teams. The total effect is still positive but lower than the direct effect, suggesting a negative indirect effect. This is due to the fact that experienced entrepreneurs set higher goals, and that campaigns with higher goals are less likely to succeed. We think of the direct effects as driven by investor preferences, as they measure what investment amounts are allocated to different characteristics of the entrepreneurial teams. The indirect effects are driven by the entrepreneurs' fundraising

choices, namely the investment goal and valuation. The entrepreneurs will naturally have expectations about what they think they can get from the crowd, but then make strategic choices about what they actually want to ask from their investors. Thanks to the richness of the equity crowdfunding context we can thus isolate these strategic fundraising choices of the entrepreneurs. Interestingly, we also find that there are no significant gender effects, neither total nor direct. However, looking at campaign success alone is not enough to fully understand the role of gender. We also need to consider investment amounts and campaign closing decisions examined in the final two steps.

The third step considers investment amounts in the cross-section. Here we ask how different company characteristics and campaign choices affect the final amount of money raised. This analysis poses the econometric challenge of a selected sample because investment amounts are only observed conditional on success. Moreover, campaign success is not random, as shown in step 2 of the analysis. Therefore, we adopt a Heckman selection procedure that utilizes some additional variables that affect campaign success. Specifically, we consider several events that occur after the onset of a campaign, and that may affect campaign outcomes, such as the number of competing campaigns on the platform. We also consider weather conditions, based on prior research on how weather affects participation in online and indoor activities (Cardona et al., 2013; Gilchrist and Sands, 2016; Xu, 2017). After controlling for selection, we find that while the total effect of experience on investment amounts is positive, the direct effect is actually negative. This suggests that the overall positive relationship between experience and investment amounts is (more than) fully explained by the higher goals set by experienced entrepreneurs. In contrast, female teams raise significantly less money. This is largely driven by the indirect effect of asking for less money in the first place. The direct effect itself is more nuanced: it is negative for all female teams, but positive for mixed gender teams.

The analysis so far suggests that the initial ask of the entrepreneurs is a key determinant of final investment outcomes. However, while the cross-sectional analysis can control for the entrepreneurs initial strategic choices, it still contains some confounding effects in terms of the final amounts. Specifically, is the final amount lower because investors did not want to provide more money, or because the entrepreneurs did not want to take additional money? For the final step, we therefore leverage the dynamic structure of our data. We estimate a dynamic panel model with two equations. The first equation measures investor demand for shares, as a function of company characteristics and other intervening events (similar to competing campaigns and weather conditions discussed above). The second equation estimates the decisions to close their campaign, which is the final choice made by the entrepreneurs. This optimal stopping decision takes into account the endogenous investment flows which are estimated in the first equation.

We find that entrepreneurs facing higher investment flows are slower to close their campaigns, thus taking advantage of the continued investor interest. Once we control for this, we can separate out how company characteristics separately affect investment flows and the stopping decision. After reaching the investment goal, experienced teams no longer receive higher investment flows. This can be interpreted as saying that by the time they reached their investment goals, inexperienced teams have caught up with experienced teams in terms of investor interest. In terms of optimal stopping decisions, we do not find any significant differences between experienced and inexperienced teams either. The results for gender are markedly different. All-female teams face significantly lower investor demand after reaching the goal. However, they are not likely to close their campaigns earlier. On the contrary, they actually hold out for money longer. This is an important finding as it suggests that the lower investment amounts found in step three of the analysis are not only explained by the entrepreneurs choices. While it is true that all-female teams chose lower investment

goals, it is also true that they receive less interest from the crowd. Most interestingly, at the end of the campaign, all-female teams try to compensate for the lower investment flows and the lower initial investment goals, by keeping their campaigns open for longer.

In this paper, we use the context of equity crowdfunding to revisit the fundamental question of how much money is raised by different entrepreneurs. The general phenomenon of crowdfunding is well established by now. It is based on an electronic platform that enables a wide variety of individuals and organizations to raise funding from the general public (the crowd), for a wide variety of causes. Broadly speaking, there are three types of crowdfunding models. First, there is crowdfunding based on donations and rewards. Prominent examples of such platforms include Kickstarter, IndieGoGo, GoFundMe, and CrowdRise. The works of Agrawal et al. (2014) and more recently Strausz (2017) examine the fundamental economics of this type of financing. Second, there are so-called peer-to-peer lending platforms, where individuals and organizations seek debt funding from the crowd. Prominent examples include Lending Club, Proper, and Funding Circle. The recent academic work of Mohammadi and Shafi (2016) and Havrylchyk and Verdier (2018) discuss the role of these lending platforms. Third, there is equity crowdfunding, which is solely available to companies, mostly entrepreneurial ventures that are developing novel products and services. The crowd invests on the basis of receiving equity in the company. An important distinction is between accredited investors (who satisfy a requirement of minimum wealth or income) and non-accredited investors (who we can think of as retail investors). The leading equity crowdfunding platform for accredited investors is AngelList, which allows these investors to find deals, and in some cases, the platform also intermediates the actual investments. For unaccredited investors (the true crowd), the UK is widely acknowledged as the most developed market. This is largely because the Financial Conduct Authority (the relevant regulator) adopted a laissez-faire approach in the

early days of the industry. A 2017 report estimated over 300 successful investment campaigns in 2017, making crowdfunding the second largest investor category in the UK (by the number of companies), after venture capital, but ahead of corporate investors or angel networks (see Halmari et al., 2017). The two largest platforms in the UK are CrowdCube and SEEDRS.

While there are some similarities between equity- and reward-based crowdfunding, namely the information flow between investors, it is fundamentally a very different situation. For example, an investor may invest in a company without the intention of ever purchasing the company's products, simply because she thinks it is a good investment. Viewed in this light, the equity crowdfunding environment is closer in many ways to traditional angel and VC investments than to reward-based crowdfunding.

An important contribution to understanding equity crowdfunding is the aforementioned paper by Bernstein et al. (2017) which uses data from AngelList. The work of Estrin and Khavul (2015) is based on the UK data from CrowdCube, and examines how rational investments from the crowd are. In a related vein, Åstebro et al. (2017) use the SEEDRS data to examine the question of herding and information cascades. Ahlers et al. (2015) focus on signalling from the investor's side, and other scholars focus on the importance of entrepreneurial networks and geography (Lukkarinen et al., 2016; Vismara, 2016a; Agrawal et al., 2015; Mollick, 2014). There is also theoretical scope for entrepreneurs to learn from equity crowdfunding about project quality and investor demand (Belleflamme et al., 2014; Chen et al., 2016). This is also the case for rewards-based crowdfunding (Chemla and Tinn, 2017; Ellman and Hurkens, 2015; Strausz, 2017). Another question is whether equity crowdfunding is complementing or competing with angels. Hornuf and Schwienbacher (2016) and Estrin et al. (2017) find that it provides additional sources of entrepreneurial finance.

Our analysis builds on this nascent literature in crowdfunding and also on the established literature in entrepreneurial finance (see Da Rin et al., 2013). One im-

portant question is the role of entrepreneurial experience and the benefits of being a serial entrepreneur. Gompers et al. (2010) find that around 10% of all VC-backed founders are serial entrepreneurs. Hsu (2007) find that serial entrepreneurs have better chances to obtain venture financing, and obtain better valuations. Colombo and Grilli (2005) look at a broader spectrum of human capital characteristics and find prior entrepreneurial experience to be positively related to company growth. Distinguishing between different types of human capital, Gimmon and Levie (2010) highlight the role of experience over education. Unger et al. (2011) provide a useful meta-analysis of the related literature. The work of Wernerfelt (1984), Rajan and Zingales (2001), and Rajan (2012) also speaks to the broader role of human assets in start-ups. Agarwal and Gort (2002) consider learning by doing and endowments as key characteristics of start-up quality in the model of firm life cycle survival.

Another strand of the finance literature investigates the role of gender. Experimental evidence suggests that women are more risk-averse than men in the population (Eckel and Grossman, 2008; Croson and Gneezy, 2009), which may or may not be the case once we consider self-selection into the occupation in business and finance (Adams and Ragunathan, 2017; Atkinson et al., 2003; Huang and Kisgen, 2013; Adams and Ferreira, 2009). Specifically, in crowdfunding, there is evidence that female investors are more risk-averse (Hervé, 2017). There is also large literature that considers differences by gender in entrepreneurial finance, by examining start-up capital, bank loans, external equity (Verheul and Thurik, 2001; Coleman and Robb, 2009; Robb and Coleman, 2010) using various sources like the Kauffman Firm Survey and a large panel of Dutch entrepreneurs. These studies find differences in fundraising strategies by gender: female entrepreneurs have a smaller amount of start-up capital and use external debt more. These factors, in addition to less prior business experience, explain why female-owned businesses are less successful than male-owned businesses (Fairlie and Robb, 2009). The gender of investors seems to

matter as well, particularly in the VC industry (Gompers and Wang, 2017b; Raina, 2017; Gompers and Wang, 2017a).

There is also recent literature that discusses the role of gender in crowdfunding. Marom, Robb and Sade (2016) show that on the rewards-based Kickstarter platform, men seek significantly higher levels of capital and raise more funds than women. Yet there is also gender-based matching between entrepreneurs and investors (Marom et al., 2016; Ewens and Townsend, 2018) and the extent to which this affects success rates depends on the number of female investors on the platform: many on Kickstarter, few on AngelList. Curiously, Johnson et al. (2018) find that investors tend to have an implicit positive bias towards women on Kickstarter. Our analysis complements this literature and sheds light on differences in fundraising strategies, investment flows, and outcomes by gender in the UK market for equity in early-stage companies.

The remainder of this paper is structured as follows. Section 5.2 introduces our data, variables, and descriptive statistics. Section 5.3 contains the main empirical analysis. Section 5.4 looks at the robustness and extensions. It is followed by a brief conclusion.

5.2 Data

5.2.1 Data sources

We use three main data sources: (i) proprietary data from SEEDRS, a UK based equity crowdfunding platform; (ii) publicly available LinkedIn profiles of entrepreneurs; and (iii) incorporation documents and annual returns from Companies House, a publicly available register of all companies incorporated in the UK.

SEEDRS provided us with the data for the period 2012-2017, which covers 1,125 campaigns, with 135,053 investments made by 39,555 investors. We exclude 55 fund

campaigns and convertible campaigns, 25 campaigns that were still ongoing in June 2017 and 1 campaign that was missing valuation data. The first equity crowdfunding campaign in our sample received its first investment on July 4th, 2012 and the last campaign closed on June 3rd, 2017.

The SEEDRS data includes identifying information about the founders (entrepreneurs) that were running these crowdfunding campaigns. This includes company name, names of entrepreneurs, their titles and roles in the start-up, and their equity shares. We use this information to identify the founders (entrepreneurs) manually. We define a founder as an individual with a management role and an equity stake that exceeds 5%. We do not count Non-Executive Directors, entrepreneurs with advisory roles, or entrepreneurs with a small equity stake as part of the founding team. Since most of these are British companies, we also manually looked up their Companies House UK incorporation record and first annual return, where available. We used this as a supplementary data source to identify and confirm the original start-up founders and their equity shares in the company[1].

For each founder of a company that ran a SEEDRS equity crowdfunding campaign, we hand-collected information about their educational and professional background from their publicly available LinkedIn profiles. We also used the information about the team that was included on the SEEDRS campaign page. As such, we gathered a number of data on entrepreneurs education and professional experience, as well as gender. The professional experiences include both the current company of the SEEDRS crowdfunding campaign, as well as other employment, prior, concurrent, and possibly after the campaign. We gathered and coded this data for 1,427 entrepreneurs from 792 companies; for 36 companies, none of the founders had LinkedIn or any information on their SEEDRS profile. These 1,427 entrepreneurs ran 1,006 SEEDRS

[1]We used Companies House data as a supplement only in case of companies that ran successful campaigns. Companies that never ran a successful campaign were likely short-lived, not incorporated or incorporated and dissolved. Therefore we decided not to gather Companies House incorporation data for companies that only ever ran unsuccessful campaign(s).

campaigns, which constitutes our overall data sample.

We also use the data on campaign goals, pre-money valuation, equity offered, and all investment pledges associated with the campaign, which directly come from SEEDRS. It should be noted that relative to traditional databases of entrepreneurial finance (such as Thomson One for venture capital deals), the quality of the SEEDRS data is clearly superior. Valuation data, for example, is notoriously hard to obtain in these traditional databases. Moreover, the pledge-level investment data is very granular and allows us to look at daily dynamics of investment flows into each campaign.

We also collected the data on weather, such as daily temperatures and rainfall, from OGIMET Weather Information Service. Most investors and founders on the platform are based in the United Kingdom, and in London in particular. Thus, we focus on daily temperatures and rainfall recorded by the London Heathrow weather station, or else by the London St James Park weather station if the Heathrow data is not available.

5.2.2 Variable definitions

We consider both successful and unsuccessful campaigns. A campaign is successful when it raises at least as much money as the campaign goal. SEEDRS average campaign success rate is approximately 35% overall. Including the data on both successful and unsuccessful campaigns allows us to avoid biased regression estimates due to the selection problem, which will be explained in more detail later.

We aggregate entrepreneur-level variables into campaign-level data to be able to run the analysis at campaign-level. We focus on team averages and dummy variables that measure the presence of certain characteristics, such as gender and experience, which are the main focus of this paper. For gender, for example, we distinguish three categories of companies: all-male, all-female, and mixed gender. We also calculate

the share of founders that are female. We similarly calculate our main measures of past start-up experience. We calculate shares and dummies for whether any founders have both (1) experience founding a company and (2) prior experience with start-up success in a different venture. Start-up success experience is defined as an IPO, acquisition, private investment, or business growth[2].

Next let us explain how investment pledges work on the SEEDRS platform, and how we use the pledge-level investment data to calculate the amounts of funds raised every day by each campaign. Most investors show an interest by registering a pledge, and then pay the money if the campaign is successful. However, investors are also allowed to either cancel the investment (at any time) or to leave the investment unpaid once the campaign is successful. Indeed, some investors cancel pledges or don't pay them. We exclude such pledges that were cancelled or left unpaid, so as to focus on realized investment amounts. An analysis of strategic investor behaviour is outside the scope of this paper, but the work by Åstebro et al. (2017) uses the same dataset to examine exactly that.

After founders close their campaigns, they can also reject some of the funding. In the data, this is common, but the amounts the entrepreneurs reject are very small. We exclude the rejected pledges from the analysis. Lastly, we also exclude pledges that were rejected by SEEDRS due to concerns about money fraud, or due to investors' input errors. As such, we use the universe of non-cancelled and non-rejected investment pledges to calculate daily investment flows into each campaign. We also calculate platform-level daily aggregates that measure competition, such as the average investment amount on the platform and the number of open campaigns.

In the design of our empirical analysis, we also leverage the fact that majority of the companies on the SEEDRS platform are eligible for EIS (Enterprise Investment

[2]In Section 5.4 on robustness we also look at three more entrepreneur characteristics, namely commitment to the venture (entrepreneurs working part-time, i.e., elsewhere during the SEEDRS campaign), ownership structure (presence of a majority owner), and business education through an MBA degree.

Scheme) and SEIS (Seed Enterprise Investment Scheme) tax breaks for investors. In the UK, investors, including directors, can receive initial tax relief of 50% on investments up to £100,000 and Capital Gains Tax (CGT) exemption for any gains on the SEIS shares. Beyond this investment level, there is EIS: investors can invest up to £1,000,000 in unlisted companies in any tax year receive 30% tax relief. We therefore also calculate a tax year dummy variable covering the last three weeks of the UK personal income tax year, which ends around April 5th each year.

Definitions and notes about the main variables follow in Table 5.1 below. These include all dependent variables and independent variables, and include campaign strategies and outcomes, founder team characteristics, as well as instrumental and control variables.

5.2.3 Descriptive statistics

We consider a balanced data sample, such that none of the dependent or explanatory variables relevant for our analysis are missing. For this, we can only use 823 campaigns from the complete sample of 1,006 campaigns. Our balanced sample of 823 campaigns includes 339 successful campaigns and 484 unsuccessful campaigns.

In this balanced sample, an average campaign goal is £190K, and average pre-money valuation of a company is £2.5M, with 10% equity offered at goal (Figure 5.1 Panel A). Turning to campaign outcomes in Figure 5.1 Panel B, teams have an average success rate of 41%[3]. Teams from the 339 successful campaigns in our balanced data sample were able to raise £296K on average, which corresponds to a 139% funding-to-goal ratio.

There is substantial heterogeneity in these fundraising strategies and outcomes,

[3]This is marginally higher than the platform-wide campaign success rate of 35% due to the fact that entrepreneurs that ran unsuccessful campaigns had less information listed on their LinkedIn profiles and were more likely to be excluded from the balanced data sample. Unsuccessful campaigns are also associated with the unknown sector (industry) of the company more frequently, which also led to the exclusion.

both by experience and by gender. Experienced teams set higher goals and company valuations relative to inexperienced teams and offer the same ownership share as equity in the campaign. A team that has at least one experienced founder has an average goal of £243K and average valuation of £3.8M, whereas a team of inexperienced founders asks for £74K less as a target fundraising amount for equity in the companies that are valued £1.7M less (Figure 5.1 Panel A). Female teams, especially females-only teams, set lower goals and have lower company valuations compared to all-male teams: all-female teams ask for £76K less than all-male teams, for roughly the same equity in companies that are valued £1M less.

These descriptive statistics alone already suggest that experienced teams pursue more ambitious fundraising strategies, whereas female teams pursue safer fundraising strategies. Some of this variation could naturally be because these teams are running equity crowdfunding campaigns in different sectors, or because all-male teams are also the experienced ones. We will explore this more deeply in our formal regressions analysis.

There is heterogeneity in fundraising outcomes by experience and gender (Figure 5.1 Panel B). While there are no substantial differences in the likelihood of success by gender, there are differences by experience. Among all 823 campaigns, experienced teams reach the goal 50% of the time on average, as compared to 38% for inexperienced teams. Experienced teams also raise more money: a team that has at least one experienced founder raises £75K more than a team with inexperienced founders. This comparison is in a sample of 339 successful campaigns, by definition, since campaigns that do not reach their goal target and that are therefore unsuccessful do not raise any money at all.

Female teams have similar prospects for campaign success as all-male teams, but they raise less money. An all-female team raises £226K less on average, a magnitude that is three times more than the difference in goals between the two types of teams.

The data thus suggests that different fundraising strategies may be linked to different outcomes, by experience, and by gender.

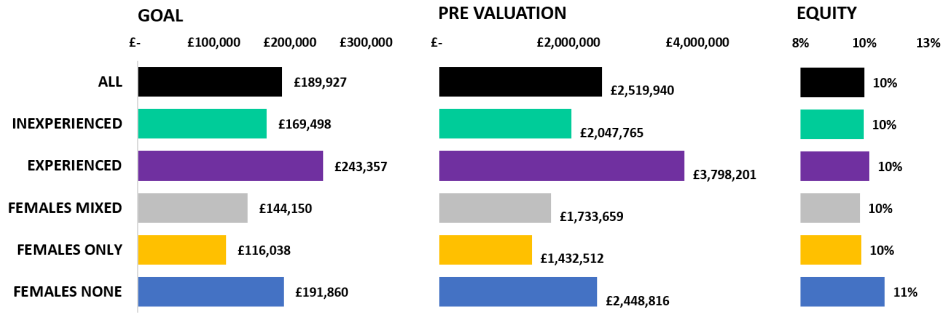
The descriptive evidence that all-female teams pursue safer fundraising strategies and raise less money does not really tell us whether all-female teams want less money and it is their choice to raise less, or whether they receive less money from the investors on the platform. That is because we don't know the direction of causality. However, in addition to the choice of campaign goal, pre-money valuation, and equity to offer at goal, founders of campaigns that are successful also have a choice over when to stop the campaign. Once a campaign reaches the target, investors can still pledge funds, at the same price per share as before. There is a cooling off period that lasts 7 days, and after 7 days the founders are allowed to close the campaign. In other words, founders can stop the campaign immediately, on the 8th day after reaching the goal, or they can wait for another week, or two, or even longer.

Table 5.1: Description of variables

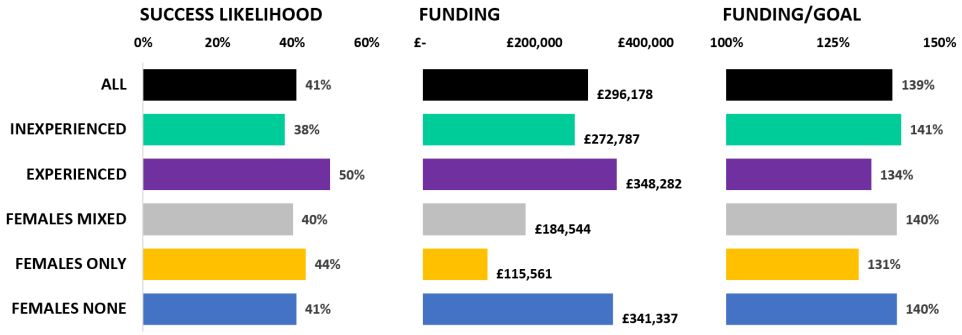
Variable	Variation	Description
GOAL	Cross-sectional	Desired campaign investment amount, as natural log.
PRE VALUATION	Cross-sectional	Pre-money valuation of the company, as natural log.
EQUITY OFFERED	Cross-sectional	Equity offered, as GOAL / (GOAL + PRE VALUATION), in percent.
SUCCESS	Cross-sectional	Campaign success dummy.
FUNDING	Cross-sectional	Total amount invested in campaign, as natural log.
OVERFUNDING	Cross-sectional	Funding amount divided by the campaign goal, in excess of 125%.
INVESTMENT FLOW	Cross-sectional, time-series	Daily campaign investment flow, as natural log.
STOP	Cross-sectional, time-series	Daily dummy whether the campaign has stopped.
EXPERIENCE	Cross-sectional	Share of entrepreneurs in the company's founding team with past founder experience in a successful startup, and as dummy (any). Successful startup experience includes an IPO, acquisition, private investment rounds, or business growth.
FEMALES	Cross-sectional	Share of female entrepreneurs in the company's founding team, and as dummy (all females, mixed gender, all males).
PRIOR SEEDRS	Cross-sectional	Dummy variable indicating whether the company had already raised some equity from SEEDRS.
TEAM SIZE	Cross-sectional	The number of company founders at the time of the SEEDRS campaign.
INVESTMENT STOCK	Cross-sectional	Cumulative investment stock on the last day of the campaign's cooling off period, as natural log.
OTHER NUMBER	Cross-sectional, time-series	The number of other open campaigns on the platform, daily.
	Cross-sectional	Average no. other open campaigns on the platform during the first week of each campaign.
OTHER FLOW	Cross-sectional, time-series	The average flow to other open campaigns on the platform, daily.
	Cross-sectional	Average flow to other open campaigns on the platform during the first week of each campaign.
TEMPERATURE	Time-series	Average daily temperature in London, as dummies in 5C increments.
	Cross-sectional	Dummies for average London temperatures during the first week of each campaign.
RAIN	Time-series	Average daily rainfall in London, as natural log.
	Cross-sectional	Average rainfall in London during the first week of each campaign, as natural log.
TAX YEAR	Time-series	Last three weeks of the UK tax year, as a dummy.
	Cross-sectional	Dummy whether first week of campaign coincides with the last three weeks of the UK tax year.
SECTOR	Cross-sectional	A series of dummy variables for Arts, Music & Design, Consumer Products, E-Commerce & Markets, Finance, Food & Drink, High Technology, and Other.
QUARTER	Cross-sectional	A series of dummy variables denoting the quarter of campaign start, for the period Q2 2012 to Q2 2017.
DAY	Cross-sectional, time-series	A series of dummy variables denoting the day of the week.

Figure 5.1: Heterogeneity in fundraising strategies and outcomes by experience and gender

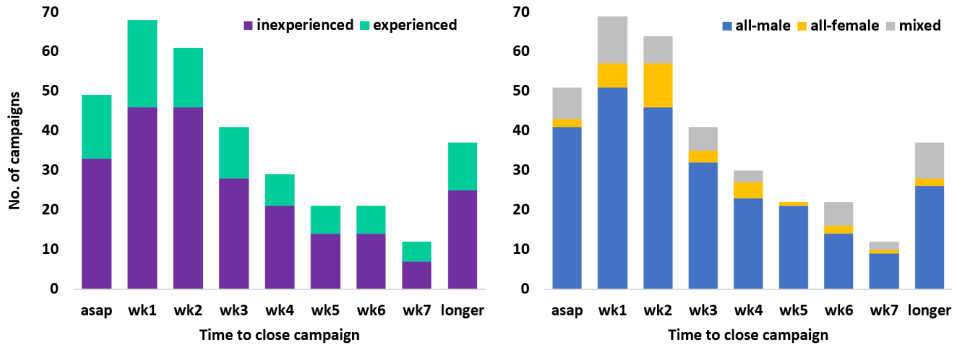
Panel A: Campaign goal, Pre-money valuation and Equity offered



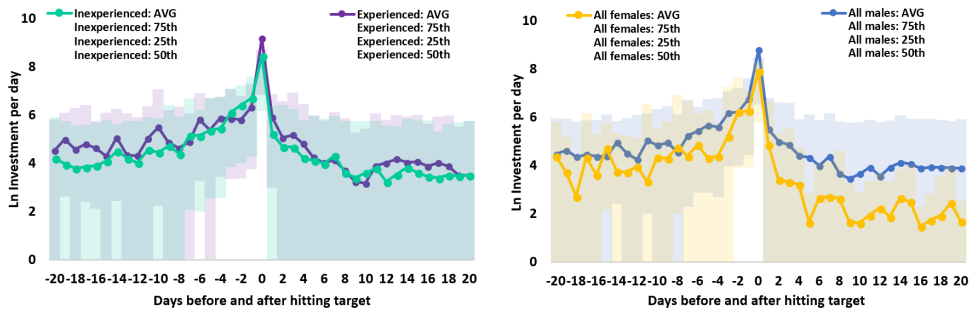
Panel B: Success likelihood, Funding raised, and Funding-to-Goal ratio



Panel C: Length of time to close the campaign



Panel D: Daily investment flow



When looking at the time the founders take to close the campaign, shown in Figure 5.1 Panel C, there are again possible differences by gender and by experience. The data suggests that inexperienced and experienced teams take a similar length of time to close their campaigns. All-female teams, on the other hand, mostly close campaigns within the first two weeks, whereas all-male teams mostly close campaigns within the first week. However, these differences are hard to detect with simple descriptive statistics, so they will be under scrutiny in the empirical analysis that follows.

The founders' decision whether to stop the campaign is likely a function of investment flows the founders have observed—whether the funds keep flowing in, or whether the investment flows have dried up—in addition to how much money the founders wish to raise. The investment flows may in turn be a function of entrepreneur characteristics, such as experience and gender, as well as a function of other events and variables. Modeling the investment flows and founder's stopping decision is the focus of our panel analysis that follows. In our panel analysis, we consider a balanced sample of 328 successful campaigns that stopped on or after the 7th day the campaign reached the target goal[4]. This corresponds to a total of 7,892 observations.

We observe some heterogeneity in investment flows by gender and not by experience (Figure 5.1) Panel D. Average investment flows to all-female campaigns appear to be lower than investment flows to all-male teams, once these campaigns have hit the target goal. Once again, it is difficult to say whether these flows are statistically significantly different, and this is the question we consider in the panel analysis.

Table 2 provides the summary statistics for the variables that are used in the analysis of cross-sectional variation of fundraising strategies, outcomes, and founding team characteristics. Monetary variables are in natural logs. In our balanced data sample of 823 campaigns, about 25% of campaigns have at least one experienced

[4]For 11 campaigns, it is not clear whether they stopped during the cooling off period, so these are excluded from this analysis

founder. The split by gender is representative of what is usually observed among start-ups: 76% of campaigns are led by all-male teams, only 8% of campaigns are led by all-female teams, and 16% are led by mixed-gender teams. As discussed earlier, we consider both dummy and shares team-level aggregates, so in addition to gender dummies, we also look at the share of females in a team. On average this share of females is 15%, in a typical team that is composed of 1.8 founders.

Some companies tap into equity crowdfunding on the SEEDRS platform multiple times and run several sequential campaigns at different stages of the same start-up venture. In our balanced sample, about 25% of campaigns are repeat campaigns.

Table 5.3 shows the differences in campaign strategies between unsuccessful and successful campaigns. Successful campaigns on average have higher goals and higher pre-money valuations, but are the ones where at goal, founders offer less equity (not taking into account any other factors such as campaign sector, quarter effects, or founder characteristics). As illustrated earlier, there are also some differences between successful and unsuccessful campaigns in terms of founder experience, but no differences in terms of the gender composition. Experienced teams are more likely to be successful, whereas the distinction by gender does not seem to matter.

Table 4 provides the summary statistics for the variables that are used in the panel analysis, with monetary variables as a natural log. Pairwise correlation coefficients and their significance, for all variables used in the cross-sectional analysis and panel analysis, are reported in the subsequent tables. There is a negative correlation of r=-0.12 between the share of experienced founders and the share of female founders, which is statistically significant but not a major concern regarding collinearity. The correlation between campaign goal and equity offered (calculated as goal/(goal+pre-money valuation)) is higher and also statistically significant, with r=0.35. In the robustness section, we further address potential multicollinearity concerns.

Table 5.2: Summary statistics for cross-sectional variables

Variable	**Mean**	**(Std. Dev.)**	**Min.**	**Max.**	**N**
GOAL	11.518	(1.187)	6.533	14.732	823
PRE VALUATION	13.984	(1.14)	4.382	18.317	823
EQUITY OFFERED	10.004	(6.947)	0.043	99.98	823
SUCCESS	0.412	(0.492)	0	1	823
FUNDING	11.654	(1.473)	6.566	15.286	339
OVERFUNDING	0.398	(0.49)	0	1	339
EXPERIENCE SHARE	0.181	(0.34)	0	1	823
EXPERIENCE ANY	0.254	(0.436)	0	1	823
FEMALES SHARE	0.153	(0.305)	0	1	823
FEMALES ONLY	0.086	(0.281)	0	1	823
FEMALES MIXED	0.16	(0.367)	0	1	823
PRIOR SEEDRS	0.247	(0.431)	0	1	823
TEAM SIZE	1.887	(0.907)	1	6	823
OTHER N	37.439	(14.813)	5.286	65.286	823
OTHER FLOW	6.51	(1.35)	1.218	10.826	823
TEMP $<$ 5C	0.095	(0.293)	0	1	823
TEMP 5C-10C	0.273	(0.446)	0	1	823
TEMP 10C-15C	0.341	(0.474)	0	1	823
TEMP 15C-20C	0.259	(0.438)	0	1	823
TEMP $>$ 20C	0.032	(0.175)	0	1	823
RAIN	-0.013	(1.241)	-2.303	3.096	823
SECTOR: Clothing, Home	0.154	(0.361)	0	1	823
SECTOR: E-Commerce, Marketing	0.165	(0.372)	0	1	823
SECTOR: Finance	0.131	(0.338)	0	1	823
SECTOR: Food, Drink	0.124	(0.33)	0	1	823
SECTOR: Games, Entertainment	0.101	(0.301)	0	1	823
SECTOR: Technology	0.18	(0.384)	0	1	823
SECTOR: Transport, Travel	0.145	(0.352)	0	1	823

Table 5.3: By success

	Mean Values Across Campaigns			
	All	Unsuccessful	Successful	P-Val Dif.
GOAL	£189,927 (267,139)	£179,146 (227,839)	£205,318 (314,593)	0.167
PRE VALUATION	£2,519,940 (6,038,181)	£2,176,503 (5,459,127)	£3,010,274 (6,758,038)	0.051
EQUITY OFFERED	10% (6.95)	10.7% (5.81)	8.95% (8.2)	0.000
FUNDING			£296,178 (522,002)	
OVERFUNDING			139% (61.3)	
EXPERIENCE SHARE	.181 (.34)	.148 (.31)	.227 (.374)	0.001
EXPERIENCE ANY	.254 (.436)	.215 (.411)	.31 (.463)	0.002
FEMALES SHARE	.153 (.305)	.153 (.302)	.154 (.309)	0.954
FEMALES ONLY	.0863 (.281)	.0826 (.276)	.0914 (.289)	0.659
FEMALES MIXED	.16 (.367)	.163 (.37)	.156 (.364)	0.792
TEAM SIZE	1.89 (.907)	1.87 (.912)	1.91 (.901)	0.569
PRIOR SEEDRS	.247 (.431)	.112 (.315)	.44 (.497)	0.000
Observations	823	484	339	

Table 5.4: Summary statistics for panel variables

Variable	Mean	(Std. Dev.)	Min.	Max.
STOP	0.041	(0.199)	0	1
INVESTMENT FLOW	2.994	(3.469)	0	13.687
GOAL	11.719	(1.259)	6.533	14.732
EQUITY OFFERED	8.609	(6.616)	0.056	99.98
EXPERIENCE SHARE	0.227	(0.37)	0	1
EXPERIENCE ANY	0.32	(0.466)	0	1
FEMALES SHARE	0.156	(0.298)	0	1
FEMALES ONLY	0.076	(0.265)	0	1
FEMALES MIXED	0.184	(0.388)	0	1
PRIOR SEEDRS	0.425	(0.494)	0	1
TEAM SIZE	1.916	(0.858)	1	5
INVESTMENT STOCK	11.574	(1.395)	6.533	15.227
OTHER N	38.313	(13.831)	1	67
OTHER FLOW	6.933	(1.342)	0	12.493
TEMP $<$ 5C	0.114	(0.318)	0	1
TEMP 5C-10C	0.27	(0.444)	0	1
TEMP 10C-15C	0.327	(0.469)	0	1
TEMP 15C-20C	0.248	(0.432)	0	1
TEMP 20C-25C	0.039	(0.193)	0	1
TEMP $>$ 25C	0.002	(0.045)	0	1
RAIN	-0.833	(1.653)	-2.303	3.873
TAX YEAR	0.045	(0.207)	0	1
SECTOR: Clothing, Home	0.186	(0.389)	0	1
SECTOR: E-Commerce, Marketing	0.141	(0.348)	0	1
SECTOR: Finance	0.198	(0.398)	0	1
SECTOR: Food, Drink	0.147	(0.354)	0	1
SECTOR: Games, Entertainment	0.082	(0.275)	0	1
SECTOR: Technology	0.091	(0.288)	0	1
SECTOR: Transport, Travel	0.155	(0.362)	0	1
TIME TREND	25.751	(21.954)	2	126
WEEK DAY: Sunday	0.143	(0.35)	0	1
WEEK DAY: Monday	0.144	(0.351)	0	1
WEEK DAY: Tuesday	0.145	(0.352)	0	1
WEEK DAY: Wednesday	0.144	(0.351)	0	1
WEEK DAY: Thursday	0.141	(0.348)	0	1
WEEK DAY: Friday	0.141	(0.348)	0	1
WEEK DAY: Saturday	0.143	(0.35)	0	1
N		7892		

Table 5.5: Cross-correlations, successful campaigns, all cross-sectional variables

Variables	G	EQ	EXP-S	EXP-A	FEM-S	FEM-O	FEM-M	P-S	SIZE	AGE	OTH-N	OTH-FLOW	RAIN
G	1.00												
EQ	0.35	1.00											
	(0.00)												
EXP-S	0.12	-0.02	1.00										
	(0.03)	(0.76)											
EXP-A	0.15	0.00	0.91	1.00									
	(0.01)	(0.97)	(0.00)										
FEM-S	-0.16	-0.05	-0.10	-0.12	1.00								
	(0.00)	(0.32)	(0.06)	(0.03)									
FEM-O	-0.12	-0.03	-0.04	-0.08	0.87	1.00							
	(0.03)	(0.64)	(0.44)	(0.14)	(0.00)								
FEM-M	-0.08	-0.07	-0.14	-0.10	0.34	-0.14	1.00						
	(0.13)	(0.23)	(0.01)	(0.08)	(0.00)	(0.01)							
P-S	-0.30	-0.27	0.07	0.04	-0.03	-0.01	-0.02	1.00					
	(0.00)	(0.00)	(0.18)	(0.50)	(0.53)	(0.81)	(0.70)						
SIZE	0.07	-0.01	-0.22	-0.05	-0.09	-0.26	0.44	-0.04	1.00				
	(0.17)	(0.86)	(0.00)	(0.38)	(0.11)	(0.00)	(0.00)	(0.51)					
AGE	0.23	0.10	0.13	0.16	-0.06	-0.07	-0.00	0.06	-0.05	1.00			
	(0.00)	(0.09)	(0.02)	(0.01)	(0.35)	(0.21)	(0.96)	(0.34)	(0.37)				
OTH-N	0.11	-0.08	0.04	0.09	-0.01	-0.07	0.10	0.04	0.12	0.04	1.00		
	(0.04)	(0.17)	(0.46)	(0.12)	(0.86)	(0.17)	(0.07)	(0.46)	(0.02)	(0.49)			
OTH-FLOW	0.18	-0.16	-0.02	0.01	-0.03	-0.08	0.06	0.06	0.10	0.07	0.49	1.00	
	(0.00)	(0.00)	(0.65)	(0.92)	(0.60)	(0.15)	(0.29)	(0.26)	(0.06)	(0.27)	(0.00)		
RAIN	-0.01	0.03	0.06	0.06	0.12	0.10	0.04	0.06	-0.00	0.03	-0.02	-0.15	1.00
	(0.81)	(0.55)	(0.25)	(0.24)	(0.03)	(0.06)	(0.49)	(0.26)	(0.93)	(0.66)	(0.76)	(0.01)	

Table 5.6: Cross-correlations, successful campaigns, all panel variables

Variables	G	EQ	EXP-S	EXP-A	FEM-S	FEM-O	FEM-M	P-S	SIZE	INV-S	OTH-N	OTH-F	RAIN	TAX
G	1.00													
EQ	0.31	1.00												
	(0.00)													
EXP-S	0.09	-0.04	1.00											
	(0.00)	(0.00)												
EXP-A	0.06	-0.02	0.89	1.00										
	(0.00)	(0.06)	(0.00)											
FEM-S	-0.16	-0.10	-0.11	-0.14	1.00									
	(0.00)	(0.00)	(0.00)	(0.00)										
FEM-O	-0.11	-0.05	-0.02	-0.06	0.82	1.00								
	(0.00)	(0.00)	(0.05)	(0.00)	(0.00)									
FEM-M	-0.09	-0.09	-0.14	-0.11	0.44	-0.14	1.00							
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)								
P-S	-0.19	-0.12	0.08	0.09	-0.12	-0.04	-0.11	1.00						
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)							
SIZE	-0.02	-0.05	-0.15	0.07	-0.05	-0.22	0.33	0.01	1.00					
	(0.09)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.19)						
INV-S	0.85	0.21	0.12	0.06	-0.13	-0.06	-0.14	-0.30	-0.07	1.00				
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)					
OTH-N	-0.00	-0.17	-0.03	-0.01	-0.01	-0.09	0.06	-0.00	0.10	0.04	1.00			
	(0.72)	(0.00)	(0.00)	(0.40)	(0.18)	(0.00)	(0.00)	(0.79)	(0.00)	(0.00)				
OTH-F	0.08	-0.06	-0.02	-0.01	0.03	-0.01	0.06	-0.01	0.04	0.01	0.13	1.00		
	(0.00)	(0.00)	(0.03)	(0.23)	(0.00)	(0.50)	(0.00)	(0.33)	(0.00)	(0.31)	(0.00)			
RAIN	-0.04	0.01	0.00	-0.00	-0.02	-0.01	-0.03	0.00	-0.01	-0.02	0.04	-0.07	1.00	
	(0.00)	(0.28)	(0.78)	(0.95)	(0.04)	(0.54)	(0.00)	(0.87)	(0.18)	(0.05)	(0.00)	(0.00)		
TAX	0.07	0.05	-0.07	-0.07	-0.02	-0.00	-0.02	0.06	0.01	-0.01	-0.14	0.13	-0.05	1.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.16)	(0.75)	(0.07)	(0.00)	(0.57)	(0.50)	(0.00)	(0.00)	(0.00)	

5.3 Empirical results

Our empirical analysis is comprised of four parts. First, we examine how founder team characteristics relate to fundraising strategies. Next, we examine how founder team characteristics and fundraising strategies affect the likelihood of campaign success. Third, we examine the fundraising outcomes. Lastly, we also analyze founder's choice to stop their campaigns, by modeling it simultaneously with daily investment flows.

5.3.1 Fundraising strategies

We start off by modeling the relationship between founder team characteristics X and fundraising strategies G, which include the campaign goal, pre-money valuation and equity offered at goal. Note that entrepreneurs are advised to set company valuation, equity offered, and funding goal to be realistic and comparable to other similar companies. This is the advice from CrowdCube (2018), and similar advice is given by SEEDRS. We can informally think of entrepreneurs setting the goal and equity offered as the result of an optimization problem that is a function of how much funding they need and what they think they can get, and constrained to be realistic.

Recall that we have two types of measures for founder team characteristics X: shares and dummies. The continuous model regresses the fundraising strategy variables G on the share of founders that are experienced, and also on the share of founders that are female. The dummy model regresses G on one experience dummy (whether the team has at least one experienced founder) and two gender dummies (whether the team is female-only and mixed-gender). Both models control for team size and whether this is a repeat campaign. Both models also include industry and calendar fixed effects. They allow us to control for possible inter-industry and calendar time variation in campaign goals and start-up valuations.

Throughout the analysis, we assume that founder team characteristics (i.e., prior

experience and gender) are exogenous to the fundraising choices made. A natural limitation of the data is that we cannot observe what happens prior to companies listing on SEEDRS. As a consequence, we cannot exclude the possibility that team formation and fundraising strategies are jointly determined. That is, it is conceivable that founders make choices about their team with future fundraising strategy in mind. For example, an inexperienced founder could bring on board someone else with prior entrepreneurial experience, or an all-female team might decide to add a male founder. This kind of endogeneity of teams does not pose an econometric problem since the experience and gender variables are clearly pre-determined at the beginning of the campaign. However, it suggests a more nuanced interpretation of the results.

We estimate the model using OLS regressions and Table 5.7 reports the results. We find that experienced teams have higher campaign goals and company valuations even after controlling for gender composition. Teams with female founders have lower goals and company valuations, even after controlling for their experience. Experienced teams thus pursue more ambitious fundraising strategies while female teams pursue safer fundraising strategies. These differences cannot be explained by different industry sectors of the ventures, team size, calendar time effects, and a host of other team characteristics[5].

It is noteworthy that even though Table 5.7 represents three choices, namely campaign goals, pre-money valuations, and equity offered at goal, founders actually only choose two out of three, the third is mechanically given.[6] We also note that a higher goal and a higher pre-money valuation automatically translate into similar equity offered. Consequently, we find in the last column of Table 5.7 no significant

[5]In the robustness section we also add controls for ownership structure, education, and commitment to the venture as measured by part-time vs. full-time work. The results are not qualitatively different when these additional founder characteristics are also included in the regressions.

[6]Specifically, the post-money valuation at goal is mechanically determined by diving the goal amount by the equity ownership. This formula is the same as the one to be used to value bids in the famous Dragon's Den show, where participants also quote their desired investment amount and the fraction of equity they are willing to give for it.

coefficients for experience or gender.[7] For the remainder of the analysis, we focus on campaign goal and equity offered as the two main fundraising choice variables. The reason for choosing those two is that this combination has the lowest possible multicollinearity, as can be verified in Table 5.5.

Table 5.7: Determinants of initial fundraising strategies

This table reports OLS regressions of campaign goal, pre-money valuation and equity offered at goal on founder team characteristics, control variables, and fixed effects. Goal and Pre-money valuation are in natural logs. In the continuous model, team characteristics are included as shares (share of founders that are experienced, share of founders that are female). In the dummy model, these are included as dummies (any experienced founders plus two gender dummies, one for a female-only team and a mixed gender team). Experience variable captures founders of start-ups prior to the SEEDRS venture, who also experienced start-up success, defined as exit (IPO, acquisition), private investment, or business growth, as listed on founders' LinkedIn profiles. Team size and a dummy indicating whether this is a repeat SEEDRS campaign are included as control variables. Regressions also include fixed effects for the industry sector (Clothing and Home, E-Commerce and Marketing, Food and Drink, Games and Entertainment, Technology, Transport and Travel, with the Finance category as the base), and calendar fixed effects using 20 quarters from 2012:Q3 to 2017:Q1, with 2017:Q2 as the base. T-statistics are in parentheses and standard errors are clustered at the company-level to take into account repeat campaigns.

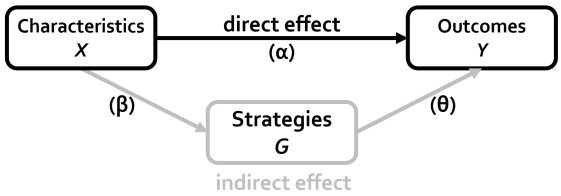
		Goal		Pre-money Valuation		Equity Offered	
		Continuous model	Dummy model	Continuous model	Dummy model	Continuous model	Dummy model
EXPERIENCE	(%, D)	0.441**	0.332**	0.344**	0.233**	0.805	0.635
		(3.25)	(3.29)	(2.50)	(2.24)	(1.16)	(1.19)
FEMALES	(%)	-0.390**		-0.148		-0.928	
		(-2.72)		(-1.31)		(-1.22)	
FEMALES ONLY	(D)		-0.256		-0.058		-0.808
			(-1.60)		(-0.47)		(-0.96)
FEMALES MIXED	(D)		-0.367**		-0.173*		-0.855
			(-2.68)		(-1.66)		(-1.31)
PRIOR SEEDRS		-0.829***	-0.809***	0.002	0.017	-3.772***	-3.733***
		(-7.78)	(-7.63)	(0.03)	(0.21)	(-7.55)	(-7.52)
TEAM SIZE		0.143***	0.167***	0.124**	0.131**	0.004	0.057
		(3.59)	(3.65)	(2.97)	(2.71)	(0.02)	(0.21)
N		823	823	823	823	823	823
Adjusted R-squared		0.238	0.240	0.281	0.279	0.107	0.107

[7]Given the simultaneity of choice, we also re-estimated SUR models with any two of the three equations. We found that the results were always very similar. We thus conclude that the correlation of residuals in these regression models is not a concern.

5.3.2 Campaign success

Let's now turn to the question of how founder team characteristics and fundraising strategies affect the likelihood of success. Here it is important to distinguish between the direct effect that founder team characteristics have on any fundraising outcome—success, funding amount raised and the extent of overfunding, controlling for the fundraising strategy—and the indirect effect. An indirect effect arises when team characteristics X determine campaign strategy, such as goal G, and a particular goal G in turn influences the outcome of interest Y. This indirect effect, for example, is the influence of experience on success via experienced team's choice of a higher goal. After controlling for the fact that experienced teams set higher goals, what remains is the direct effect of experience on success. This decomposition of total effect into direct effect and indirect effect is visually illustrated in Figure 5.2. Chapter 4 of Hayes (2014) contains a detailed explanation of this analysis.[8] We associate the indirect effect with the strategic choices of the entrepreneur, and the remaining direct effect with investor preferences. It should however be noted that strategic choices of entrepreneurs themselves are naturally also influenced by their expectations of what interest their project will garner amongst investors.

Figure 5.2: Simple model of mediation



$$\text{TOTAL EFFECT} = \alpha + \beta \times \theta = \delta$$

[8]The typical method to distinguish between direct and indirect relationships is to run regressions with and without the mediator variable G. This amounts to the basic idea in what is formally called mediation analysis in the statistics literature. In our analysis, we follow Baron and Kenny (1986) and other scholars to examine such mediation effects. The total effects is a coefficient δ in regression $Y = \delta_c + \delta X + \epsilon_x$. The direct effect is the coefficient α in regression $Y = \alpha_c + \alpha X + \alpha_g G + \epsilon_{x,g}$. If all the relationships are linear, the indirect effect can be estimated as the product $\beta \times \theta$, where β and θ are the coefficients from two regressions, $G = \beta_c + \beta X + \psi_x$ and $Y = \theta_c + \theta G + \epsilon_g$. Thus the indirect effect is $\delta - \alpha$, i.e., the difference between the total effect and the direct effect.

The prior crowdfunding literature shows that early momentum can have an important effect on eventual campaign success (see for example Vulkan et al., 2016; Mollick, 2014; Åstebro et al., 2017). Early campaign momentum can be affected by random variation in weather, by influencing internet use and investment activity (see Cardona et al., 2013; Gilchrist and Sands, 2016; Xu, 2017). We expect that the effect of weather on investor's investment activity is non-linear (cold enough to stay at home in front of the computer, but not too cold that you want to pursue winter leisure activities outdoors) and include weather temperature variables as dummies. Additional factors like competition on the platform may also affect momentum at the beginning of the campaign. Hence, we model the likelihood of success both as a function of entrepreneur characteristics, fundraising strategies, various controls as before, and the following variables that proxy for early campaign momentum:

1. average number of other open campaigns during the first week of each campaign
2. average daily flows to other open campaigns, as a natural log, during the first week of each campaign
3. average daily rainfall, as a natural log, during the first week of each campaign

Table 5.8 reports the total and direct effects of experience and gender on campaign success, as well as these additional easly-stage momentum proxies. We find that both the total and the direct effect of experience are positive and statistically significant. This is in line with the literature that concludes that entrepreneurial success breeds more success, and there is a positive experience premium. Going further, we find that total effect of experience on success is smaller than the direct effect. The indirect effect of experience on success, via experienced team's ambitious choice of a higher goal is negative. That is because in general, a higher goal corresponds to lower campaign success, hence the fundraising strategy of setting a higher goal is ambitious because it corresponds to a greater risk of campaign failure. The indirect effect of experience on

success via equity offered at goal is negligible since the coefficient on equity in Table 5.8 is nearly zero, hence so is the indirect effect.

We thus find that the direct success premium that comes from experience outweighs the risk of campaign failure due to a higher goal chosen by the experienced team since the total effect of experience on the probability of success is positive and statistically significant. Experienced founders not only pursue more ambitious fundraising strategies but are also more likely to be successful even after considering the fact that their higher goals hurt their chances of campaign success. For experienced founders, the risk of pursuing a more ambitious strategy pays off.

The situation for female teams is different. They pursue safer fundraising strategies by setting lower goals, and lower goals in general are correlated with a greater likelihood of success. However, this effect of a safer fundraising strategy is too small to make a difference. The total and direct effects of female gender, as seen in Table 5.8, are not statistically significantly different from zero. Female teams pursue safer fundraising strategies, but it does not translate into a significantly greater likelihood of success, although it helps to keep success probabilities up.

Running a repeat campaign on SEEDRS is naturally associated with a statistically significantly higher probability of success. Competition on the SEEDRS platform and weather both predict cross-sectional variation in funding flows and the subsequent campaign success. The result is statistically significant at the 5% level for rainfall and average flow to other open campaigns in all specifications: an increase in either phenomenon during the first week of the campaign reduces the likelihood of eventual campaign success.[9]

[9]Although the temperature dummies, the number of competing campaigns, and the end-of-tax-year effect are not significant in every specification, the joint test of coefficients on selection variables rejects the null hypothesis that all selection variables simultaneously have zero coefficients, with p-values of 0.057-0.071.

5.3.3 Fundraising amounts

Besides campaign success, another key outcome of equity crowdfunding is how much money the founder teams raise. One econometric issue to consider is that outcomes like funding raised, or the funding-to-goal ratio, are only observed for successful campaigns. If the campaign is unsuccessful because it is unable to reach the campaign goal within 60 days, founders get no money. If the probability of success is a function of team characteristics X, the sample of successful campaigns is not representative of the founder population. Then regressing amounts raised on team characteristics in a censored sample of successful campaigns might produce biased estimates due to this selection bias.

Specifically, we use the two-step Heckman (1979) model to correct for potential selection bias, which relies on the notion that this is a form of omitted variable bias. We formulate a model for the probability of having a successful campaign, which is a function of both founder characteristics as well as other selection variables, i.e., a function of $Z = X, S$. This first stage captures the probability of being in the sample of successful campaigns. In the second stage, we estimate our variables of interest Y in the sample of successful campaigns and include an additional term that represents the probability of being in the sample given founder characteristics and other selection variables. This term, called the Inverse Mills Ratio, thus corrects for selection bias.[10]

The identification of unbiased coefficients under this Heckman procedure relies on the non-linearity in the Inverse Mills Ratio that comes from having at least one

[10]The specification is:

$$\begin{aligned} Pr(\text{Success} = 1|Z) &= \Phi(Z\gamma) + \epsilon \\ (Y|X, \text{Success} = 1) &= X\beta + (\rho\sigma)\lambda(Z\hat{\gamma}) + \psi \end{aligned} \tag{5.3.3.1}$$

where λ is the Inverse mills ratio evaluated at $Z\hat{\gamma}$ and $\rho\sigma$ is it's coefficient. Here ρ is the correlation between unobserved determinants of campaign success and unobserved determinants of variable of interest Y and σ is the standard deviation of the unoberserved determinants of variable of interest Y. The coefficient on the Inverse mills ratio $\rho\sigma$ indicates whether selection bias is present when tested against the null of zero. The coefficients β on explanatory variables of interest X are unbiased, such as those in the model $Y = X\beta + \psi$, which is unconditional on success, and this is the relationship of interest.

selection variable S included in the first stage but not the second. This selection variable S should predict the probability of success with a coefficient that is statistically significantly different from zero but should not directly explain Y except through the probability of success. We use competition on the SEEDRS platform and weather during the first week of a campaign as selection variables for eventual campaign success, in line with the evidence that early momentum can have an important effect on campaign success (see for example Vulkan et al., 2016; Mollick, 2014; Åstebro et al., 2017). The argument is that the competition on the SEEDRS platform and weather during campaign start does not influence eventual funding amount raised except through the probability of success.

Thus, we estimate a Heckman two-stage model, to account for the effect that investment amounts are only observable conditional on campaign success. We already estimated the first step regression in Table 5.8. The second stage of the two-step Heckman selection model for total funding raised is reported in Table 5.9. The dependent variable is the investment amount raised, transformed with a natural log.

A first important finding is that the coefficient on campaign goal is highly significant, and in fact not statistically different from 1. This suggests that there is a one-to-one relationship between campaign goal and investment raised for successful campaigns: "be careful what you ask for, as you might get it!"

Next, we find that the total effect of experience is positive: teams with at least one experienced founder raise 53% more than inexperienced teams.[11] However, after controlling for the fundraising goal (and the equity offered), the direct effect is that experienced teams raise 10% less than inexperienced teams. This suggests that the indirect effect is very different from the direct effect. The total effect includes the initial choice of a more ambitious fundraising strategy. This indirect effect pays off for experienced teams. The remaining direct effect is negative but makes relatively

[11]Regression output is $ln(Funding) = \alpha + \hat{\beta}X + E$, thus we use $e^{\hat{\beta}}$ when interpreting coefficients.

little difference.

Turning to the gender variable, we find that for the total effect all-female teams raise 45% less than all-male teams. When we decompose this total effect, both the direct and indirect effect of gender is negative. The indirect effect comes from the lower campaign goals chosen by all-female teams, which account for 29% of the effect. After controlling for fundraising strategy, the direct effect is that all-female teams still raise 16% lower amounts. This suggests that all-female teams choosing a safer fundraising strategy is one part of what explains why female teams raise less money when compared to their all-male counterparts that are operating in the same sector, calendar time, and are similar in their experience and other characteristics. The other part of the explanation is the direct effect, which suggests that investors provide less funding to all-female teams, even after controlling for all these other factors, including the choices of funding goals and equity offered.

We have similar but not identical findings for mixed-gender teams. They raise 38% less than all-male teams. However, for them the direct effect is positive: after controlling for fundraising strategy, mixed-gender teams raise 11% more than all-male counterparts. Yet it is the indirect effect of a much lower goal that accounts for why mixed-gender teams raise less than all-male teams. That is, the indirect effect accounts for 49% less funding raised by mixed gender teams.

We also consider the likelihood that teams raise much more than the original goal. For this one can use many thresholds, and we focus here on whether companies go more or less than 25% over goal. 40% of all successful campaigns pass this threshold of overfunding, suggesting this is not a rare event. Moreover, the results are qualitatively similar for other funding-to-goal thresholds such as 115% and 150%. The second stage of the two-step Heckman selection model for the likelihood of overfunding is reported in Table 5.10. We find that there is again a positive and statistically significant relationship between campaign goal and likelihood to overfund for successful

campaigns.

The total and direct effect of experience on the likelihood to raise more than 125% of the goal is negative and statistically significantly different at 5% level. As seen from Table 5.9, after controlling for fundraising strategy, experienced teams raise 10% less funding than inexperienced teams. They also set much higher goals than inexperienced teams. These two factors together explain why experienced teams have lower funding-to-goal ratios than inexperienced teams, both overall and after controlling for fundraising strategy.

All-female teams are also less likely to raise more than 125% of goal, but this is primarily due to the fact that all-female teams raise significantly lower amounts of money than their male counterparts.[12] Mixed gender teams, on the other hand, are more likely to exceed 125% of their goal target as compared to all-male teams. This is due to their choice of a significantly lower campaign goal, as seen in Table 5.7, which translates into higher funding-to-goal ratios, together with a positive direct effect of mixed gender on fundraising amount as seen in Table 5.9.

Finally, note that controlling for selection is somewhat relevant in these regressions. The coefficient on the Inverse Mills Ratio is sometimes borderline insignificant or else statistically significant at 10% level, both in Tables 5.9 and 5.10. Hence the use of a two-step Heckman correction is justified in this analysis of cross-sectional variation.

[12] All-female teams have lower goals, which mechanically translates into higher funding-to-goal ratios, but their goals are not nearly as much lower as for mixed-gender teams (Table 5.7). The direct negative effect of all-female teams on funding raised, from Table 5.9, is the main reason why all-female teams are less likely to overfund above 125% of goal.

Table 5.8: Team Characteristics and Campaign Success

This table reports Probit regressions of campaign success on founder team characteristics, control variables, fixed effects, and selection variables (competition on the SEEDRS platform and weather variables). In the continuous model, team characteristics are included as shares (share of founders that are experienced, share of founders that are female). In the dummy model, these are included as dummies (any experienced founders plus two gender dummies, one for a female-only team and a mixed gender team). Experience variable captures founders of start-ups prior to the SEEDRS venture, who also experienced start-up success, defined as exit (IPO, acquisition), private investment, or business growth, as listed on founders' LinkedIn profiles. Team size and a dummy indicating whether this is a repeat SEEDRS campaign are included as control variables. Regressions also include fixed effects for the industry sector (Clothing and Home, E-Commerce and Marketing, Food and Drink, Games and Entertainment, Technology, Transport and Travel, with the Finance category as the base), and calendar fixed effects using 20 quarters from 2012:Q3 to 2017:Q1, with 2017:Q2 as the base. T-statistics are in parentheses and standard errors are clustered at the company-level to take into account repeat campaigns.

		Campaign Success			
		Total effect		Direct effect	
		Continuous model	Dummy model	Continuous model	Dummy model
EXPERIENCE	(%, D)	0.373**	0.240**	0.417**	0.273**
		(2.30)	(1.96)	(2.56)	(2.22)
FEMALES	(%)	-0.080		-0.121	
		(-0.48)		(-0.48)	
FEMALES ONLY	(D)		-0.018		-0.040
			(-0.10)		(-0.22)
FEMALES MIXED	(D)		-0.097		-0.142
			(-0.66)		(-0.95)
GOAL				-0.105**	-0.104**
				(-2.03)	(-2.01)
EQUITY OFFERED				-0.000	-0.000
				(-0.06)	(-0.05)
PRIOR SEEDRS		1.104***	1.117***	1.027***	1.044***
		(9.43)	(9.42)	(8.14)	(8.15)
TEAM SIZE		0.069	0.066	0.085	0.085
		(1.12)	(0.97)	(1.37)	(1.24)
COMPETITION NO		0.006	0.006	0.006	0.006
		(0.51)	(0.48)	(0.54)	(0.50)
COMPETITION FLOW		-0.156**	-0.152**	-0.173**	-0.167**
		(-2.19)	(-2.12)	(-2.38)	(-2.30)
RAIN		-0.088**	-0.089**	-0.087**	-0.088**
		(-2.08)	(-2.08)	(-2.01)	(-2.03)
TEMP: <5C		-0.015	-0.002	-0.022	-0.010
		(-0.06)	(-0.01)	(-0.09)	(-0.04)
TEMP: 5C to 10C		0.165	0.172	0.184	0.192
		(1.04)	(1.09)	(1.15)	(1.21)
TEMP: 15C to 20C		0.061	0.075	0.060	0.075
		(0.36)	(0.44)	(0.35)	(0.44)
TEMP: >20C		0.507*	0.514*	0.470	0.480
		(1.69)	(1.71)	(1.56)	(1.58)
N		823	823	823	823
Pseudo R-squared		0.181	0.179	0.185	0.183

Table 5.9: Team Characteristics and Funding

This table reports OLS regressions of funding raised on founder team characteristics, control variables, fixed effects, and a selection term which represents the probability of being in the sample of successful campaigns and corrects for selection bias via the Heckman (1979) procedure. Competition on the SEEDRS platform and weather variables are included as selection variables. In the continuous model, team characteristics are included as shares (share of founders that are experienced, share of founders that are female). In the dummy model, these are included as dummies (any experienced founders plus two gender dummies, one for a female-only team and a mixed gender team). Experience variable captures founders of start-ups prior to the SEEDRS venture, who also experienced start-up success, defined as exit (IPO, acquisition), private investment, or business growth, as listed on founders' LinkedIn profiles. Team size and a dummy indicating whether this is a repeat SEEDRS campaign are included as control variables. Regressions also include fixed effects for the industry sector (Clothing and Home, E-Commerce and Marketing, Food and Drink, Games and Entertainment, Technology, Transport and Travel, with the Finance category as the base), and calendar fixed effects using 20 quarters from 2012:Q3 to 2017:Q1, with 2017:Q2 as the base. T-statistics are in parentheses based on Heckman (1979) standard errors.

		Investment				
			Total effect		Direct effect	
		Base model	Continuous model	Dummy model	Continuous model	Dummy model
EXPERIENCE	(%, D)		0.532**	0.423**	-0.126*	-0.104**
			(2.25)	(2.39)	(-1.89)	(-2.13)
FEMALES	(%)		-0.728**		-0.065	
			(-3.17)		(-1.02)	
FEMALES ONLY	(D)			-0.597**		-0.171**
				(-2.35)		(-2.53)
FEMALES MIXED	(D)			-0.476**		0.105*
				(-2.19)		(1.84)
GOAL		1.040***			1.045***	1.048***
		(60.39)			(55.17)	(56.46)
EQUITY OFFERED		-0.006**			-0.006**	-0.006**
		(-2.25)			(-2.21)	(-2.18)
PRIOR SEEDRS		-0.127	-1.051**	-1.043**	-0.143	-0.132
		(-1.20)	(-2.56)	(-2.50)	(-1.41)	(-1.30)
TEAM SIZE		-0.029	0.076	0.099	-0.040*	-0.065**
		(-1.32)	(0.89)	(1.05)	(-1.74)	(-2.59)
IMR lambda		-0.240	-0.149	-0.162	-0.279*	-0.264*
		(-1.45)	(-0.24)	(-0.25)	(-1.73)	(-1.64)
N total		823	823	823	823	823
N successful		339	339	339	339	339

Table 5.10: Team Characteristics and Overfunding above 125% of Goal

This table reports Probit regressions of the likelihood of raising more than 125% of goal on founder team characteristics, control variables, fixed effects, and a selection term which represents the probability of being in the sample of successful campaigns and corrects for selection bias via the two-step Heckman (1979) procedure. Competition on the SEEDRS platform and weather variables are included as selection variables. In the continuous model, team characteristics are included as shares (share of founders that are experienced, share of founders that are female). In the dummy model, these are included as dummies (any experienced founders plus two gender dummies, one for a female-only team and a mixed gender team). Experience variable captures founders of start-ups prior to the SEEDRS venture, who also experienced start-up success, defined as exit (IPO, acquisition), private investment, or business growth, as listed on founders' LinkedIn profiles. Team size and a dummy indicating whether this is a repeat SEEDRS campaign are included as control variables. Regressions also include fixed effects for the industry sector (Clothing and Home, E-Commerce and Marketing, Food and Drink, Games and Entertainment, Technology, Transport and Travel, with the Finance category as the base), and calendar fixed effects using 20 quarters from 2012:Q3 to 2017:Q1, with 2017:Q2 as the base. T-statistics are in parentheses based on two-step Heckman (1979) standard errors.

		Overfunding above 125% of Goal				
			Total effect		Direct effect	
		Base model	Continuous model	Dummy model	Continuous model	Dummy model
EXPERIENCE	(%, D)		-0.169*	-0.166**	-0.216**	-0.208**
			(-1.78)	(-2.39)	(-2.16)	(-2.84)
FEMALES	(%)		-0.205**		-0.165*	
			(-2.22)		(-1.74)	
FEMALES ONLY	(D)			-0.394***		-0.365***
				(-3.96)		(-3.62)
FEMALES MIXED	(D)			0.165**		0.205**
				(1.98)		(2.39)
GOAL		0.066**			0.070**	0.080**
		(2.48)			(2.47)	(2.87)
EQUITY OFFERED		-0.002			-0.002	-0.002
		(-0.56)			(-0.56)	(-0.49)
PRIOR SEEDRS		-0.219	-0.254	-0.249	-0.224	-0.210
		(-1.35)	(-1.56)	(-1.55)	(-1.47)	(-1.39)
TEAM SIZE		-0.021	-0.032	-0.082**	-0.042	-0.095**
		(-0.64)	(-0.94)	(-2.23)	(-1.20)	(-2.53)
IMR lambda		-0.410	-0.393	-0.386	-0.434*	-0.427*
		(-1.63)	(-1.60)	(-1.59)	(-1.79)	(-1.79)
N total		823	823	823	823	823
N successful		339	339	339	339	339

5.3.4 Stopping the campaign

The next question that needs to be addressed is the one of causality. The amount of funding that teams raise after the campaign is successful is a function of both investors' demand for shares in the company, and founders' supply of shares. When we observe that experienced teams receive more funding and female teams receive less funding, we don't know whether this is due to entrepreneurs taking less, or investors giving less. In this section we disentangle the role of founders' choices from investors' choices, by modeling simultaneously the founder's decision to stop the campaign, and the funding flows by investors.

We model these two decisions by separately estimating investors' demand for shares, and the entrepreneurs' decisions to stop the supply of shares. Starting from day 8 after campaign reaches the target goal, i.e., after the cooling off period, the founders face a decision whether to stop or not, $STOP_{t+1}$. This decision is a function of their time-invariant characteristics X and the continued investor pledges that founders observe, I_t. We model founders' stopping choice using a linear probability model:

$$STOP_{i,t+1} = f(X_i, I_{i,t}) + e_{i,t} \tag{5.3.4.1}$$

Investors' demand for shares in the company, investment flow I_t, is in turn a function of the same time-invariant founder team characteristics X and some time-varying variables S_t, hence the continued investor pledges to each campaign are:

$$I_{i,t} = g(X_i, S_t) + u_{i,t} \tag{5.3.4.2}$$

where $i = 1, ..., N$ are individual campaigns and $t = 8, ..., T$ are days from hitting the target.

There may be reasons to suspect endogeneity in 5.3.4.1 in the sense that $E[I_{i,t}e_{i,t}] \neq 0$ in this panel setting. Endogeneity can arise due to simultaneity bias, since $STOP$

and I are jointly determined, at the same time.

We proceed by using the instrumental variable technique and the two-stage GLS (random effects) estimator. We include daily weather and competition on the platform as candidate instruments S_t. We also include a daily dummy with respect to the last three weeks of the UK tax year to capture investors' incentives to fully use up the tax break allowance at the end of each tax year. The instruments need to explain investor' demand for shares ($E[S_t I_{i,t}] \neq 0$) but not the founders' stopping decision directly, ($E[S_t e_{i,t+1}] = 0$), i.e., the instruments may affect the founders' stopping decision only through the instrumented investment flow variable $I_{i,t}$. We use the standard test for instrument informativeness in the first stage. Since the number of instruments exceeds the number of endogenous variables and our IV model is overidentified, we are also able to test whether the excluded instruments are independent of the error process, and thus valid.

We are only using the successful campaigns sample, so the concern from section 3.2 and 3.3 about selection bias remains present too. We thus need to consider the possibility that the estimates in the equation for founders' optimal stopping decision and in the equation of investors' demand are biased due to selection effects. To estimate a model with both selection effects and a system of endogenous equations we follow the procedure suggested in Woodridge (2002, p.568). He suggests to first correct for selection bias by including the inverse Mills ratio in both the endogenous investment equation 5.3.4.2 and the stopping equation 5.3.4.1. Thus, it is a system of three equations, as below.

$$\begin{aligned} Pr(\text{Success}_i = 1|Z_i) &= \Phi(Z_i\gamma) + \epsilon_i, \ \forall i \\ (I_{i,t}|X_i, S_t, \text{Success}_i = 1) &= X_i\pi + S_t\phi + t\delta + (\rho\sigma)\lambda(Z_i\hat{\gamma}) + u_{i,t}, \text{ for } t \geq 8 \\ (STOP_{i,t+1}|X_i, \hat{I_{i,t}}, \text{Success}_i = 1) &= X_i\beta + \alpha\hat{I_{i,t}} + t\delta + (\rho\sigma)\lambda(Z_i\hat{\gamma}) + e_{i,t}, \text{ for } t \geq 8 \end{aligned} \tag{5.3.4.3}$$

where $i = 1, ..., N$ are individual campaigns and $t = 8, ..., T$ are days from hitting

the target.

The null hypothesis of no selection effects can be tested using standard t-tests for the coefficients of the Inverse Mills ratio. Under the null of no selection, the standard errors are correct (Woodridge, 2002, p.568). We find that selection effects are not a concern in this particular system of equations. Specifically, the estimated coefficient $\hat{\rho\sigma}$ on the Inverse Mills Ratio $\lambda(Z_i\hat{\gamma})$ is not statistically different from zero. Therefore, in Tables 5.11 and 5.12 we present the results from instrumental variable regressions without the selection correction. The point estimates and qualitative results are very similar with and without the selection correction.

Table 11 shows the base first stage model for investment flows which represent investors' demand for equity. The instrumental variables and selection variables used in this paper measure the same phenomena such as weather, competition and tax incentives. The only difference is that before we measured these variables during the first week of the campaign (capturing cross-sectional variation), whereas now we use the daily variables in panel regression (also capturing time-series variation). Competition on the SEEDRS platform, weather, and tax incentives all predict the time-series variation in funding flows coming from investors. Many other competing campaigns are associated with lower investment flows, while high average flow to other competing campaigns (after controlling for their number) is associated with higher investment flows, and all these results are statistically significant at the 10% level. Furthermore, successful campaigns receive higher investment flows during three weeks in the run-up to the end of the UK tax year, significant at the 5% level. This in line with the hypothesis that investors rush to use up the tax incentives allowance limit as the tax deadline approaches, and this effect is particularly strong for campaigns that have already reached their target goal and are successful.

Note that we also include the investment stock on the last day of the cooling off period as a control, which captures additional variation in investment flows that

happened during the cooling off period. Similar to the model of funding in Table 5.9, a higher campaign goal is positively related to investment flows. There is a high correlation between campaign goal and investment stock on the last day of the cooling off period, so we cannot reliably comment on the significance of these two variables. This multicollinearity problem, however, does not affect the inference on the variables of interest—founder team characteristics.

The two-stage generalized least squares model in Table 5.12 does not show any presence of misspecification. Instrumental variables used are strongly associated with investment flows based on the Cragg-Donald Wald F-test for informativeness. Also, the instrumental variables are valid based on non-rejection of the null hypothesis in the Sargan-Hansen overidentification test, so we have no concerns that instrumental variables may be correlated with the residual term in the second stage. The endogeneity problem is addressed. Thus, we interpret the two equations as investors' choices (first stage) and founders' choices (second stage).

When we examine the effect of founder team characteristics in Table 5.12, we find that experienced teams neither receive significantly lower investment flows after the cooling off period nor are they more likely to stop their campaign. Female teams, and especially all-female founder teams, do receive significantly lower investment flows after the cooling off period, and they are less likely to stop their campaigns. This represents the direct effect of gender in response to exogenous variation in investment flows, after controlling for fundraising strategy, variation in post-success funding flows during the cooling off period, team size, repeat campaign, sector of the venture, calendar time effects, and the time trend.

The main result from these instrumental variable regressions which disentangle founders' stopping choice from investors' choices is that female teams hold out for more investment by keeping their campaign open longer. Experienced teams do not. Female teams may be holding out for more investment to compensate for lower invest-

ment flows after reaching the campaign target goal, whereas the experienced teams do not need to compensate in this manner. This result matters in the light of the findings from Table 5.7 which showed that all-female teams ask for less money. At the end of the campaign is thus appears that they try to partly overcome their lower ask by holding out for longer before closing their campaigns. However, since the investor interest is lower too, as shown in the first stage regressions, the longer wait does not result in larger investment amounts (as shown in Table 5.9) or even more overfunding (as shown in Table 5.10).

Table 5.11: Base model for Investment Flow once Can Stop

This table reports GLS (random effects) panel regression of daily investment flows on instrumental variables, fundraising strategies, control variables, fixed effects. Competition on the SEEDRS platform, weather, and year-end tax break incentives are included as instrumental variables for investment flow. Investment stock on the last day of the cooling off period and time trend are included as controls, in addition to team size and a repeat campaign dummy. Regressions also include fixed effects for the industry sector (Clothing and Home, E-Commerce and Marketing, Food and Drink, Games and Entertainment, Technology, Transport and Travel, with the Finance category as the base), and calendar fixed effects using 20 quarters from 2012:Q3 to 2017:Q1, with 2017:Q2 as the base. T-statistics are in parentheses and are clustered at company level to take into account repeat campaigns.

	Flow		
	Base model (1st stage)		
GOAL	0.376*		0.396*
	(1.70)		(1.79)
EQUITY OFFERED	0.007		0.006
	(0.44)		(0.39)
COMPETITION NO		-0.031*	-0.031*
		(-1.89)	(-1.90)
COMPETITION FLOW		0.073*	0.073*
		(1.80)	(1.80)
RAIN		0.002	0.003
		(0.10)	(0.11)
TEMP: <5C		-0.253	-0.254
		(-1.31)	(-1.31)
TEMP: 5C to 10C		-0.077	-0.077
		(-0.66)	(-0.65)
TEMP: 15C to 20C		0.062	0.063
		(0.33)	(0.33)
TEMP: 20C to 25C		0.015	0.017
		(0.05)	(0.06)
TEMP: >25C		1.022	1.024
		(1.24)	(1.24)
TAX YEAR		0.667**	0.665**
		(2.21)	(2.20)
PRIOR SEEDRS	-0.047	-0.111	-0.093
	(-0.16)	(-0.39)	(-0.32)
TEAM SIZE	-0.169	-0.153	-0.164
	(-1.32)	(-1.21)	(-1.29)
STOCK	0.362*	0.696***	0.351*
	(1.78)	(7.95)	(1.73)
N	7892	7892	7892
N successful	327	327	327
Overall R-squared	0.198	0.197	0.199

Table 5.12: Team Characteristics, Flow Constraints and Stopping Choice

This table reports two-stage GLS (random effects with IV) panel regression of stopping dummy on instrumented lag daily investment flows, founder team characteristics, fundraising strategies, control variables, fixed effects. Competition on the SEEDRS platform, weather, and year-end tax break incentives are included as instrumental variables for investment flow. Investment stock on the last day of the cooling off period and time trend are included as controls, in addition to team size and a repeat campaign dummy. Regressions also include fixed effects for the industry sector (Clothing and Home, E-Commerce and Marketing, Food and Drink, Games and Entertainment, Technology, Transport and Travel, with the Finance category as the base), and calendar fixed effects using 20 quarters from 2012:Q3 to 2017:Q1, with 2017:Q2 as the base. Conventional T-statistics are in parentheses.

		Flow (1st stage) direct effect		Stop (2nd stage) direct effect	
		Continuous model	Dummy model	Continuous model	Dummy model
INSTRUMENTED FLOW				-0.050***	-0.050***
				(-3.98)	(-3.99)
EXPERIENCE	(%, D)	-0.085	-0.274	0.025	0.013
		(-0.22)	(-0.90)	(0.74)	(0.47)
FEMALES	(%)	-0.621		-0.072*	
		(-1.41)		(-1.86)	
FEMALES ONLY	(D)		-1.121**		-0.079*
			(-2.31)		(-1.80)
FEMALES MIXED	(D)		0.382		-0.014
			(0.93)		(-0.39)
GOAL		0.398	0.444	-0.102***	-0.100***
		(1.32)	(1.48)	(-3.85)	(-3.80)
EQUITY OFFERED		0.005	0.005	0.002	0.002
		(0.26)	(0.28)	(1.50)	(1.50)
PRIOR SEEDRS		-0.13	-0.081	0.021	0.023
		(-0.43)	(-0.27)	(0.79)	(0.91)
TEAM SIZE		-0.186	-0.336*	-0.009	-0.014
		(-1.14)	(-1.87)	(-0.66)	(-0.86)
STOCK		0.326	0.313	0.100***	0.100***
		(1.19)	(1.15)	(4.15)	(4.18)
N		7892	7892	7892	7892
N successful		327	327	327	327
Between R-squared				0.125	0.123
Sargan-Hansen overidentification test				0.247	0.246
Cragg-Donald Wald F informativeness test				0.000	0.000

5.4 Robustness and extensions

In this section, we consider many empirical robustness checks, models, and extensions that speak to the representativeness of the results presented in Section 3.

The analysis presented earlier includes all founder team characteristics jointly. Including multiple characteristics together allows us to separate effects. For example, we can separate the effect of gender from the effect of experience. As robustness, we check that the results are qualitatively similar when team characteristics are included one-by-one. As another robustness check, we considered whether multicollinearity of goal and equity were affecting our results and the interpretation of indirect effects through the choice of fundraising strategy. Results are qualitatively similar when the equity variable is replaced by the pre-money valuation variable, residual pre-money valuation variable (obtained by regressing pre-money valuation on goal) in line with the sequential regression approach to deal with multicollinearity, and also when the equity variable is excluded from the specification, which is the other approach to dealing with multicollinearity. See Dormann et al. (2013) for a review of these and other methods.

One concern is that there are confounding entrepreneur characteristics besides gender and experience. We have considered a host of variables that were available from LinkedIn profiles, including education (highest degree, MBA, relevant education in finance/econ/business, etc.). Founders with MBAs have higher valuations and goals. This translates into a higher total effect on investment, but no direct effect from MBA per se. Having controlled for start-up experience, an MBA did not have an effect on campaign success. Also, there is no significant correlation between having an MBA and being a founder with past start-up success or with gender. Hence the results of the analysis presented earlier are not driven by education of the founders.

We performed considerable robustness analysis on our measure of experience. Our base measure is based on whether founders have prior founding experiences. For

robustness, we further consider whether their prior founding experience was successful or not. Moreover, we also ask whether founder have prior experience working in a (successful or unsuccessful) start-up as a non-founder. This gives us a 2×2 matrix to work with, distinguishing successful vs. unsuccessful in one dimension, and founders vs. non-founders in the second dimension. In our base model, we presented results of being a founder AND having prior success. This left out founders with no identifiable success, and people that worked in a successful start-up. We find that adding either or both categories does not substantially change our main results.

We also considered the importance of founders having sector-specific experience but did not find that this mattered. Besides these human capital measures, we have also considered the ownership split of the firm, as well as the time commitment of the founders, proxied by whether they held another job during the SEEDRS campaign. We find that teams with majority owners had higher goals and higher valuations (controlling for experience, gender, and commitment), but we find no difference in terms of equity offered. This higher goal translated into a lower investment-to-goal ratio. Teams where founders were working part-time offered more equity and had lower company valuations. They were also less likely to be successful. Among the teams that did reach the goal target, teams with lower time commitment to the project stopped their campaigns sooner. This analysis of looking at additional characteristics adds potential extra dimensions when considering entrepreneur's fundraising strategies. However, for tractability of the paper, these additional dimensions are not part of the main analysis. This is because we want to focus on the effect of experience and gender, on which there is large prior literature. Our robustness analysis thus helps to dismiss doubts about confounding factors.

Another concern is that our controls are not sufficient for capturing capital intensity, even within the industry categories. We therefore consider additional data that measure the companies' business models (distinguishing B2B vs. B2C vs. mixed

models) and the companies' model of delivery (digital, non-digital, or mixed model). This data is incomplete for unsuccessful campaigns, so we can only use it for successful campaigns. Adding it as controls there does not affect any of our main results.

We also consider several interaction effects. One question concerns whether the role of the goal is different across gender categories, but we find that none of the interaction effects are significant. Another question concerns the interaction of gender and experience, but again we find that none of the interaction effects are significant.

Finally, there is the question of repeat campaigns. Our main analysis controls for having a past SEEDRS campaign (approximately 25% of the sample do) but makes no distinction whether they were successful or not. As robustness, we therefore look at the number of successful and unsuccessful campaigns of SEEDRS. We find that both types of previous campaigns positively predict current campaign success. However, there might be a problem of multicollinearity, as the correlation between these two variables is 0.28, suggesting caution in how to interpret coefficient estimates. Moreover, a more complete analysis of these dynamics is beyond the scope of this paper.

5.5 Conclusion

This paper uses the context of equity crowdfunding to address a fundamental question in entrepreneurial finance about the relationship between founder team characteristics, fundraising strategies, and fundraising outcomes. We leverage the fact that the entrepreneurs' choices are observable in crowdfunding campaigns, and therefore allow us to distinguish between what the entrepreneurs ask for, versus what the investors give. We find that experienced founder teams ask for more money, at higher valuations, are more likely to succeed in their campaigns, and end up raising more money. Female teams ask for less money, have the same probability of campaign success, but

end up raising less money. Interestingly, they also hold out for money longer before closing their campaigns. The analysis suggests that different founder teams pursue different fundraising strategies, but also face different levels of investor demand.

While we believe that our analysis breaks new ground in terms of leveraging the detailed data from equity crowdfunding platforms to gain insights into entrepreneurs' fundraising strategies, we also acknowledge that it has some limitations. An important one is that while we can observe fundraising strategies in the context of equity crowdfunding, we naturally cannot generalize these findings to other fundraising contexts, such as traditional venture capital or "off-line" angel investing. Closely related, we cannot address questions about how entrepreneurs chose between these alternative modes of financing. Finally, any study involving a cross-section of companies and founder teams raises the question of what other characteristics remain unmeasured. While our data already contains rich controls, it is inherently impossible to exclude the possibility that our experience and gender effects are still affected by unobservable correlates.

Our findings on the importance of fundraising strategies raise a host of new questions about the underlying reasons as to why different entrepreneurs chose different strategies. One important question for future research concerns the question of how entrepreneurs determine their fundraising goals. Of particular interest is why female founder teams ask for less money: is this mainly driven by expectations of what they think they can get (rightly or wrongly?), or do other factors also influence this behaviour?

Another interesting question concerns the dynamics across multiple fundraising events. Our current analysis focuses on individual fundraising campaigns as a unit of analysis. Future research might investigate how entrepreneurs dynamically adjust their fundraising strategies across multiple campaigns, and how they adjust based on the underlying business developments in their start-up companies. Finally, there is

always the question of how fundraising strategies are related to the ultimate success of the company at the time of exit.

Chapter 6

Conclusion

Each paper presented in this dissertation examines a distinct asset class and financial environment. Yet all three papers have in common the theme that securities and financial intermediaries allocate resources in an economy and that efficient allocation of resources is a desirable outcome. All three papers contribute to our understanding of the resource allocation process. The first paper does so by exploring the role of information in the market for public equity, by focusing on security analysts that issue investment recommendations. In this paper, I find that it may be possible to infer the analysts' information sets and argue that information diversity is a prevalent, economically significant, and tradable phenomenon. The second paper contributes to our understanding of how securities and intermediaries allocate resources by exploring the objectives of government when it comes to sovereign debt issuance. In this second paper, I find that governments in advanced economies both minimize the cost of servicing their national debts as well as potentially facilitate sharing of risk across generations. The third paper examines the equity crowdfunding market and how the fundraising strategies of startup founders interplay with investor preferences. I find that the initial fundraising strategies of entrepreneurs are particularly important in determining the eventual allocation of investment resources to different

start-ups. Taken together, these three studies contribute to our understanding of how intermediaries and securities allocate resources between consumers, investors, and the government in various financial settings. This final chapter summarizes the results of the three papers and provides their conclusions and policy implications.

6.1 Diversity of information in the market for public equity

In the first paper, I apply a recently developed statistical technique by Satopää, Jensen, Pemantle, and Ungar (2016) to combine investment recommendations of many sell-side security analysts into one aggregate recommendation. This allows me to examine the information content behind these recommendations. To better understand this aggregation technique, in one part of the paper I quantify and test it by explicitly modeling dissemination of information among agents, letting the agents predict outcomes based on information they have, and then applying the statistical technique to aggregate their heterogeneous predictions. I thus use a calibrated simulation to explore the usefulness and limitations of this method. I also illustrate how this method can infer differences in information across agents and how it can combine agents' heterogeneous predictions more efficiently than a simple average. Finally, I also confirm that this technique is well-suited to data properties and the environment of security analysts.

In another part of the paper, I provide the first application of this statistical method using financial data—stock recommendations of security analysts. I find a trading strategy which not only outperforms a consensus-based strategy by 3-4% on an annualized basis but also delivers an economically significant net abnormal return of 2-3%. This strategy is profitable after adjusting for its risk characteristics and after incorporating trading costs that arise through portfolio rebalancing.

These findings allow me to reject the null hypothesis that all security analysts are equally well-informed. Moreover, the finding of abnormal returns suggests that some of the differences in information among security analysts could reflect private information that some analysts have and others do not. The novel aggregation method infers different information sets of analysts, potentially enabling trading on private information. As such, this paper shows that analyst recommendations have valuable information content. Moreover, different analysts know different aspects of the total information that is relevant when predicting the future performance of the firm. In other words, there is a diversity of information among analysts.

This result directly contributes to the prior literature on security analysts and the seeming contrast between academic theory and Wall Street practice. On the one hand, according to the semi-strong market efficiency hypothesis, investors should not be able to trade profitably on publicly available information such as the analyst stock recommendations. On the other hand, research departments of brokerage houses continue employing many security analysts to this day. Profits from private information are theoretically possible, but insider trading is illegal. Furthermore, the literature is fairly conclusive that trading on the level of consensus recommendation scores of sell-side analysts is not profitable once transaction costs are taken into account (Barber and Odean, 2001; Jegadeesh et al., 2004) but there may be positive returns associated with changes in analyst recommendations (Stickel, 1995; Womack, 1996; Jegadeesh et al., 2004). The profitable trading strategy that I present in my paper only uses the contemporaneous level of security analyst recommendations (not upgrades or downgrades), which I aggregate differently from consensus, with a recently developed statistical technique that can infer differences in information. Therefore, the results of this paper shed new light on uncovering additional information content in analyst predictions by looking at the level of their recommendations alone.

More generally, this paper highlights the prevalence of information diversity as a

phenomenon. This paper shows how information diversity can be empirically measured and demonstrates the economic significance of this phenomenon by means of a profitable trading strategy developed for the public equity market. It is plausible that this kind of information diversity is prevalent in other settings in finance, such as mutual fund investments or betting markets, and this calls for future research. The analysis presented in this paper also supports the need for future development of statistical aggregation techniques. One extension would be to incorporate time-series forecasts, rather than contemporaneous forecasts, taking into account the cross-sectional dimension of these time-series forecasts. This could be applied to examine information content in the flows of mutual fund investors or investment choices of mutual fund managers, which are often dismissed as noise traders and lacking "skill".

Furthermore, the analysis in this paper speaks to the idea that disaggregated and crowdsourced forecasts are useful. Crowdsourced forecasts can include prediction markets, betting markets, as well as markets for online retail investors, i.e., where the "crowd" provides rewards- or equity-based funding. There is already growing literature that documents the phenomenon that crowd choices are comparable to choices of experts (Mollick and Nanda, 2016; Chalmers et al., 2013). This paper provides additional evidence in this regard by documenting the wisdom of crowds phenomenon in a finance setting.

The research agenda into the wisdom of crowds, such as this first paper, is focused on exploring the ways to extract and efficiently use all available information. As discussed in Chapter 1, information plays a key role in facilitating the allocation of resources, both through security prices and through intermediaries such as security analysts. In the second paper, I explore another aspect of the resource allocation problem. I turn to government and explore the objectives behind the choice of maturity structure when a government issues bonds.

6.2 Demographic clienteles in the market for long-term government bonds

There are several theoretical prescriptions regarding how the government should choose the maturity structure of the debt it issues. In this second paper, I first review these formal models and derive reduced-form specifications for the determinants of maturity structure of government debt. Next, I test these different reduced-form predictions as competing hypotheses. I ask what explains the cross-country variation in maturity structure of advanced economies during the 2000-2010 decade, and what explains the dynamics in the maturity structure of U.S. government debt in 1972-2013. The sample period is driven by data availability.

Three main findings emerge as a result of the analysis. The first finding is that debt managers choose maturity structure so as to minimize expected cost of debt service, which is in line with cost-risk portfolio optimization theory. In the example of the United States, this entails responding to term premium in the short-run. I find that the US Treasury dynamically issues more of the security that is relatively cheaper, controlling for the level of rollover risk. In cross-country terms, countries with a steeper long-run slope of the yield curve have shorter maturity structure, controlling for the volatility of interest rates and covariance between the short- and the long-term rates. The second finding is that demographic clienteles play an important role. In the U.S., there is a positive long-run equilibrium relationship between maturity structure and the share of prime savers in the population. Among developed economies, there is a positive association between the size of the working-age population and maturity structure. Moreover, debt managers respond to the growing assets of institutional investors (pension funds and insurance corporations), even after controlling for demographic changes. This suggests that debt managers cater to demographic clienteles when choosing the maturity structure of debt.

The clienteles theory prescribes some catering to demographic clienteles, but not enough to fully offset demand pressures (Guibaud et al., 2013). Imperfect catering allows the government to facilitate the sharing of risk between the generation of savers and the generation of retirees. Without government's choice of a particular maturity structure that facilitates this risk-sharing, newly issued debt imposes the risk of an uncertain future tax burden only on the generation of young savers, but not on the generation of old retirees. The third finding of my paper is that government debt managers do choose the maturity structure with intergenerational risk-sharing considerations in mind.

The two findings regarding the role of demographic clienteles raise many interesting questions for further academic inquiry and policy implications. For example, this finding is significant in the context of the current pension fund environment, now that many pension funds face funding pressures. One of the contributing factors is a prolonged period of low interest rates: discounting promised pension fund payments at lower interest rates translates into a higher present value of pension fund liabilities; this in turn requires a higher present value of assets; consequently, falling interest rates translate into widening pension fund deficits. This is particularly acute for defined benefit pension funds. Furthermore, in the aftermath of the financial crisis, pension funds seeking safe returns that have good liability-matching properties have been fuelling the demand for long-term government debt (Bird, 2016). As a result, sovereign debt managers have been facing downward pressure on long-term interest rates. With imperfect catering to increased pension fund demand, as prescribed by theory and what the debt managers appear to be doing, downward pressure on long-term interest rates persists, and long-term interest rates remain low.

Therefore, this leads to a negative feedback loop which exacerbates the problem of pension fund solvency. This negative feedback loop is due to a combination of three factors: the zero lower bound, high demand for long-term government bonds by

pensions funds, and a debt management policy that caters to such clientele demand imperfectly. The imperfect catering is indeed prescribed by the clienteles theory. However, what may be optimal from the perspective of enabling risk-sharing across generations may no longer be optimal once the expected welfare cost of potential pension fund bankruptcy is taken into account in the formal model. Future theoretical research should address the way in which demographic clienteles and stability of the institutional investor sector should influence debt manager's choice of maturity structure more generally. Governments issuing fixed-income securities have a role to play in the allocation of resources in the economy, and welfare-improving policies are desirable.

6.3 Startup founder choices in the crowdfunding market for private equity

Following the analysis of government debt issuance and intergenerational risk-sharing, the third paper examines a different asset class and a different resource allocation problem. In the markets that provide finance to early-stage companies, as discussed earlier in Chapter 1, a desirable outcome is one where startups that can use investment resources in the best way are the ones that receive it. In the third paper, I focus on the equity crowdfunding market, which is a fairly well-established market in the UK. The eventual allocation of resources in the crowdfunding market is a function of both investor preferences and the fundraising strategies of startup founders. In other words, demand and supply factors both matter. The third paper in this dissertation investigates the interplay between fundraising strategies, investor demand, and how each factor contributes to eventual funding outcomes.

Specifically, my co-authors and I examine the relationship between founder team characteristics, fundraising strategies, and fundraising outcomes. Our analysis lever-

ages the fact that the entrepreneurs' choices are observable in crowdfunding campaigns, and therefore allow us to distinguish between what the entrepreneurs ask for versus what the investors give. We find that experienced founder teams ask for more money, at higher valuations, are more likely to succeed in their campaigns, and end up raising more money. Female teams ask for less money, have the same probability of campaign success, but end up raising less money. After reaching the campaign target, female teams face lower investment flows and hold out for money longer before closing their campaigns. Thus we show that different founder teams pursue different fundraising strategies, but also face different levels of investor demand.

Our findings on the importance of fundraising strategies raise a host of new questions about the underlying reasons as to why different entrepreneurs chose different strategies. One important question for future research concerns the question of how entrepreneurs determine their fundraising goals. Of particular interest is why female founder teams ask for less money: is this mainly driven by expectations of what they think they can get (rightly or wrongly?), or do other factors also influence this behaviour? An investigation along these lines would yield important findings and policy implications, especially as it relates to promoting gender equality in finance.

Another interesting question that emerges from our current research concerns the dynamics of fundraising across multiple campaigns. Current analysis focuses on individual fundraising campaigns as a unit of analysis, and future research might investigate how entrepreneurs dynamically adjust their fundraising strategies across multiple campaigns, and how they adjust based on the underlying business developments in their start-up companies. Finally, there is always the question of how fundraising strategies are related to the ultimate success of the company at the time of exit. These questions are left for future research.

Taken together, these three studies presented in this thesis contribute to our understanding of how intermediaries and securities allocate resources in various financial

settings. The analyses contribute to many fields in finance, including empirical asset pricing, financial intermediation, and entrepreneurial finance. The analyses complement existing literature where a large body of research already exists and fill the gap where research is scarce. The results break new ground concerning the information content of analyst recommendations with a trading strategy that uses the level of publicly-available security recommendations but aggregates them in a novel way. The analysis of equity crowdfunding is also the first to leverage the dynamic nature of campaign data and disentangle investor demand from founder choices in this fundraising environment. Last but not least, all three papers raise many interesting questions and pave the way for future research.

Appendix A

Appendix: Information Diversity and Security Analysts

A.1 Proof of proposition

This directly relies on Jensen et al. (2016) online supplementary material, plus my own intermediary steps and justifications.

Denote the mean of the target random variable Y with $\mu_0 := \mathbb{E}(Y)$. It is assumed that $\mathcal{F}_j = \sigma(X_j)$ such that $\mathbb{E}(Y \mid \mathcal{F}_j) = \mathbb{E}(Y \mid X_j) = X_j$ for all $j = 1, ..., N$

i Since $X_j = \mathbb{E}(Y \mid X_j)$, then $\mathbb{E}(X_j) = \mathbb{E}(\mathbb{E}(Y \mid X_j))$.

By the law of iterated expectations, $\mathbb{E}(\mathbb{E}(Y \mid X_j)) = \mathbb{E}(Y)$.

Hence $\mathbb{E}(Y) = \mathbb{E}(X_j)$.

iii By definition, $Cov(X_j, X_i) = \mathbb{E}(X_j X_i) - \mathbb{E}(X_j)\mathbb{E}(X_i)$.

$Cov(X_j, X_i) = \mathbb{E}(X_j X_i) - \mu_0^2$

By the law of iterated expectations $\mathbb{E}(X_j) = \mathbb{E}(\mathbb{E}(X_j \mid X_i))$:

$= \mathbb{E}(\mathbb{E}(X_j \mid X_i)X_i) - \mu_0^2$

Since $X_j = \mathbb{E}(Y \mid X_j)$:

$= \mathbb{E}(\mathbb{E}(\mathbb{E}(Y \mid X_j) \mid X_i)X_i) - \mu_0^2 = \mathbb{E}(\mathbb{E}(\mathbb{E}(Y \mid \mathcal{F}_j) \mid \mathcal{F}_i)X_i) - \mu_0^2$

Since $\mathcal{F}_i \subseteq \mathcal{F}_j$, then $\sigma(X_i) \subseteq \sigma(X_j)$ and the smallest sigma-field wins

$= \mathbb{E}(\mathbb{E}(Y \mid \mathcal{F}_i)X_i) - \mu_0^2 = \mathbb{E}(\mathbb{E}(Y \mid X_i)X_i) - \mu_0^2$

Also assumed that $X_i = \mathbb{E}(Y \mid X_i)$:

$= \mathbb{E}(X_i X_i) - \mu_0^2 = \mathbb{E}(X_i^2) - \mu_0^2$

By definition, $Var(X_i) = \mathbb{E}((X_i - \mu_0)^2) = \mathbb{E}(X_i^2) - (\mathbb{E}(X_i))^2$:

$= Var(X_i)$

ii By item iii, $Var(X_i) = Cov(X_j, X_i)$ if $\mathcal{F}_i \subseteq \mathcal{F}_j$.

By definition, $Cov(X_j, X_i) = \mathbb{E}((X_i - \mu_0)(X_j - \mu_0))$.

By Cauchy-Schwarz inequality, $|\mathbb{E}(X_j X_i)|^2 \leq \mathbb{E}(X_j^2)\mathbb{E}(X_i^2)$:

$$\mathbb{E}((X_i - \mu_0)(X_j - \mu_0)) \leq \sqrt{\mathbb{E}((X_i - \mu_0)^2)\mathbb{E}((X_j - \mu_0)^2)}$$

All together, if $\mathcal{F}_i \subseteq \mathcal{F}_j$:

$$\begin{aligned} Var(X_i) &\leq \sqrt{Var(X_i)Var(X_j)} \\ \sqrt{Var(X_i)Var(X_i)} &\leq \sqrt{Var(X_i)Var(X_j)} \\ Var(X_i) &\leq Var(X_j) \end{aligned}$$

This inequality is tight because for $\mathcal{F}_i = \mathcal{F}_j$,

have $X_i = X_j$ and $Var(X_i) = Var(X_j)$.

A.2 Derivation of implied weights

Overall, the revealed aggregator $X'' = \mathbb{E}(Y \mid \mathcal{F}'')$ for particular stock k is:

$$X''_k = \sigma_0 diag(\Sigma_{cov})' \Sigma_{cov}^{-1} Z_k + \frac{X'_k \Sigma_{cov}^{-1} 1_N}{1'_N \Sigma_{cov}^{-1} 1_N}$$

where:

$Z_k = [Z_{1k}, ..., Z_{Nk}]'$ is a vector such that

$$Z_{jk} = \frac{1}{\hat{\sigma}_0}(X_{jk} - \hat{\mu}_{0k}) = \frac{1}{\hat{\sigma}_0}(X_{jk} - \frac{1}{N}\sum_{j=1}^{N} X_{jk})$$

Collapsing:

$$X''_k = \sigma_0 w'_k Z_k + \phi X'_k \tilde{w}_k$$

where:

$w'_k = diag(\Sigma_{cov})' \Sigma_{cov}^{-1} = [w_{1k}, ..., w_{Nk}]$ is a vector of weights for Z_k

$\phi = \frac{1}{1'_N \Sigma_{cov}^{-1} 1_N}$ is a scalar

$\tilde{w}_k = \Sigma_{cov}^{-1} 1_N = [\tilde{w}_{1k}, ..., \tilde{w}_{Nk}]'$ is a vector of weights for X_k

Re-writing:

$$X''_k = \sigma_0 w'_k \frac{1}{\hat{\sigma}_0}(X_k - M_k) + \phi \tilde{w}'_k X_k = w'_k X_k - w' M_k + \phi \tilde{w}'_k X_k$$

where:

$X_k = [X_{1k}, ..., X_{Nk}]'$

$M_k = [\hat{\mu}_{0k}, ..., \hat{\mu}_{0k}]'$ is a vector with average recommendation

$\hat{\mu}_{0k} = \frac{1}{N}\sum_{j=1}^{N} X_{jk}$

Finally, the revealed aggregator can be represented with implied weights for each recommendation:

$$X_k'' = W_k' X_k$$

where:

$$W_k' = [w_k + \phi\tilde{w}_k - S_k]' \text{ with } S_k = [s_k, ..., s_k]' \text{ such that} s_k = \frac{1}{N}\sum_{j=1}^{N} w_{jk}$$

A.3 When revealed aggregator nests the simple average

Consider a simplified case of $K = 2, N = 2$, and $\Sigma_{cov} = \begin{bmatrix} \sigma^2 & \rho \\ \rho & \sigma^2 \end{bmatrix}$.

This represents symmetric information overlap (since all covariance terms are the same ρ) and information sets of the same size (since variances are the same σ^2). It follows that:

$$w_k' = diag(\Sigma_{cov})'\Sigma_{cov}^{-1} = \begin{bmatrix} \sigma^2, & \sigma^2 \end{bmatrix} \frac{1}{\sigma^4 - \rho^2} \begin{bmatrix} \sigma^2 & -\rho \\ -\rho & \sigma^2 \end{bmatrix}$$

$$= \frac{1}{\sigma^4 - \rho^2} \begin{bmatrix} \sigma^4 - \rho\sigma^2, & -\rho\sigma^2 + \sigma^4 \end{bmatrix}$$

$$s_k = \frac{1}{N}\sum_{j=1}^{N} w_{jk} = \frac{1}{\sigma^4 - \rho^2}(\sigma^4 - \rho\sigma^2)$$

$$\phi = \frac{1}{1_N'\Sigma_{cov}^{-1}1_N} = \frac{1}{\begin{bmatrix} 1, & 1 \end{bmatrix} \frac{1}{\sigma^4-\rho^2} \begin{bmatrix} \sigma^2 & -\rho \\ -\rho & \sigma^2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}} = \frac{1}{2}\left(\frac{\sigma^4 - \rho^2}{\sigma^2 - \rho}\right)$$

$$\tilde{w}_k = \Sigma_{cov}^{-1}1_N = \frac{1}{\sigma^4 - \rho^2} \begin{bmatrix} \sigma^2 & -\rho \\ -\rho & \sigma^2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sigma^4 - \rho^2} \begin{bmatrix} \sigma^2 - \rho \\ -\rho + \sigma^2 \end{bmatrix}$$

In this case, the revealed aggregator is the simple average. The implied weights W_k for $X_k'' = W_k' X_k$ are both equal to one-half:

$$
\begin{aligned}
W_k &= [w_k + \phi \tilde{w}_k - S_k] \\
&= \frac{1}{\sigma^4 - \rho^2} \begin{bmatrix} \sigma^4 - \rho\sigma^2 \\ \sigma^4 - \rho\sigma^2 \end{bmatrix} + \frac{1}{2} \left(\frac{\sigma^4 - \rho^2}{\sigma^2 - \rho} \right) \frac{1}{\sigma^4 - \rho^2} \begin{bmatrix} \sigma^2 - \rho \\ -\rho + \sigma^2 \end{bmatrix} - \frac{1}{\sigma^4 - \rho^2} \begin{bmatrix} \sigma^4 - \rho\sigma^2 \\ \sigma^4 - \rho\sigma^2 \end{bmatrix} \\
&= \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}
\end{aligned}
$$

Therefore, in this special case of equal information diversity and overlap, the revealed aggregator nests the consensus aggregator.

A.4 Trimming algorithm

The target share of missing observations is set to 50%. I use two types of trimming, hard trim which excludes least prolific brokers and least popular stocks, and recursive trim, which excludes brokers that don't overlap with other brokers.

Hard trim:

i Exclude brokers that provide coverage for 1 stock only.

ii Calculate how many brokers is each stock covered by, and find the bottom quartile of that distribution. Exclude these least popular stocks.

iii Exclude brokers that provide coverage for 1 stock only.

Recursive trim:

i For each broker, count with how many other brokers is there a zero overlap. Remove broker with the most zero overlaps, i.e. broker which has the least overlap with others.

ii Stop when distribution of overlap is substantially reduced, all brokers have pairwise overlap, or target missing share has been reached.

The trimming algorithms proceeds as follows:

1 Hard trim

2 Recursively trim

3 Continue repeating hard trim, then recursive trim, until all brokers have pairwise overlap.

4 If missing share target has not been reached, repeat recursive trim until it is reached.

A.5 Impact of an upgrade

In the main part, using stylized simplifications, I showed how the revealed aggregator maps the information sets of brokers based on the variance-covariance matrix of their recommendations across stocks, as of April 15th. Next I show the impact of a strong upgrade by April 22nd.

What happens to revealed aggregator due to an upgrade or downgrade? Since the recommendations for ALLY and TDC have not changed from April 15, 2016 to April 22, 2016, let's consider another change – an upgrade from 4 (sell) to 1 (strong buy) for Kansas City Southern (KSU:NYSE) by broker 32. Note that broker 32 was identified in last section as taking a stand against the majority. While this upgrade took place in the real data, all other upgrades and downgrades were ignored to isolate its impact.

Table A.1: Add caption

		Consensus		Revealed	
	Number of brokers	Score	Percentile	Score	Percentile
Before Upgrade	18	2.56	0.67	2.11	0.46
After Upgrade	18	2.39	0.54	1.51	0.22

This upgrade led to a significant improvement in the consensus recommendation, and an even bigger improvement based on the revealed aggregator, in terms of both scores and percentile ranks. As a result of the upgrade, the information set of broker 32 became closer to Stephens Inc and further from Buckingham Research Corp and broker 21648. The resulting small changes in the variance-covariance matrix impacted the revealed aggregate scores of other stocks, but not in any significant way, as Figure A.2 shows. However, in terms of the realized stock performance of KSU, it was not a good bet. KSU underperformed relative to the market for the following months, and it appears that broker 32's upgrade was just a mistake.

Figure A.1: Changes in Revealed Aggregator

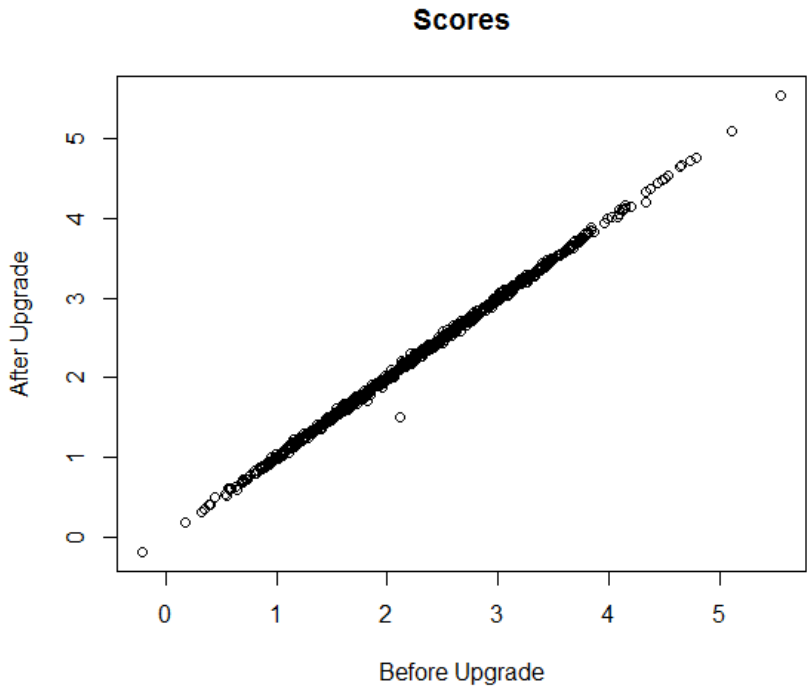
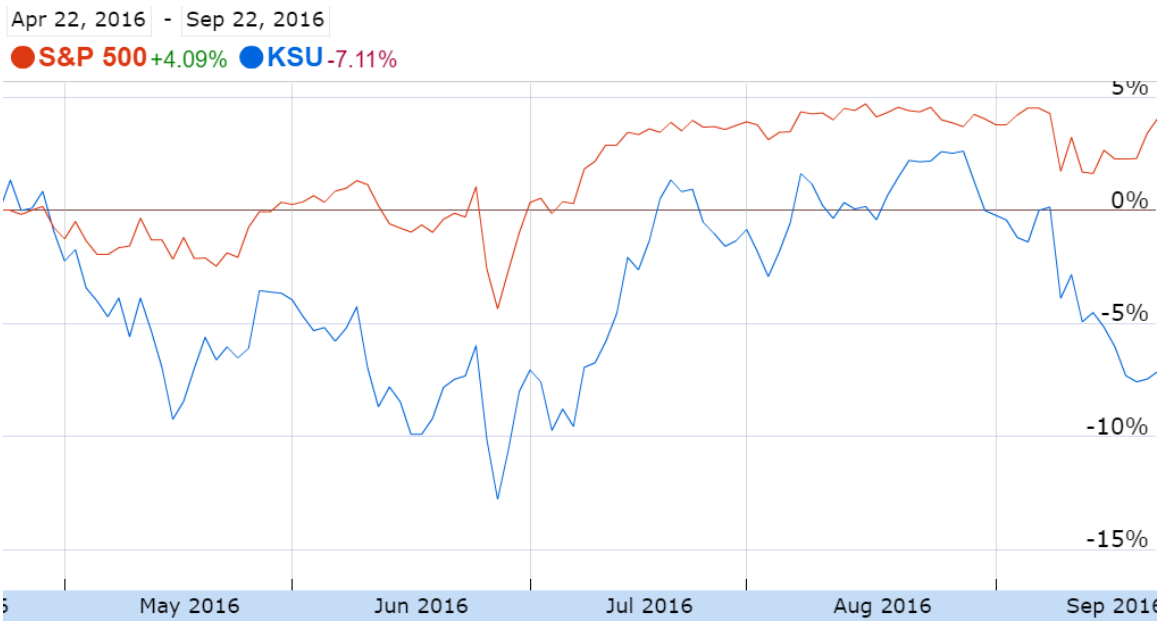


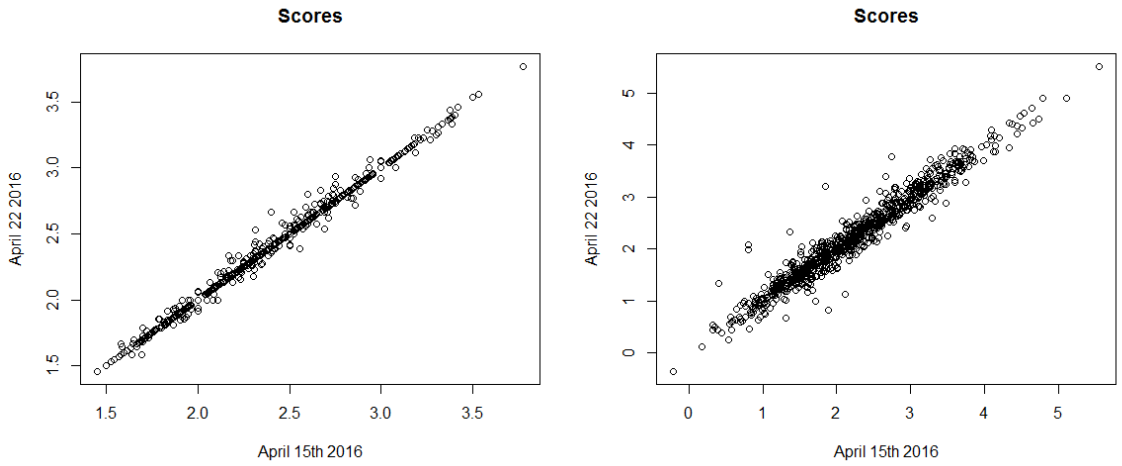
Figure A.2: Performance of KSU stock, April 15 - September 15, 2016



A.6 Further observations

As previously shown, any upgrade or downgrade changes the covariance matrix, which propagates to changes in the revealed aggregator. Considering jointly all the upgrades and downgrades that took place between April 15th and 22nd, many of the individual stocks changed, both in terms of the consensus recommendation and the revealed recommendation. The figure below shows these changes. This illustrates that the revealed aggregator is by design more volatile than the consensus aggregator. One can also observe how in case of the revealed aggregator the distribution is imposed by design.

Figure A.3: Upgrades and Downgrades in the Data



(a) Changes in Consensus Aggregator

(b) Changes in Revealed Aggregator

A.7 Using hard cut-offs

An alternative for portfolio construction is to consider hard cut-offs, such that $1 \leq Score_{k,\tau-1} \leq 1.5$ for the first portfolio, $1.5 \leq Score_{k,\tau-1} \leq 2$ for the second portfolio, $2 \leq Score_{k,\tau-1} \leq 2.5$ for the third portfolio, $2.5 \leq Score_{k,\tau-1} \leq 3$ for the fourth portfolio, and $Score_{k,\tau-1} > 3$ for the fifth portfolio, as the least favorably rated stocks.

This results in very different portfolio sizes, since the revealed aggregate scores are standardized to a set distribution, while the distribution of consensus scores can significantly change over time. Some of these portfolios have very low number or very high number of stocks included, and results should be interpreted with caution.

Table A.2: Returns and Sharpe Ratios earned by portfolios formed on the basis of consensus and revealed recommendations using hard cut-offs.

	Portfolios				
Formed on the basis of:	1st (best)	2nd	3rd	4th	5th (worst)
	Annualized average gross returns:				
Consensus	15.29	10.13*	9.69**	10.4	7.86
Revealed	14.17	7.24*	6.94**	11.26	9.12
CRSP Market			9.33		
	Annualized average market-adjusted returns:				
Consensus	5.46	0.73	0.33	0.99	-1.35
Revealed	4.43**	-1.91	-2.19	1.77	-0.19
	Annualized Sharpe Ratio:				
Consensus	0.58	0.50	0.47	0.43	0.25
Revealed	0.68	0.35	0.33	0.50	0.34
CRSP Market			0.43		

Value-weighted portfolios, with weekly rebalancing. 1st portfolio is one with the most favorably recommended stocks (to buy), while 5th portfolio is one with the least favorably recommended stocks (to sell). Sample data is weekly, Jan 6, 2007 - Dec 23, 2016, taking into account non-trading days. Returns are expressed in percent. The stars correspond to 0.1, 0.05, and 0.01 significance levels for a paired t-test. In case of gross returns, the null is that portfolio returns formed on consensus and revealed aggregates are the same. In case of excess returns, the null is that portfolio returns formed on consensus aggregate equals zero, and similarly for the revealed aggregate.

Table A.3: Abnormal performance of portfolios formed on the basis of consensus and revealed recommendations using hard cut-offs.

		Portfolios formed on the basis of									
		Consensus					Revealed				
		1st (best)	2nd	3rd	4th	5th (worst)	1st (best)	2nd	3rd	4th	5th (worst)
CAPM:	Intercept	0.167	0.03	0.015	0.009	-0.056	0.095***	-0.024	-0.03	0.031	-0.027
		(0.125)	(0.031)	(0.018)	(0.037)	(0.095)	(0.031)	(0.027)	(0.024)	(0.029)	(0.056)
	Rm-Rf	0.588***	0.896***	0.944***	1.064***	1.189***	0.928***	0.916	0.922***	1.018	1.147***
		(0.047)	(0.012)	(0.007)	(0.014)	(0.036)	(0.012)	(0.01)	(0.009)	(0.011)	(0.021)
Fama-French:	Intercept	0.179	0.017	0.012	0.026	-0.014	0.082***	-0.03	-0.034	0.037	-0.001
		(0.124)	(0.025)	(0.016)	(0.028)	(0.076)	(0.026)	(0.025)	(0.023)	(0.028)	(0.043)
	Rm-Rf	0.493***	0.987	0.983***	0.976***	0.963***	1.01	0.969	0.959***	0.997***	1.014***
		(0.055)	(0.011)	(0.007)	(0.012)	(0.034)	(0.011)	(0.011)	(0.01)	(0.012)	(0.019)
	SMB	0.191*	-0.094***	-0.13***	-0.054**	-0.056	-0.053**	-0.106***	-0.109***	-0.084***	-0.075**
		(0.109)	(0.022)	(0.014)	(0.025)	(0.067)	(0.023)	(0.022)	(0.021)	(0.025)	(0.038)
	HML	0.285***	-0.313***	-0.093***	0.37***	0.911***	-0.297***	-0.161***	-0.099***	0.121***	0.555***
		(0.09)	(0.018)	(0.012)	(0.021)	(0.055)	(0.019)	(0.018)	(0.017)	(0.02)	(0.031)

Time series regressions of weekly portfolio returns. 1st portfolio is one with the most favorably recommended stocks (to buy), while 5th portfolio is one with the least favorably recommended stocks (to sell). Sample data is weekly, Jan 6, 2007 - Dec 23, 2016, taking into account non-trading days. Returns are expressed in percent. Standard errors are in parentheses and the stars correspond to 0.1, 0.05, and 0.01 significance levels.

A.8 Using historical percentile cut-offs

Another alternative for portfolio construction is to consider percentile cut-offs based on last year's recommendation data.

Table A.4: Returns and Sharpe Ratios earned by portfolios formed on the basis of consensus and revealed recommendations using historical percentile cut-offs.

	Portfolios				
Formed on the basis of:	1st (best)	2nd	3rd	4th	5th (worst)
	Annualized average gross returns:				
Consensus	10.24**	7.62	11.34*	10.95	9.88
Revealed	12.85**	7.78	7.96*	9.1	9.73
CRSP Market			9.33		
	Annualized average market-adjusted returns:				
Consensus	0.84	-1.56	1.84	1.49	0.51
Revealed	3.23**	-1.41	-1.26	-0.21	0.37
	Annualized Sharpe Ratio:				
Consensus	0.51	0.37	0.52	0.47	0.38
Revealed	0.62	0.38	0.39	0.41	0.37
CRSP Market			0.43		

Value-weighted portfolios, with weekly rebalancing. 1st portfolio is one with the most favorably recommended stocks (to buy), while 5th portfolio is one with the least favorably recommended stocks (to sell). Sample data is weekly, Jan 6, 2007 - Dec 23, 2016, taking into account non-trading days. Returns are expressed in percent. The stars correspond to 0.1, 0.05, and 0.01 significance levels for a paired t-test. In case of gross returns, the null is that portfolio returns formed on consensus and revealed aggregates are the same. In case of excess returns, the null is that portfolio returns formed on consensus aggregate equals zero, and similarly for the revealed aggregate.

Table A.5: Abnormal performance of portfolios formed on the basis of consensus and revealed recommendations using historical percentile cut-offs.

		Portfolios formed on the basis of									
		Consensus					Revealed				
		1st (best)	2nd	3rd	4th	5th (worst)	1st (best)	2nd	3rd	4th	5th (worst)
CAPM:	Intercept	0.032	-0.019	0.037	0.021	-0.009	0.073**	-0.015	-0.011	-0.004	-0.014
		(0.029)	(0.024)	(0.026)	(0.037)	(0.057)	(0.03)	(0.028)	(0.027)	(0.027)	(0.053)
	Rm-Rf	0.899***	0.927***	0.986***	1.045***	1.116***	0.926***	0.92	0.915	1.002***	1.133***
		(0.011)	(0.009)	(0.01)	(0.014)	(0.021)	(0.011)	(0.01)	(0.01)	(0.01)	(0.02)
Fama-French:	Intercept	0.019	-0.024	0.039	0.035	0.019	0.06**	-0.021	-0.015	0.001	0.01
		(0.022)	(0.022)	(0.026)	(0.032)	(0.043)	(0.024)	(0.026)	(0.026)	(0.026)	(0.042)
	Rm-Rf	0.99***	0.98***	0.988	0.973***	0.984***	1.008***	0.973***	0.955***	0.987	1.01***
		(0.01)	(0.01)	(0.011)	(0.014)	(0.019)	(0.011)	(0.011)	(0.011)	(0.011)	(0.018)
	SMB	-0.106***	-0.136***	-0.085***	-0.029	-0.127***	-0.063***	-0.106***	-0.116***	-0.079***	-0.066*
		(0.02)	(0.019)	(0.023)	(0.028)	(0.038)	(0.021)	(0.023)	(0.023)	(0.023)	(0.037)
	HML	-0.313***	-0.15***	0.029	0.295***	0.573***	-0.293***	-0.162***	-0.104***	0.092***	0.509***
		(0.016)	(0.016)	(0.019)	(0.023)	(0.031)	(0.018)	(0.019)	(0.019)	(0.019)	(0.03)

Time series regressions of weekly portfolio returns. 1st portfolio is one with the most favorably recommended stocks (to buy), while 5th portfolio is one with the least favorably recommended stocks (to sell). Sample data is weekly, Jan 6, 2007 - Dec 23, 2016, taking into account non-trading days. Returns are expressed in percent. Standard errors are in parentheses and the stars correspond to 0.1, 0.05, and 0.01 significance levels.

A.9 Using a larger sample

Table A.6: Returns and Sharpe Ratios earned by portfolios formed on the basis of consensus and revealed recommendations for 1994-2016.

	Portfolios				
Formed on the basis of:	1st (best)	2nd	3rd	4th	5th (worst)
	Annualized average gross returns:				
Consensus	11.93	12.02	12.06	15.32	11.02
Revealed	13.65	12.35	9.49	13.79	11.6
CRSP Market			10.89		
	Annualized average market-adjusted returns:				
Consensus	0.94	1.02	1.05	4**	0.11
Revealed	2.5*	1.31	-1.27	2.62	0.64
	Annualized Sharpe Ratio:				
Consensus	0.45	0.48	0.50	0.63	0.40
Revealed	0.54	0.50	0.36	0.55	0.42
CRSP Market			0.44		

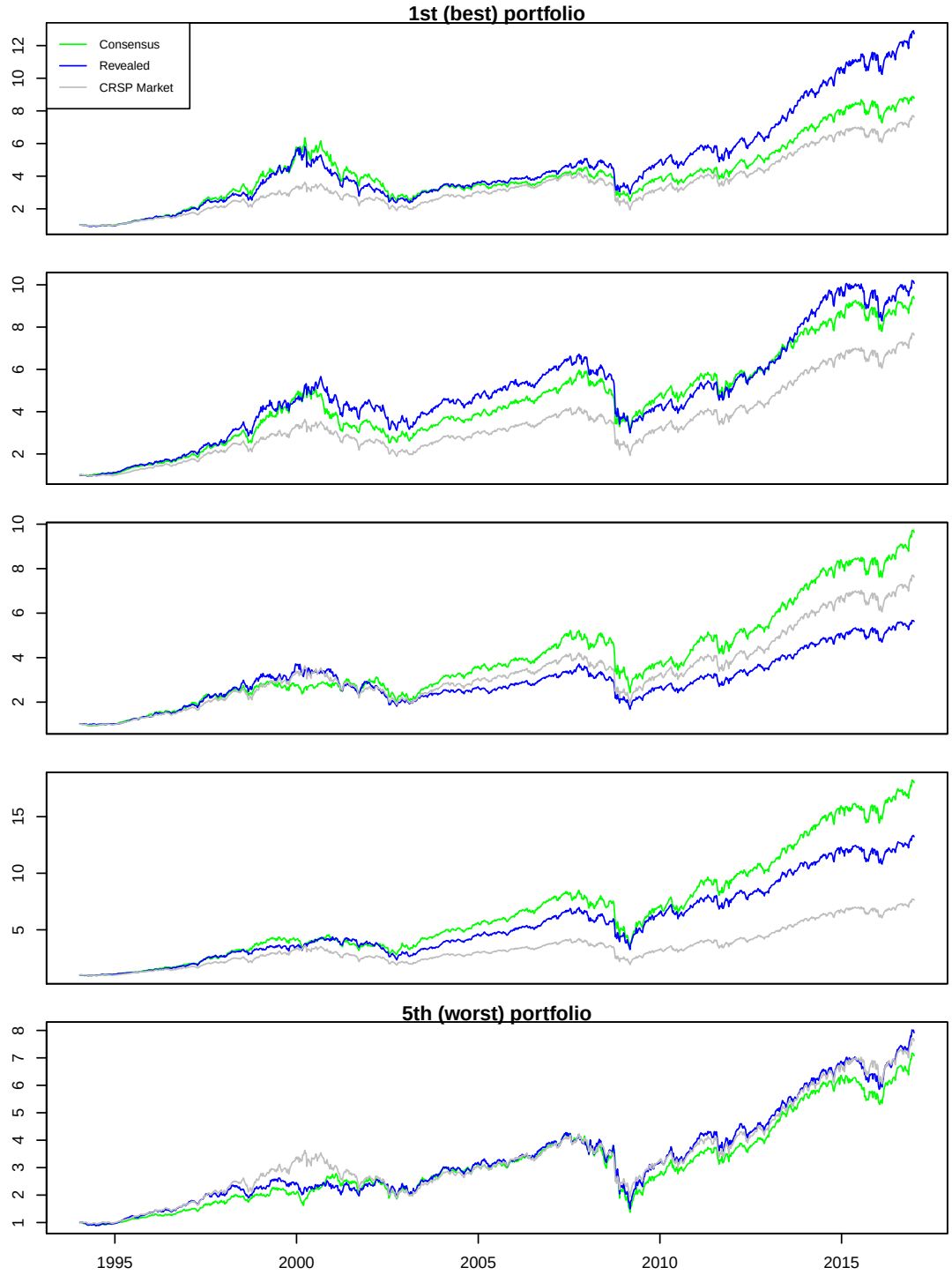
Value-weighted portfolios, with weekly rebalancing. 1st portfolio is one with the most favorably recommended stocks (to buy), while 5th portfolio is one with the least favorably recommended stocks (to sell). Sample data is weekly, Jan 7, 1994 - Dec 23, 2016, taking into account non-trading days. Returns are expressed in percent. The stars correspond to 0.1, 0.05, and 0.01 significance levels for a paired t-test. In case of gross returns, the null is that portfolio returns formed on consensus and revealed aggregates are the same. In case of excess returns, the null is that portfolio returns formed on consensus aggregate equals zero, and similarly for the revealed aggregate.

Table A.7: Abnormal performance of portfolios formed on the basis of consensus and revealed recommendations for 1994-2016.

		Portfolios formed on the basis of									
		Consensus					Revealed				
		1st (best)	2nd	3rd	4th	5th (worst)	1st (best)	2nd	3rd	4th	5th (worst)
CAPM:	Intercept	0.015	0.025	0.031	0.087**	0.012	0.048*	0.032	-0.014	0.055*	0.013
		(0.03)	(0.028)	(0.029)	(0.037)	(0.044)	(0.028)	(0.027)	(0.029)	(0.032)	(0.04)
	Rm-Rf	1.017***	0.963***	0.925***	0.924***	0.935***	0.994***	0.957***	0.929***	0.962***	0.994***
		(0.012)	(0.012)	(0.012)	(0.015)	(0.018)	(0.012)	(0.011)	(0.012)	(0.013)	(0.016)
Fama-French:	Intercept	0.04	0.04	0.027	0.067**	-0.025	0.068***	0.045*	-0.006	0.046	-0.017
		(0.026)	(0.026)	(0.026)	(0.031)	(0.035)	(0.026)	(0.025)	(0.027)	(0.029)	(0.033)
	Rm-Rf	1.028***	0.983***	0.944***	0.948***	0.955***	1.006***	0.974***	0.95***	0.984***	1.011***
		(0.011)	(0.011)	(0.011)	(0.013)	(0.015)	(0.011)	(0.01)	(0.011)	(0.012)	(0.014)
	SMB	-0.182***	-0.256***	-0.211***	-0.236***	-0.153***	-0.185***	-0.217***	-0.253***	-0.228***	-0.141***
		(0.019)	(0.019)	(0.02)	(0.023)	(0.026)	(0.019)	(0.019)	(0.02)	(0.021)	(0.024)
	HML	-0.328***	-0.169***	0.127***	0.358***	0.607***	-0.26***	-0.145***	-0.056***	0.208***	0.505***
		(0.019)	(0.019)	(0.019)	(0.023)	(0.026)	(0.019)	(0.018)	(0.019)	(0.021)	(0.024)

Time series regressions of weekly portfolio returns. 1st portfolio is one with the most favorably recommended stocks (to buy), while 5th portfolio is one with the least favorably recommended stocks (to sell). Sample data is weekly, Jan 7, 2007 - Dec 23, 2016, taking into account non-trading days. Returns are expressed in percent. Standard errors are in parentheses and the stars correspond to 0.1, 0.05, and 0.01 significance levels.

Figure A.4: Geometric returns earned by all portfolios since 1994.



Compound returns of value-weighted portfolios, with weekly rebalancing, for 1$ invested on Jan 5, 1994 to Dec 23, 2016.

A.10 With transaction costs

Table A.8: Returns and Sharpe Ratios earned by portfolios formed on the basis of consensus and revealed recommendations, net of transaction costs.

	Portfolios				
Formed on the basis of:	1st (best)	2nd	3rd	4th	5th (worst)
	Annualized average gross returns:				
Consensus	8.63***	6.94	8.78	9.91	8.11
Revealed	11.95***	5.96	6.33	9.15	8.76
CRSP Market			9.33		
	Annualized average market-adjusted returns:				
Consensus	-0.64	-2.19	-0.5	0.54	-1.12
Revealed	2.41	-3.08**	-2.75**	-0.16	-0.52
	Annualized Sharpe Ratio:				
Consensus	0.42	0.33	0.40	0.41	0.30
Revealed	0.58	0.28	0.30	0.39	0.32
CRSP Market			0.43		

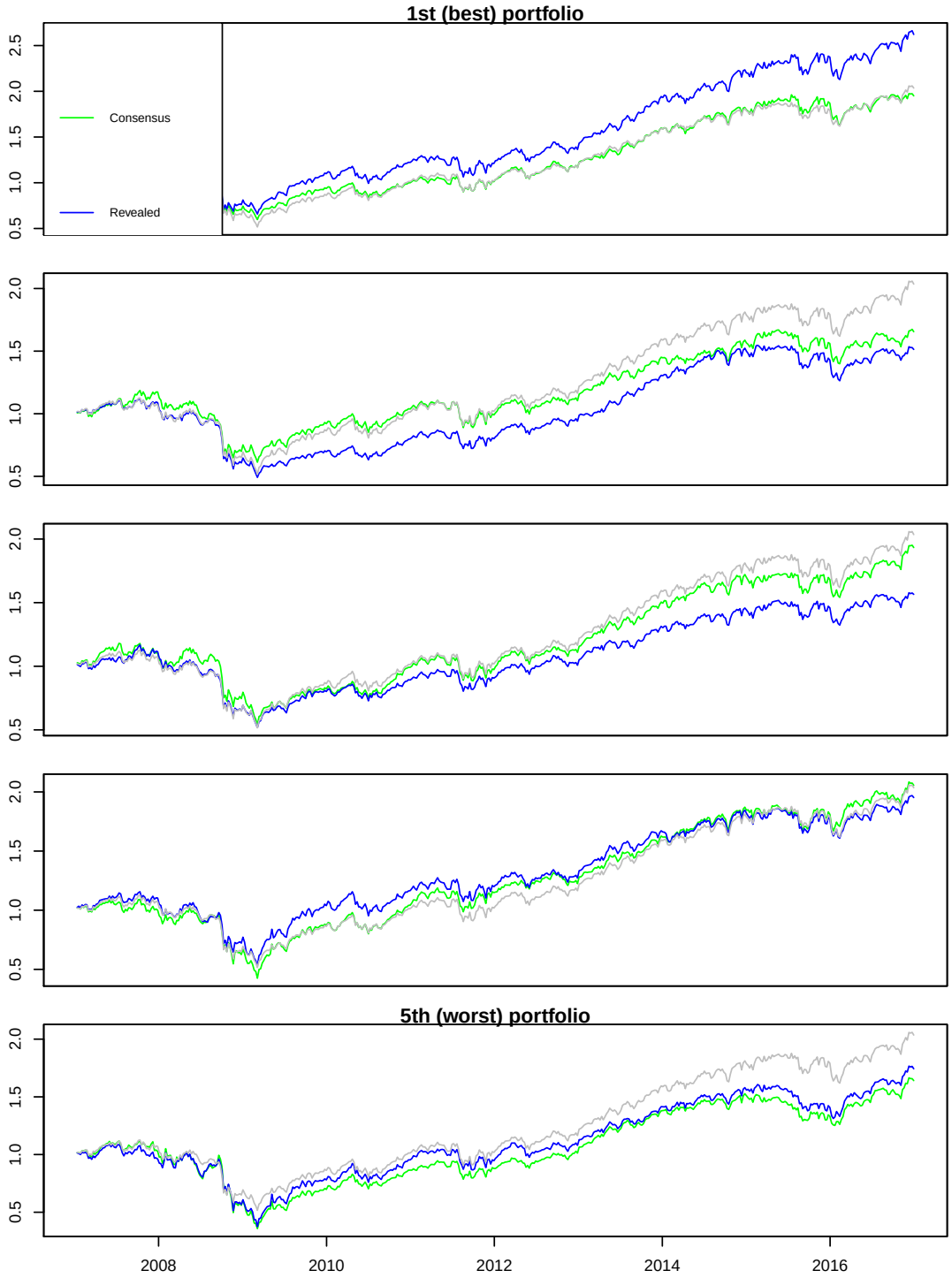
Value-weighted portfolios, with weekly rebalancing. 1st portfolio is one with the most favorably recommended stocks (to buy), while 5th portfolio is one with the least favorably recommended stocks (to sell). Sample data is weekly, Jan 6, 2007 - Dec 23, 2016, taking into account non-trading days. Returns are expressed in percent. The stars correspond to 0.1, 0.05, and 0.01 significance levels for a paired t-test. In case of gross returns, the null is that portfolio returns formed on consensus and revealed aggregates are the same. In case of excess returns, the null is that portfolio returns formed on consensus aggregate equals zero, and similarly for the revealed aggregate.

Table A.9: Abnormal performance of portfolios formed on the basis of consensus and revealed recommendations, net of transaction costs.

		Portfolios formed on the basis of									
		Consensus					Revealed				
		1st (best)	2nd	3rd	4th	5th (worst)	1st (best)	2nd	3rd	4th	5th (worst)
CAPM:	Intercept	0.003	-0.032	-0.007	0	-0.045	0.058*	-0.047*	-0.043*	-0.008	-0.034
		(0.029)	(0.026)	(0.025)	(0.041)	(0.056)	(0.031)	(0.028)	(0.024)	(0.034)	(0.056)
	Rm-Rf	0.906***	0.932***	0.98***	1.063***	1.146***	0.92***	0.917	0.932***	1.03***	1.152***
		(0.011)	(0.01)	(0.01)	(0.016)	(0.021)	(0.012)	(0.011)	(0.009)	(0.013)	(0.021)
Fama-French:	Intercept	-0.01	-0.039*	-0.005	0.016	-0.019	0.045*	-0.053**	-0.045**	0.001	-0.008
		(0.023)	(0.023)	(0.025)	(0.035)	(0.043)	(0.025)	(0.026)	(0.023)	(0.031)	(0.044)
	Rm-Rf	0.991***	0.995***	0.982	0.989***	1.016***	1.006	0.968***	0.967***	1.001***	1.02***
		(0.01)	(0.01)	(0.011)	(0.015)	(0.019)	(0.011)	(0.012)	(0.01)	(0.014)	(0.019)
	SMB	-0.104***	-0.146***	-0.068***	-0.088***	-0.089**	-0.069***	-0.097***	-0.129***	-0.126***	-0.089**
		(0.021)	(0.02)	(0.022)	(0.031)	(0.038)	(0.022)	(0.023)	(0.02)	(0.027)	(0.038)
	HML	-0.288***	-0.183***	0.024	0.332***	0.552***	-0.305***	-0.155***	-0.08***	0.17***	0.556***
		(0.017)	(0.016)	(0.018)	(0.025)	(0.031)	(0.018)	(0.019)	(0.017)	(0.022)	(0.032)

Time series regressions of weekly portfolio returns. 1st portfolio is one with the most favorably recommended stocks (to buy), while 5th portfolio is one with the least favorably recommended stocks (to sell). Sample data is weekly, Jan 6, 2007 - Dec 23, 2016, taking into account non-trading days. Returns are expressed in percent. Standard errors are in parentheses and the stars correspond to 0.1, 0.05, and 0.01 significance levels.

Figure A.5: Geometric returns earned by all portfolios, net of transaction costs.



Compound returns of value-weighted portfolios, with weekly rebalancing, for 1$ invested on Jan 6, 2007 to Dec 23, 2016.

A.11 Sensitivity analysis with the noisy information percolation model

Table A.10: Comparison of the Root Mean Square Error in case of Consensus and Revealed estimates, generated from the simulated information percolation model

		Low noise			σ			High noise
	λ	0.01	0.05	0.1	0.25	0.5	0.75	1
Low overlap	0.3	5.07, 4.96	5.55, 5.42	5.84, 5.71	5.07, 5.01	6.82, 6.78	7.19, 7.14	6.97, 6.97
(high diversity)	1.5	4.85, 4.66	5.28, 5.09	5.1, 4.94	5.51, 5.34	5.3, 5.2	5.96, 5.94	6.83, 6.81
	3	4.26, 4.05	4.41, 4.19	4.88, 4.66	4.62, 4.44	5.24, 5.05	5.71, 5.6	6.7, 6.62
	6	3.74, 3.47	3.55, 3.3	3.62, 3.34	4.46, 4.17	3.85, 3.66	4.41, 4.26	4.96, 4.86
	9	2.98, 2.66	2.92, 2.67	2.91, 2.65	2.56, 2.33	3.48, 3.17	3.64, 3.46	4.17, 4.05
	12	2.33, 2.12	2.06, 1.87	2.01, 1.84	2.02, 1.84	2.13, 2.01	2.52, 2.43	2.95, 2.9
	13.5	1.68, 1.58	1.82, 1.72	1.61, 1.52	1.71, 1.64	1.8, 1.74	2.2, 2.18	2.31, 2.42
	15	1.18, 1.21	1.44, 1.47	1.28, 1.29	1.62, 1.68	1.63, 1.65	1.7, 1.79	2.08, 2.3
High overlap	16.5	1.06, 1.21	1.06, 1.2	1.02, 1.21	1.07, 1.21	1.21, 1.39	1.36, 1.55	1.69, 1.96
(low diversity)	22.5	0.32, 0.86	0.3, 0.79	0.36, 0.83	0.4, 0.89	0.48, 0.8	0.64, 1	0.81, 1.28

Table reports pairs (Consensus RMSE, Revealed RMSE), i.e. root mean square error, calculated from 100 simulations of the noisy information percolation model. Lower number means a more precise estimate. Increasing σ means increasing noise of the signals that agents receive. Increasing Poisson intensity of meeting arrivals λ means increasing information overlap and concentration. The number of agents and number of stocks is kept constant, at $J = K = 30$.

Appendix B

Appendix: Debt Management and Demographics

Table 1.1: Measurements of maturity and debt

Share of short-term vs. medium- and long-term debt in total outstanding debt	
Short-term debt is defined as debt with original maturity of one year or less, and medium- and long-term (MLT) debt has maturity greater than one year. A more granular distinction is below, based on outstanding securities in a portfolio, or newly-issued securities.	
Portfolio shares	*Issuance shares*
$$PS_m = \frac{D_m^{TOT}}{D^{TOT}} \forall m$$ where D_m^{TOT} is the face value of outstanding securities with maturity m and D^{TOT} is the nominal volume of total debt outstanding, at the end of the period, measured in months, years. Maturity m can be an interval of any granularity.	$$IS_m = \frac{D_m^{NEW}}{D^{NEW}} \forall m$$ where D_m^{NEW} is the face value of newly-issued securities with maturity m and D^{NEW} is the nominal volume of total debt issued in each period, measured in months, years. Maturity m can be an interval of any granularity.
Weighted average remaining maturity:	*Weighted average original maturity:*
is the weighted average time to repay the principals of the outstanding securities where the weights reflect their relative nominal values: $$WARM = \frac{\sum_{s=1}^{N} m_s X_s}{\sum_{s=1}^{N} X_s} = \frac{\sum_{s=1}^{N} m_s X_s}{D^{TOT}}$$ *N* is the number of outstanding securities at the end of the period, m_s is the time to maturity of security *s*, and X_s is face value. $WARM$ is in units of the period (months, years). The sum of X_s across all outstanding securities *N* is equivalent to the nominal volume of total debt outstanding at the end of the period, D^{TOT}.	is the weighted average weighted time to repay the principals of the newly-issued securities where the weights reflect their relative nominal values: $$WAOM = \frac{\sum_{i=1}^{K} m_i X_i}{\sum_{i=1}^{K} X_i} = \frac{\sum_{i=1}^{K} m_i X_i}{D^{NEW}}$$ *K* is the number of newly-issued securities during the period, m_i is the time to maturity of newly-issued security *i*, and X_i is its face value. Similarly to $WARM$, $WAOM$ is in units of the period (months, years). The sum X_i across all the newly-issued securities K is equivalent to the nominal volume of newly-issued debt at the end of the period, D^{NEW}.
Effective Maturity:	
Elasticity of the overall debt to changes in unexpected inflation. This measure makes the following modifications to the average residual maturity: "(i) zero maturity is assigned to foreign currency debt and price-index-linked debt; (ii) the time before the next coupon adjustment is considered for variable-rate securities; (iii) the earliest redemption date is assigned to putable and convertible securities and (iv) the latest redemption date is assigned to callable securities." Missale (1999, chap. 4)	
Macaulay Duration:	*Modified Macaulay Duration:*
Linear approximation of the elasticity of the value of fixed-rate securities to unexpected changes in interest rates. Calculated as the weighted average time to receive all interests and principal, where the weights reflect the relative present values of the cash flows: $$M = \sum_{j=1}^{n} \left[\frac{CP_j}{(1+i)^{T_j}} \right] T_j \Big/ \sum_{j=1}^{n} \left[\frac{CP_j}{(1+i)^{T_j}} \right]$$ where CP_j is the cash payment (interest plus principal) at period *j*, T_j is the time until cash payment *j* is made, *n* is the number of payments and *i* is the yield to maturity.	Linear approximation of the elasticity of the value of the fixed-rate securities, assuming that the bond's expected cash payment does not change when the yield changes. For fixed payment bonds, Macaulay duration and modified duration are equal. Formally, $$MM = \frac{M}{1 + \frac{YTM}{k}}$$ Where M is the Macaulay duration, YTM is the yield-to-maturity of the bond, and k is the frequency of coupon payments per year.

B.1 Optimal share of long-term debt under the tax smoothing approach

Bohn (1990) shows that $\Sigma_r d_t + \Sigma_{\hat{g},r} - w_t \Sigma_{\hat{g},r} = 0$ and only output fluctuations are quantitatively significant:

$$d_t = w_t \Sigma_r^{-1} \Sigma_{\hat{g},r}$$

Following Bohn (1990), I limit the scope to $K = 0, ST, LT$ securities. Hence r_0 is the risk-free overnight interbank interest rate, r_{ST} is the interest rate on three-month interbank securities, while r_{LT} is the interest rate on a 10-year government bond. Excess rates are therefore $\hat{r}_{ST} = r_{ST} - r_0$ and $\hat{r}_{LT} = r_{LT} - r_0$, with $k = ST, LT$. This implies the following.

$$\begin{bmatrix} d^{ST} \\ d^{LT} \end{bmatrix} = w_t \times \begin{bmatrix} \sigma^2_{\hat{r}_{ST}} & \sigma_{\hat{r}_{ST},\hat{r}_{LT}} \\ \sigma_{\hat{r}_{ST},\hat{r}_{LT}} & \sigma^2_{\hat{r}_{LT}} \end{bmatrix}^{-1} \times \begin{bmatrix} Cov(\hat{r}_{ST}, \sum_{j\geq 0} \psi^j \hat{y}_{t+1+j}) \\ Cov(\hat{r}_{LT}, \sum_{j\geq 0} \psi^j \hat{y}_{t+1+j}) \end{bmatrix} = w_t \times \begin{bmatrix} s_{ST} \\ s_{LT} \end{bmatrix}$$

Since the optimal structure of debt is proportional to the vector s, the optimal share of long-term debt is

$$\beta^{TS} = \frac{s_{LT}}{s_{LT} + s_{ST}}$$

Noting present value expression $\sum_{j\geq 0} \psi^j \hat{y}_{t+1+j}$, with innovations that represent the change in expectation of far out realizations of output growth, Bohn (1990) argues that vector-autoregregression (VAR) techniques are ideally suited for the task of identifying innovations from the multivariate process. I follow the exact same approach. I run bivariate VARs that include only GDP growth rate and a single excess return series, using quarterly data as far back as possible, including a constant and four lags. Therefore for $z = [y, \hat{r}_k]'$, I obtain output innovations $\hat{e}^y = [\hat{e}^y_5, ..., \hat{e}^y_T]$ from

the following $VAR_k(4)$:

$$z_t = \mu + \alpha_1 z_{t-1} + ... + \alpha_4 z_{t-4} + e_t \quad \text{for } t = 1, ..., T$$

I calculate the present-value expressions $\sum_{j\geq 0} \psi^j \hat{y}_{t+1+j}$ in each period, using $\psi = 0.99$ which corresponds to 1% time discounting rate, in the following way:

$$PV_k(\hat{y})_5 = \sum_{j=4}^{T-1} \psi^{j-4} \hat{e}^y_{j+1}$$

$$...$$

$$PV_k(\hat{y})_{T-1} = \sum_{j=T-2}^{T-1} \psi^{j-(T-2)} \hat{e}^y_{j+1}$$

Hence I obtain a vector of present-value expressions $PV_k(\hat{y}) = [PV_k(\hat{y})_5, ..., PV_k(\hat{y})_{T-1}]'$. I then calculate the covariance of excess returns and output innovations as $Cov(\hat{r}_k, PV_k(\hat{y}))$ based on all data available. This is done separately for $k = ST$ and $k = LT$.

I also calculate the variance-covariance matrix of excess returns, Σ_r:

$$\Sigma_r = \begin{bmatrix} \sigma^2_{\hat{r}_{ST}} & \sigma_{\hat{r}_{ST},\hat{r}_{LT}} \\ \sigma_{\hat{r}_{ST},\hat{r}_{LT}} & \sigma^2_{\hat{r}_{LT}} \end{bmatrix}$$

Therefore I obtain:

$$\begin{bmatrix} s_{ST} \\ s_{LT} \end{bmatrix} = \Sigma_r^{-1} \times \begin{bmatrix} Cov(\hat{r}_{ST}, PV_{ST}(\hat{y})) \\ Cov(\hat{r}_{LT}, PV_{LT}(\hat{y}) \end{bmatrix}$$

and fully written out, the expression for vector s is:

$$\begin{bmatrix} s_{ST} \\ s_{LT} \end{bmatrix} = \frac{1}{det} \times \begin{bmatrix} \sigma^2_{\hat{r}_{LT}} Cov(\hat{r}_{ST}, PV_{ST}(\hat{y})) - \sigma_{\hat{r}_{ST},\hat{r}_{LT}} Cov(\hat{r}_{LT}, PV_{LT}(\hat{y}) \\ -\sigma_{\hat{r}_{ST},\hat{r}_{LT}} Cov_(\hat{r}_{ST}, PV_{ST}(\hat{y})) + \sigma^2_{\hat{r}_{ST}} Cov(\hat{r}_{LT}, PV_{LT}(\hat{y}) \end{bmatrix}$$

where $det = \sigma^2_{\hat{r}_{ST}} \sigma^2_{\hat{r}_{LT}} - \sigma^2_{\hat{r}_{ST},\hat{r}_{LT}}$.

Lastly, I calculate the optimal share of long-term debt as $\beta^{TS} = \frac{s_{LT}}{s_{LT}+s_{ST}}$.

Note that the results of the empirical analysis do not change if I use backward-looking 10-year overlapping windows and derive a time-variant version instead.

Table B.2: Summary statistics for 2000-2010 annual data on tax smoothing determinants of maturity structure

		Summary statistics for 2000-2010 annual data				
	Bivariate VARs	Covariance of output innovations with excess returns		Optimal composition by maturity		Optimal share of long-term debt
Country	sample starts in	$Cov(\hat{r}_{LT}, PV_{LT}(\hat{y}))$	$Cov(\hat{r}_{ST}, PV_{LT}(\hat{y}))$	s_{LT}	s_{ST}	β^{TS}
Australia	Q3-1991	1.82	-0.05	1.58	-1.44	11.30
Austria	Q1-2000	0.11	-0.35	0.48	-4.33	-0.12
Belgium	Q1-2000	-0.28	-0.23	-0.12	-2.62	0.04
Canada	Q1-1976	-0.44	-1.23	0.19	-0.36	-1.17
Chile	Q3-2005	1.80	0.17	0.59	0.02	0.97
Czech Republic	Q2-2001	-0.18	-0.31	-0.23	-0.12	0.75
Denmark	Q1-1998	0.62	-0.02	0.91	-0.15	1.19
Estonia						
Finland	Q1-2000	-0.22	-0.69	0.14	-8.05	-0.02
France	Q1-1971	0.23	-0.13	0.23	-0.77	-0.44
Germany	Q2-1971	1.43	0.45	0.57	0.86	0.40
Greece	Q3-1998	9.54	-0.33	0.22	-0.05	1.33
Hungary	Q2-2000	3.56	1.12	0.60	0.22	0.73
Iceland	Q2-1999	10.61	-0.30	2.21	-1.89	6.82
Ireland	Q1-2000	12.21	0.48	2.35	-0.63	1.37
Israel	Q1-1997	2.92	0.22	0.15	0.17	0.47
Italy	Q1-2000	0.72	-0.20	0.51	-2.96	-0.21
Japan	Q1-1997	-0.52	-0.16	-0.97	-10.16	0.09
Korea	Q4-2001	-0.09	0.49	-0.19	0.45	-0.74
Luxembourg	Q1-2000	-1.87	-0.64	-0.59	-6.63	0.08
Mexico	Q4-2002	0.65	0.01	0.86	-0.05	1.07
Netherlands	Q1-2000	0.07	-0.27	0.28	-3.17	-0.10
New Zealand	Q2-1989	1.27	0.12	0.25	-0.25	-49.16
Norway	Q1-1986	-2.27	0.73	-1.42	1.44	-78.23
Poland	Q1-2002	-1.42	-0.64	-0.31	-0.03	0.91
Portugal	Q1-2000	4.11	-0.32	0.61	-5.22	-0.13
Slovak Republic	Q4-2001	-3.98	2.07	-1.72	0.12	1.07
Slovenia	Q1-2005	3.19	-0.32	0.78	-3.23	-0.32
Spain	Q1-1978	-1.92	0.38	-0.44	0.27	2.64
Sweden	Q1-1988	1.84	2.22	0.64	0.88	0.42
Switzerland	Q1-1975	0.16	0.34	-0.01	0.24	-0.05
Turkey						
United Kingdom	Q1-1979	0.64	0.10	0.18	0.13	0.57
United States	Q1-1962	-2.11	-0.44	-0.61	-0.46	0.57

B.2 Descriptive statistics of variables in panel estimation

Table B.3: Descriptive statistics of variables in panel estimation

Variable	Mean	Std. Dev.	Min.	Max.	N
GovtMaturity	5.749	2.11	0	14.5	337
c	0.06	0.497	-1.318	1.701	352
cETermSpread	0.132	1.01	-7.515	6.665	304
TaxSmoothingLT	-3.06	16.256	-78.233	11.304	352
WorkingSavers	65.508	2.453	58.273	69.783	374
ContractSaveAssets	0.046	0.13	-0.560	1.38	363
CorpMaturity	14.089	7.361	2.128	37.01	363
RiskSpread	0.455	1.499	-5.020	6.16	335
InflationCredibility	5.642	22.395	-15.884	371.309	333
Inflation	3.379	4.882	-5.39	52.851	374
Debt	39.698	28.628	0	165.414	370

B.3 Specification of GMM models

The model specification is:

$$WARM_{j,t} = \alpha WARM_{j,t-1} + \gamma x_{j,t} + \mu + (\eta_j + v_{j,t})$$

for $t = [1, ..., 11]$ period, j countries, with $\gamma = [\gamma_1, ..., \gamma_9]$ a 1×9 vector and $x_t = [c_j; c_j E\bar{ts}_t; \beta_t^{TS}; ws_t; csa_t; pmat_t; risk_t; (r_t d_t)^{-1}; \pi_t;]'$ a 9×1 vector.

In the First Difference GMM model with $|\alpha| < 1$, I follow the standard three assumptions which lead to a set of valid instruments Z_i for moment conditions $E[Z_i' \Delta v_i] = 0$. These standard three assumptions are regarding the error components, serially uncorrelated shocks, and predetermined initial conditions:

(1) error components: $E[\eta_i] = E[v_{i,t}] = E[\eta_i v_{i,t}] = 0$

(2) serially uncorrelated shocks: $E[v_{i,s} v_{i,t}] = 0 \forall st$

(3) predetermined initial conditions: $E[y_{i,1}v_{i,t}] = 0 \forall t = 2, T$

With one additional assumption, I treat the dependent variables as either exogenous, predetermined, or endogenous:

Table B.4: Moment conditions corresponding to assumptions on the explanatory variables

Strictly exogenous	Predetermined explanatory variables	Endogenous explanatory variables
$E[x_{i,s}v_{i,t}] = 0 \; \forall \; s, t$	$E[x_{i,s}v_{i,t}] = 0 \; \forall \; s \leq t$	$E[x_{i,s}v_{i,t}] = 0 \; \forall \; s < t$
Following the first-differencing transformation of the model, $\Delta y_{i,t} = \alpha \Delta y_{i,t-1} + \beta \Delta x_{i,t} + \Delta v_{i,t}$, these assumptions imply the following set of additional moment conditions, such that $E[Z_i' \Delta v_i] = 0$: $E[y_{i,t-s} \Delta v_{i,t}] = 0 \; \forall \; t = 3, \ldots T, s \geq 2$		
$E[x_{i,s} \Delta v_{i,t}] = 0$ $\forall \; t = 3, \ldots T$ $s = 1,2, \ldots T$	$E[x_{i,t-s} \Delta v_{i,t}] = 0$ $\forall \; t = 3, \ldots T$ $s = 1,2, \ldots T$	$E[x_{i,t-s} \Delta v_{i,t}] = 0$ $\forall \; t = 3, \ldots T$ $s = 2, \ldots T$
This is a very large set of instruments.	This is a subset of strict exogeneity moment conditions.	This is a further subset of moment conditions.

In case of System GMM, I make a further assumption concerning the initial conditions, namely that change in the initial condition is uncorrelated with the unobserved individual effects, i.e., $E[\Delta y_{i,2} \eta_i] = 0$. I use the additional non-redundant linear moment conditions for the equations in levels. The levels equation is

$$y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t} + v_{i,t}$$

and the associated moment conditions are:

$$E[\Delta y_{i,s} u_{i,T}] = 0 \forall s = 2, \ldots, T - 1 \text{ where } u_{i,T} = \eta_i + v_{i,T}$$

These additional non-redundant moment conditions provide an improvement in efficiency and reduction in finite sample bias when data exhibit high time-series persistence, and if the additional assumption is valid (Blundell and Bond, 1998).

B.4 Selection of the panel model

For the purposes of model selection, I apply specification tests described and presented in Table B.5.

Table B.5: Specification tests for model selection

Test	Description
Arellano and Bond (1991) test, labelled AB	Test of for serial correlation in the residuals of the first-differences specification. While there is expected to be AR(1) correlation in the residuals, it is expected that there will not be any AR(2) correlation, so this is a key test of the model.
Hansen J test of overidentifying restrictions (Hansen 1982; Sargan 1958)	Under the null hypothesis, the set of moment conditions used is valid. High test statistic and small p-value implies rejection of the validity of the set of moment conditions and equivalent overidentifying restrictions. Although this test of overidentifying restrictions is robust to heteroskedasticity in the errors, it is not robust to too many instruments, and a problem is confirmed if the Hansen p-value is close to 1 (Roodman 2006). I use the Hansen J test since p-value of the Hansen J test statistic < 1 in all models.[1]
Difference-in-Hansen test of nested hypotheses[2]	I use this test when only a subset of the moment conditions that are valid under the null hypothesis remain valid under a less restrictive alternative hypothesis[3]. Large values of DH test statistic imply that null hypothesis cannot be accepted, which suggests that either some of the additional moment conditions are invalid, or that moment conditions in both sets are invalid.
Hausman test (1978; 1981).	In this test, I focus on the difference in the estimated parameters, tested individually, in order to test the validity of additional moment conditions [4]. Large values of the Hausman test statistic suggests that some of the additional moment conditions are invalid.

[1] The Sargan test of overidentifying restrictions is based on similar logic but is not robust to heteroskedasticity in the errors, yet it is robust to too many instruments (Arellano and Bond 1991). Sargan test is not reported since Hansen J test is preferred.

[2] The difference Hansen and Hausman test procedures are likely to have low power in cases where the moment conditions available for the equations in first-differences provide only weak identification, when the true value of $\alpha \to 1$. Therefore these tests should be treated with caution.

[3] For Z_i^A that is a strict subset Z_i, under the null hypothesis that both subsets have valid moment conditions, the difference between the Hansen J test statistics is $DH = H - H^A \sim \chi^2(m - m^a)$, where m is the number of instruments.

[4] The estimated parameters are tested individually and not as a vector, for computational simplicity. Under the null hypothesis that additional moment restrictions present in Z_i but not in Z_i^A are valid, the Hausman test statistic of single parameter estimate is $h = \frac{\hat{\beta}^A - \hat{\beta}}{\sqrt{se_A^2 - se^2}} \sim N(0,1)$.

Main results are presented in Table 4.3. First Difference GMM models (5) and (6) are adequately specified with the instrument choice established above, with no AR(2)

serial correlation of residuals and no rejection of the null hypothesis of valid overidentifying restrictions using the Hansen J test, at the 5% significance level. The two-step weight matrix estimation procedure reflected in model (6) should theoretically produce lower standard errors, and is asymptotically more efficient. The standard errors in the two models are similar. Therefore, between the two models (5) and (6), the two-step estimation procedure in model (6) is marginally preferred.

Given near-unit root properties of the dependent and possibly explanatory variables, the System GMM estimator should be superior (Blundell and Bond 1998). However, the Hausman test suggests that not all of the additional moment conditions present in the System GMM model, beyond the moment conditions present in the First Diffference GMM model, are valid. Specifically, several of the $|h|$ statistics exceed 1.96, the critical value at 5% significance level. Furthermore, in case of System GMM, $\alpha < 0.9$ is the boundary for a consistent estimate. The system GMM results are reported in Table B.6. The Hausman test results in Table B.7. Therefore, the First Difference GMM model (6) is preferred.

Table B.6: Cross-country determinants of maturity structure

	(1) FDGMM	(2) FDGMM	(3) SYSGMM	(4) SYSGMM
L.GovtMaturity	0.779***	0.773***	0.954***	0.925***
	(0.168)	(0.213)	(0.052)	(0.066)
cETermSpread	-0.191**	-0.318***	-0.124*	-0.156**
	(0.095)	(0.078)	(0.067)	(0.073)
WorkingSavers	0.212***	0.208**	-0.001	0.001
	(0.061)	(0.091)	(0.032)	(0.037)
ContractSaveAssets	1.222**	1.601**	1.353***	1.295**
	(0.525)	(0.649)	(0.387)	(0.497)
CorpMaturity	-0.011	-0.007	0.004	0.007
	(0.017)	(0.020)	(0.010)	(0.010)
RiskSpread	0.010	0.015	-0.032	-0.046*
	(0.020)	(0.034)	(0.022)	(0.028)
InflationCredibility	0.000	-0.001	-0.002**	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)
Inflation	0.081**	0.101	0.050**	0.046**
	(0.040)	(0.067)	(0.016)	(0.022)
c			0.174	0.213
			(0.278)	(0.299)
TaxSmoothingLT			-0.000	-0.002
			(0.015)	(0.013)
Constant			0.202	0.202
			(2.218)	(2.476)
Observations	213	213	241	241
Adjusted R^2				
No. countries	27	27	27	27
No. instruments	16	16	26	26
AB:AR(1)	0.048	0.032	0.009	0.019
AB:AR(2)	0.499	0.593	0.579	0.621
Hansen J	0.167	0.167	0.099	0.099

Notes: ***p<0.01, **p<0.005, *p<0.1. Year effects included in models (1) and (3), excluded from models (2) and (4). One-step weight matrix for models (5) and (7), and two-step weight matrix for models (6) and (8). Robust standard errors in parentheses. Windmeijer(2005) corrected errors are reported for two-step estimation; p-values reported for AB and Hansen J tests.

Table B.7: Hausman test for vailidity of additional moment conditions in System GMM, beyond First Difference GMM

	One-step estimation models					Two-step estimation models				
	FDGMM		SYSGMM			FDGMM		SYSGMM		
	beta	se2	beta	se2	h	beta	se2	beta	se2	h
L.GovtMaturity	0.78	0.03	0.95	0.00	-1.10	0.77	0.05	0.92	0.00	-0.75
cETermSpread	-0.19	0.01	-0.12	0.00	-0.99	-0.32	0.01	-0.16	0.01	-5.75
WorkingSavers	0.21	0.00	0.00	0.00	4.11	0.21	0.01	0.00	0.00	2.48
ContractSaveAssets	1.22	0.28	1.35	0.15	-0.37	1.60	0.42	1.29	0.25	0.74
CorpMaturity	-0.01	0.00	0.00	0.00	-1.07	-0.01	0.00	0.01	0.00	-0.82
RiskSpread	0.01	0.00	-0.03	0.00		0.02	0.00	-0.05	0.00	3.12
InflationCredibility	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	
Inflation	0.08	0.00	0.05	0.00	0.82	0.10	0.00	0.05	0.00	0.87
c			0.17	0.08				0.21	0.09	
TaxSmoothingLT			0.00	0.00				0.00	0.00	
Constant			0.20	4.92	0.00			0.20	6.13	

Some of the h statisics exceed 1.96, the critical value at 5% significance level. This test implies rejection of the null hypothesis that all the additional moment conditions present in the System GMM model, beyond the moment conditions present in the First Difference GMM model, are valid. In particular, this rejection is in place for the share of working savers in the population.

B.5 Specification tests for the panel model

To test the power of instruments, consider the regression of the instrumented variable on its instrument, controlling for the exogenous explanatory variables, as applicable in case of first-difference equation. In the first-difference equation, the instrumented variable is $\Delta y_{i,t-1}$, with instruments in Z_i. Hence the equivalent of a first-stage regression is $\Delta y_{i,t-1} = \pi_1 y_{i,t-2} + \pi_2 \Delta x_{i,t} + r_{i,t}$. The instrument is informative and have power if π_1 and π_2 do not equal 0. In the levels equation, the instrumented variable is $y_{i,t-1}$ with instruments $(\Delta y_{i,t-1}, x_{i,t})$. The equivalent of a first-stage regression is $y_{i,t-1} = \phi_1 \Delta y_{i,t-1} + \phi_2 x_{i,t} + s_{i,t}$.

This regression is presented in Table B.8. The hypothesis of zero partial correlation between the instrumented variable and its instrument is rejected at 5% significance level, while the fit is good and the corresponding F-statistics are in favour of the joint significance of the variables.

Table B.8: Testing the power of instruments

First-difference equation	
	LD.GovtMaturity
L2.GovtMaturity	-0.352***
	(0.093)
D.cETermSpread	0.020
	(0.023)
D.WorkingSavers	-0.519**
	(0.241)
D.ContractSaveAssets	-0.892***
	(0.247)
D.CorpMaturity	-0.003
	(0.005)
RiskSpread	-0.057***
	(0.011)
InflationCredibility	-0.002*
	(0.001)
Inflation	-0.033*
	(0.017)
Constant	2.200***
	(0.561)
Observations	213
Number of countryid	27
Adjusted R2	0.189
F-statistic	17.64

***p<0.01, **p<0.05, *p<0.1 Robust standard errors in parentheses. Fixed-effects model, no year effects. High adjusted R^2 and F-statistics show that the instrument used in the first-difference equations used for First Difference GMM model is informative.

I further test whether some of the explanatory variables that were treated as strictly exogenous might be endogenous with the difference-in-Hansen J tests. Specifically, I consider a test where matrix Z_i corresponds to the moment conditions treating the variable x as strictly exogenous, given by both $E[x_{i,s}\Delta v_{i,t}] = 0 \forall t = 3, ..., T; s = 1, 2, ..., T$ and $E[x_{i,t-s}\Delta v_{i,t}] = 0 \forall t = 3, ..., T; s = 2, ..., T$ while matrix Z_i^A corresponds to the moment conditions treating the variable as endogenous, given only by $E[x_{i,t-s}\Delta v_{i,t}] = 0 \forall t = 3, ..., T; s = 2, ..., T$. Hence Z_i^A is a strict subset of Z_i, and $DH = H - H^A \approx \chi^2(1)$.

The clienteles theory suggests that term spread may be endogenous, but the

difference-in-Hansen J tests do not indicate that strictly exogenous specification for the term spread is invalid. This is reported in Table B.9.

Table B.9: Difference-in-Hansen J test for strict exogeneity of explanatory variable

	H		
	Endogenous	Exogenous	Difference
cETermSpread	19.63	19.94	0.31
WorkingSavers	18.27	20.47	2.20
ContractSaveAssets	21.63	22.17	0.54
CorpMaturity	18.90	20.80	1.90
RiskSpread	21.10	20.92	-0.17
InflationCredibility	15.37	18.22	2.85
Inflation	21.44	24.58	3.14

None of the $DH = H - H^a$ test statistics exceeds 3.84, the critical value at 5% significance level. Therefore this test does not reject the null hypothesis that strictly exogenous specification is valid.

For further robustness, I also perform the Hausman tests of endogeneity. I compare the moment conditions where the variable is treated as strictly exogenous, given by $E[x_{i,s}\Delta v_{i,t}] = 0 \forall t = 3, ..., T; s = 1, 2, ..., T$ and the less restrictive set of moment conditions where the variable is treated as endogenous, given by $E[x_{i,t-s}\Delta v_{i,t}] = 0 \forall t = 3, ..., T; s = 2, ..., T$. These tests are reported in Table B.10. I find that all but one of the additional moment conditions present in the First Difference GMM model (6)—beyond the moment conditions when the expected long-run term spread is treated as endogenous, one at a time—are valid. As such, I conclude against endogeneity of the expected long-run term spread for the model specification.

Table B.10: Hausman test for strict exogeneity of cETermSpread explanatory variable

	cETermSpread				
	Endogenous		Exogenous		
	beta	$\hat{se2}$	beta	$\hat{se2}$	h
L.GovtMaturity	0.46	0.03	0.62	0.03	-2.00
cETermSpread	-0.18	0.03	-0.22	0.01	0.32
WorkingSavers	0.26	0.01	0.21	0.00	0.89
InflationCredibility	0.00	0.00	0.00	0.00	-0.90
Inflation	0.06	0.00	0.06	0.00	

Only one of the h statisics marginally exceed 1.96, the critical value at 5% significance level. Therefore this test does not reject the null hypothesis that strictly exogenous specification is valid.

B.6 Unit root tests

I apply the Augmented Dickey Fuller (ADF) test to detect presence of unit roots in the variables. ADF test includes the lagged values of the dependent variable, of order d. The model in a test with a constant is $\Delta Z_t = \delta Z_{t-1} + \mu_c + \gamma_d \Delta Z_{t-d}$ and under the null hypothesis, $\delta = \mu_c = 0$, Z_t is non-stationary. Similarly, the model in a test with a constant as well as a trend, is $\Delta Z_t = \delta Z_{t-1} + \mu_c + \mu_l t + \sum_d \gamma_d \Delta Z_{t-d}$ and under the null hypothesis, $\delta = \mu_l = 0$, Z_t is non-stationary. The results of the ADF test and relevant lag selection are presented in Table B.11 and Tables B.12-B.14.

Lag length selection in the ADF unit root tests relies on the Schwarz-Bayes, Hannan-Quin, and Akaike Information Criteria, as well as on the Likelihood Ratio test, as follows. An alternative to selecting d in $ADF(d) = ADF(p-1)$ based on information crieria and likelihood ratio tests is to use the greatest statistically significant lag order as d. The conclusions of ADF test are not sensitive to this alternative. Note that the likelihood ratio and the p-value corresponding to lag l is a test of the null of l lags against an alternative of *l+1* lags.

Table B.11: Unit root tests

		Lags		ADF t-statistic in the deterministic model			Integration
	Sample	p	d	None	Constant	Linear	order
GovtMaturity	1952:Q2-2013:Q3	5	4	-0.6157	-2.093	-2.082	I(1)
D.GovtMaturity		4	3	-5.554**	-5.538**	-5.638**	I(0)
TermPremium	1962:Q3-2016:Q1	1	0	-1.463	-2.471	-2.456	I(1)
D.TermPRemium		0	0	-16.06**	-16.03**	-16.06**	I(0)
PrimeSavers	1955:Q4-2013:Q3	12	11	-1.201	-5.767**	-5.720**	I(2)
D.PrimeSavers		12	11	-0.9679	-0.9131	-0.5729	I(1)
D2.PrimeSavers		11	10	-3.675**	-3.675**	-3.863*	I(0)
TermSpread	1956(2) - 2015(4)	7	6	-2.390*	-3.742**	-4.730**	I(0)
ContractSaveAssets	1972:Q3-2015:Q3	0	0	-11.24**	-11.96**	-11.96**	I(0)
CorpMaturity	1972:Q2-2014:Q4	5	4	-0.8101	-2.881*	-3.27	I(1)
D.CorpMaturity		4	3	-6.304**	-6.292**	-6.280**	I(0)
InflationCredibility	1955:Q3-2014:Q4	7	6	-1.027	-2.325	-2.362	I(1)
D.InflationCredibility		5	4	-7.833**	-7.815**	-7.837**	I(0)
Inflation	1952:Q2-2014:Q4	6	5	-1.193	-2.079	-2.236	I(1)
D.Inflation		6	5	-7.098**	7.080**	-7.085**	I(0)

A small and statistically insignificant ADF t-statistic means that null hypothesis of a unit root cannot be rejected. The ADF unit root test finds that weighted average remaining maturity, term premium, corporate maturity, inflation, and variable capturing inflation credibility are integrated I(1), whereas share of prime savers in the population is integrated I(2).

Table B.12: Lag length selection for unit root tests

	TermPremium:				
Lags	Information Criteria			Likelihood ratio test	
p	BIC	HQC	AIC	Log likelihood	p-value
0	3.176	3.167	3.160	-333.99	0.00**
1	**1.078**	**1.060**	**1.047**	**-108.95**	**0.32**
2	1.099	1.071	1.052	-108.46	0.27
3	1.119	1.081	1.055	-107.85	0.67
4	1.143	1.096	1.064	-107.76	0.30

Lag length of 1 selected for the term premium in levels. Similarly, lag length of 0 selected for its first difference.

Table B.13: Lag length selection for unit root tests

Lags	GovtMaturity: Information Criteria			Likelihood ratio test	
p	BIC	HQC	AIC	Log likelihood	p-value
0	2.670	2.661	2.656	-316.33	0.00**
1	**-1.250**	-1.267	-1.279	154.82	0.124
2	-1.237	-1.263	-1.281	156.02	0.434
3	-1.217	-1.251	-1.275	156.33	0.006**
4	-1.225	-1.269	-1.298	160.10	0.021*
5	-1.225	**-1.277**	**-1.312**	**162.84**	**0.664**
6	-1.203	-1.264	-1.305	162.93	0.286
7	-1.185	-1.255	-1.301	163.52	0.342
8	-1.166	-1.244	-1.297	163.99	

Lag length of 5 selected for the weighted remaining maturity of government debt in levels. Similarly, lag length of 4 selected for its first difference.

Lags	PrimeSavers: Information Criteria			Likelihood ratio test	
p	BIC	HQC	AIC	Log likelihood	p-value
0	4.912	4.903	4.897	-589.10	0.00**
1	-2.226	-2.244	-2.255	273.76	0.00**
2	-6.215	-6.241	-6.259	757.15	0.00**
3	-7.158	-7.192	-7.215	873.46	0.00**
4	-8.718	-8.762	-8.791	1064.28	0.03*
5	-8.715	-8.767	-8.802	1066.59	0.00**
6	-8.742	-8.802	-8.843	1072.56	0.00**
7	-8.849	-8.918	-8.964	1088.19	0.00**
8	-9.712	-9.789	-9.842	1194.94	0.00**
9	-9.829	-9.915	-9.974	1211.82	0.46
10	-9.809	-9.904	-9.968	1212.1	0.79
11	-9.786	-9.890	-9.960	1212.14	0.00**
12	**-10.045**	**-10.157**	**-10.233**	**1246.04**	

Lag length of 12 selected for the share of prime savers. Similarly, lag length of 12 selected for its first difference, and 11 for its second difference.

Table B.14: Lag length selection for unit root tests

ContractSaveAssets:					
Lags	Information Criteria			Likelihood ratio test	
p	BIC	HQC	AIC	Log likelihood	p-value
0	**6.314**	**6.303**	**6.295**	**-534.08**	**0.28**
1	6.337	6.315	6.300	-533.49	0.53
2	6.365	6.332	6.309	-533.29	0.88
3	6.395	6.351	6.321	-533.28	0.96
4	6.425	6.370	6.333	-533.28	0.31

Lag length of 0 selected for the growth rate of contractual savings sector assets.

CorpMaturity:					
Lags	Information Criteria			Likelihood ratio test	
p	BIC	HQC	AIC	Log likelihood	p-value
0	5.681	5.671	5.663	-486.03	0.00**
1	4.512	4.490	4.475	-382.85	0.00**
2	**4.483**	**4.450**	4.428	-377.78	0.24
3	4.504	4.461	4.431	-377.08	0.73
4	4.534	4.479	4.442	-377.02	0.01*
5	4.530	4.465	**4.420**	**-374.12**	**0.52**
6	4.557	4.481	4.429	-373.91	0.82
7	4.587	4.500	4.441	-373.88	0.34
8	4.611	4.514	4.447	-373.41	

Lag length of 5 selected for the maturity of private sector bond issuances in levels, and lag length of 4 for its first difference.

InflationCredibility:					
Lags	Information Criteria			Likelihood ratio test	
p	BIC	HQC	AIC	Log likelihood	p-value
0	0.512	0.504	0.498	254.93	0.00**
1	-1.888	-1.905	-1.917	254.91	0.00**
2	-1.932	-1.958	-1.975	253.53	0.01*
3	-1.937	-1.971	-1.994	250.98	0.00**
4	**-1.960**	**-2.003**	-2.033	250.93	0.76
5	-1.938	-1.990	-2.025	**250.98**	**0.02***
6	-1.937	-1.997	-2.038	253.53	0.10
7	-1.925	-1.994	**-2.041**	254.91	0.83
8	-1.903	-1.980	-2.033	254.93	

Lag length of 7 selected for the inverse debt service variable, which captures inflation credibility. Lag length of 6 selection for its first difference.

Table B.15: Lag length selection for unit root tests

Lags	Inflation: Information Criteria			Likelihood ratio test	
p	BIC	HQC	AIC	Log likelihood	p-value
0	4.610	4.601	4.596	-578.05	0.00**
1	3.330	3.314	3.302	-414.09	0.00**
2	3.211	3.186	3.169	-396.27	0.00**
3	**3.198**	**3.164**	3.142	-391.82	0.79
4	3.219	3.177	3.149	-391.79	0.01*
5	3.215	3.165	3.131	**-388.54**	**0.05**
6	3.223	3.164	**3.125**	-386.70	0.19
7	3.238	3.171	3.126	-385.82	0.33
8	3.256	3.180	3.130	-385.35	

Lag length of 6 selected for inflation in levels as well as for the first difference.

B.7 Bivariate cointegration tests

I test whether maturity and other $I(1)$ explanatory variables follow a common stochastic trend in a bivariate setting by using the Johansen (1991) trace test. This test is based on $rank(\Pi) = r$ of the matrix $\Pi = \alpha\beta'$ in estimated $VAR(q)$:

$$\Delta Y_t = \Gamma_1 \Delta Y_{t-1} + ... + \Gamma_q \Delta Y_{t-q+1} + \Pi Y_{t-q} + \mu + u_t$$

where $Y = [y_1, y_2]$ and $r = [1, 2]$ for the bivariate case. The standard distributional assumptions apply. These are: (i) $u_1, ..., u_T$ are mutually independent, given Y_0; (ii) $u_T =$ in distribution $N(0, \Omega)$; (iii) parameter space is such that Ω is positive-definite.

In order to determine the lag in the $VAR(q)$, I apply the information criteria and likelihood ratio tests for the relevant multivariate models.

Lag length selection in the Johansen trace tests above relies on the Schwarz-Bayes, Hannan-Quin, and Akaike Information Criteria, as well as on the Likelihood Ratio test, as follows. Note that the likelihood ratio and the p-value corresponding to lag l is a test of the null of l lags against an alternative of *l+1* lags.

Table B.16: Johansen trace tests for bivariate cointegration

		Lags	Ho: rank=0, Ha: rank≥1		Ho: rank≤1, Ha: rank=2		Bivariate
	Sample	q	Statistic	p-value	Statistic	p-value	cointegration
GovtMaturity, TermPremium	1962:Q3-2013:Q3	5	13.767	0.089	4.5089	0.034*	no
GovtMaturity, D.PrimeSavers	1955:Q4-2013:Q3	12	12.37	0.141	1.4036	0.236	not clear
GovtMaturity, CorpMaturity	1971:Q2:2013:Q3	5	15.759	0.044*	2.566	0.109	yes
GovtMaturity, InflationCredibility	1955:Q3-2013:Q3	9	11.64	0.177	2.7034	0.1	no
GovtMaturity, Inflation	1951:Q4:2013:Q3	7	21.192	0.005**	3.569	0.059	yes
Inflation, CorpMaturity	1971:Q2-2014:Q4	5	13.113	0.111	4.0161	0.045*	no
Inflation, TermPremium	1963:Q3-2015:Q3	3	20.305	0.008**	6.3327	0.012*	no
Inflation, D.PrimeSavers	1953:Q4-2013:Q3	12	18.64	0.015*	3.2126	0.073	yes

Johansen trace test using the hypotheses above. With the bivariate case, when we can reject the hypothesis of rank=0 in favour of the alternative, that rank=1, but cannot reject it in favour of the alternative that rank=2, this indicates that the two series are cointegrated. This is the case for: weighted average remaining maturity of government debt and weighted average maturity of corporate debt; weighted average remaining maturity of government debt and inflation; inflation and change in the share of prime savers in total population.

Table B.17: Lag length selection for cointegration test

Lags	GovtMaturity, TermPremium: Information Criteria			Likelihood ratio test	
q	BIC	HQC	AIC	Log likelihood	p-value
0	5.717	5.697	5.684	-560.72	0.00**
1	**-0.548**	-0.608	-0.648	70.15	0.00**
2	-0.512	-0.610	-0.678	77.08	0.63
3	-0.418	-0.556	-0.651	78.40	0.01*
4	-0.373	-0.551	-0.672	84.56	0.00**
5	-0.414	**-0.632**	**-0.780**	99.18	**0.60**
6	-0.322	-0.579	-0.754	100.64	0.87
7	-0.222	-0.519	-0.720	101.30	0.52
8	-0.133	-0.469	-0.698	103.05	

Lag length of 5 selected for the bivariate VAR of weighted average remaining maturity and term premium.

Lags	GovtMaturity, D.PrimeSavers: Information Criteria			Likelihood ratio test	
q	BIC	HQC	AIC	Log likelihood	p-value
0	0.225	0.208	0.196	-20.68	0.00**
1	-7.469	-7.522	-7.558	882.75	0.000**
2	-8.408	-8.496	-8.556	1002.51	0.000**
3	-9.828	-9.952	-10.036	1178.18	0.002**
4	-9.809	-9.969	-10.077	1186.90	0.000*
5	-9.826	-10.021	-10.152	1199.68	0.000*
6	-9.899	-10.130	-10.285	1219.10	0.000*
7	-10.595	-10.861	-11.041	1310.72	0.000*
8	-10.713	-11.015	-11.219	1335.36	0.097
9	-10.656	-10.993	-11.221	1339.63	0.829
10	-10.570	-10.942	-11.194	1340.45	0.000*
11	-10.690	-11.098	-11.374	1365.36	0.000*
12	**-10.730**	**-11.173**	**-11.472**	1380.81	**0.336**
13	-10.658	-11.137	-11.460	1383.39	0.635
14	-10.577	-11.091	-11.438	1384.85	

Lag length of 12 selected for the bivariate VAR of weighted average remaining maturity and change in the share of prime savers in the population.

Table B.18: Lag length selection for cointegration test

Lags	GovtMaturity, CorpMaturity: Information Criteria			Likelihood ratio test	
q	BIC	HQC	AIC	Log likelihood	p-value
0	8.274	8.251	8.236	-669.23	0.00**
1	2.668	2.600	2.554	-202.16	0.00**
2	**2.587**	**2.474**	2.397	-185.36	0.44
3	2.688	2.530	2.422	-183.39	0.04*
4	2.750	2.547	2.409	-178.30	0.00**
5	2.735	2.487	**2.318**	-166.88	**0.37**
6	2.832	2.539	2.339	-164.59	0.99
7	2.956	2.617	2.386	-164.47	0.31
8	3.048	2.665	2.403	-161.81	

Lag length of 5 selected for the bivariate VAR of weighted average remaining maturity and maturity of private sector bond issuances.

Lags	GovtMaturity, InflationCredibility: Information Criteria			Likelihood ratio test	
q	BIC	HQC	AIC	Log likelihood	p-value
0	3.078	3.060	3.048	-348.56	0.00**
1	-3.169	-3.223	-3.259	380.77	0.00**
2	-3.191	-3.280	-3.341	394.17	0.00**
3	-3.158	-3.283	-3.368	401.28	0.00**
4	**-3.253**	-3.413	-3.522	423.00	0.01*
5	-3.218	**-3.415**	**-3.547**	429.93	0.63
6	-3.136	-3.367	-3.524	431.28	0.20
7	-3.069	-3.336	-3.517	434.45	0.96
8	-2.977	-3.280	-3.485	434.76	0.01*
9	-2.941	-3.280	-3.509	441.59	**0.17**
10	-2.877	-3.252	-3.505	445.10	0.11
11	-2.819	-3.229	-3.506	449.21	0.63
12	-2.736	-3.182	-3.484	450.64	

Lag length of 9 selected for the bivariate VAR of weighted average remaining maturity and variable capturing inflation credibility as the inverse of the effective debt service.

Table B.19: Lag length selection for cointegration test

Lags	GovtMaturity, Inflation: Information Criteria			Likelihood ratio test	
q	BIC	HQC	AIC	Log likelihood	p-value
0	6.755	6.737	6.726	-815.19	0.00**
1	1.971	1.919	1.884	-222.94	0.00**
2	**1.944**	1.858	1.800	-208.73	0.19
3	2.009	1.889	1.807	-205.60	0.00**
4	2.041	1.886	1.782	-198.50	0.00**
5	2.043	**1.854**	**1.726**	-187.74	0.90
6	2.129	1.905	1.755	-187.2	0.02*
7	2.171	1.913	1.739	-181.32	**0.75**
8	2.253	1.961	1.764	-180.32	0.47
9	2.327	2.001	1.781	-178.41	0.56
10	2.405	2.044	1.801	-176.8	

Lag length of 7 selected for the bivariate VAR of weighted average remaining maturity and inflation.

Lags	Inflation, CorpMaturity: Information Criteria			Likelihood ratio test	
q	BIC	HQC	AIC	Log likelihood	p-value
0	10.433	10.411	10.396	-871.28	0.00**
1	**7.610**	7.543	7.498	-623.84	0.00**
2	7.616	**7.505**	7.430	-614.09	0.48
3	7.716	7.561	7.456	-612.28	0.18
4	7.799	7.600	7.464	-609.00	0.00**
5	7.830	7.587	**7.421**	-601.33	**0.79**
6	7.941	7.653	7.457	-600.41	0.23
7	8.026	7.695	7.468	-597.35	0.06
8	8.089	7.713	7.456	-592.34	

Lag length of 5 selected for the bivariate VAR of inflation and maturity of private sector bond issuances.

Table B.20: Lag length selection for cointegration test

Lags	Inflation, TermPremium: Information Criteria			Likelihood ratio test	
q	BIC	HQC	AIC	Log likelihood	p-value
0	7.768	7.748	7.735	-794.72	0.00**
1	**4.241**	4.183	4.144	-420.80	0.00**
2	4.260	**4.164**	4.099	-412.16	0.03*
3	4.312	4.177	**4.086**	-406.85	**0.37**
4	4.394	4.221	4.103	-404.61	0.06
5	4.452	4.241	4.097	-399.99	0.35
6	4.533	4.283	4.113	-397.64	0.64
7	4.623	4.335	4.139	-396.29	0.20
8	4.696	4.369	4.146	-393.08	

Lag length of 3 selected for the bivariate VAR of inflation and term premium.

Lags	Inflation, D.PrimeSavers: Information Criteria			Likelihood ratio test	
q	BIC	HQC	AIC	Log likelihood	p-value
0	2.399	2.382	2.370	-272.88	0.00**
1	-2.872	-2.925	-2.961	349.46	0.00**
2	-3.901	-3.990	-4.050	479.74	0.01*
3	5.377	-5.502	-5.585	661.90	0.00**
4	-5.340	-5.499	-5.607	668.40	0.00**
5	-5.342	-5.538	-5.669	679.64	0.00**
6	-5.397	-5.628	-5.784	696.88	0.00**
7	-6.156	-6.421	-6.601	795.74	0.00**
8	-6.244	-6.546	-6.749	816.93	0.16
9	-6.181	-6.518	-6.746	820.48	0.88
10	-6.093	-6.465	-6.717	821.11	0.00**
11	-6.262	-6.670	-6.946	851.68	0.00**
12	**-6.291**	**-6.734**	**-7.033**	865.86	**0.70**
13	-6.207	-6.686	-7.010	867.11	0.29
14	-6.138	-6.652	-7.000	869.94	

Lag length of 12 selected for the bivariate VAR of inflation and change in the share of prime savers in the population.

B.8 Trivariate cointegration test

Specifications from bivariate setting easily extend to multivariate setting, with the main difference that several cointegrating relations are possible.

Table B.21: Lag length selection for cointegration test

Lags	GovtMaturity, Inflation, D.PrimeSavers: Information Criteria			Likelihood ratio test	
q	BIC	HQC	AIC	Log likelihood	p-value
0	4.356	4.330	4.312	1095.69	0.00**
1	-4.164	-4.268	-4.338	349.46	0.00**
2	-5.096	-5.277	-5.400	479.74	0.01**
3	-6.452	-6.711	-6.887	661.90	0.00*
4	-6.334	-6.672	-6.900	668.40	0.00**
5	-6.274	-6.690	-6.970	679.64	0.00**
6	-6.235	-6.728	-7.062	696.88	0.00**
7	-6.908	-7.479	-7.865	795.74	0.00**
8	**-6.908**	**-7.558**	-7.996	816.93	0.16
9	-6.749	-7.476	-7.967	820.48	0.88
10	-6.577	-7.382	-7.925	821.11	0.00**
11	-6.647	-7.531	-8.127	851.68	0.00**
12	-6.596	-7.557	**-8.206**	**865.86**	

Using the full sample,1952:Q4-2013:Q3. Lag length of either 8 or 12 selected for the trivariate VAR of government debt maturity, inflation and change in the share of prime savers in the population.

Table B.22: Johansen trace test for cointegration of maturity, inflation and change in the share of prime savers. This is for the alternative sample period, 1952:Q4-2013:Q3.

Lags	Ho: rank=0, Ha: rank¿=1		Ho: rank¡=1, Ha: rank¿=2		Ho: rank¡=2, Ha: rank=3		Single trivariate
q	Statistic	p-value	Statistic	p-value	Statistic	p-value	cointegration
8	33.644	0.016*	7.926	0.481	2.5112	0.113	yes
12	35.692	0.009**	7.8239	0.492	2.5964	0.107	yes

The corresponding error correction mechanisms are:
$WARM_t - \underset{(2.169)}{3.113}\Delta ps_t + \underset{(0.080)}{0.569}\pi_t$ for $VAR(8)$ and
$WARM_t - \underset{(2.118)}{2.602}\Delta ps_t + \underset{(0.054)}{0.645}\pi_t$ for $VAR(12)$.
This shows that there is a similarly strong single cointegrating vector as for the main sample period, 1971:Q2-2013:Q3, and parameter values are similar and with the same signs. This confirms a stable single cointegrating relationship over time period when data is available.

Table B.23: Lag length selection for cointegration test

Lags	GovtMaturity, Inflation, D.PrimeSavers: Information Criteria			Likelihood ratio test	
q	BIC	HQC	AIC	Log likelihood	p-value
0	4.456	4.423	4.400	-371.02	0.00**
1	-4.431	-4.562	-4.652	407.44	0.00**
2	-5.411	-5.641	-5.798	513.85	0.00**
3	-6.699	-7.028	-7.253	646.49	0.02*
4	-6.549	-6.976	-7.268	656.78	0.00**
5	-6.509	-7.035	-7.394	676.48	0.00**
6	-6.436	-7.061	-7.487	693.42	0.00**
7	**-7.027**	-7.750	-8.244	766.73	0.00**
8	-6.972	**-7.794**	-8.356	785.23	0.13
9	-6.797	-7.717	-8.346	793.42	0.46
10	-6.587	-7.607	-8.303	798.73	0.00**
11	-6.641	-7.759	-8.523	826.45	0.02*
12	-6.512	-7.728	**-8.559**	**838.53**	**0.77**
13	-6.284	-7.599	-8.497	842.26	0.81
14	-6.053	-7.467	-8.433	845.80	

Using the main sample,1972:Q1-2013:Q3. Lag length of either 8 or 12 selected for the trivariate VAR of government debt maturity, inflation and change in the share of prime savers in the population.

B.9 Alternative cointegration

Another variable which stands for the inflation credibility determinant—real debt service—appears to have the same stochastic trend as government debt maturity and change in the share of prime savers, as seen in 4.3. Indeed it is possible to also include real debt service in this ECM. However, there isn't always a robust single cointegrating vector, as shown in Table B.24. Furthermore, the positive relationship is contrary to inflation credibility theory which specifies that government debt maturity should be a negative function of real debt service. To keep the model simple, real debt service was excluded from the ECM.

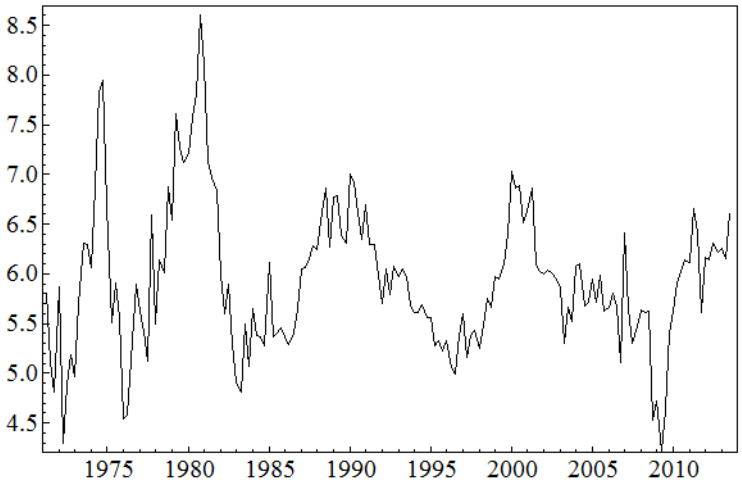
Table B.24: Johansen trace tests for multivariate cointegration between government debt maturity, inflation, change in the share of prime savers in the population, and real government debt service

Lags	Ho: rank=0, Ha: rank≥1		Ho: rank≤1, Ha: rank≥2		Ho: rank≤2, Ha: rank≥3		Ho: rank≤3, Ha: rank=4		Single cointegrating
q	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	relation
8	53.382	0.013*	26.945	0.105	8.604	0.410	0.411	0.522	yes
12	63.644	0.001**	31.495	0.031*	8.331	0.438	0.064	0.801	no

Lags lengths are based on the information criteria and likelihood ratio tests. The trace test using the hypotheses above suggests that there may be a single cointegrating relationship, but it is not robust to presence of second cointegrating vector as well. The sample period is 1970:Q2-2013:Q3.

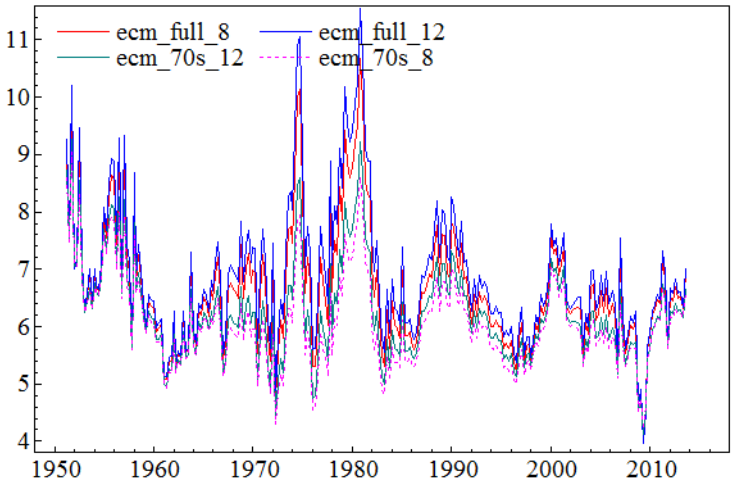
B.10 Equilibrium correction mechanisms

Figure B.1: Equilibrium Correction Mechanism used, for 1972:Q1-2013:Q3, based on $VAR(8)$



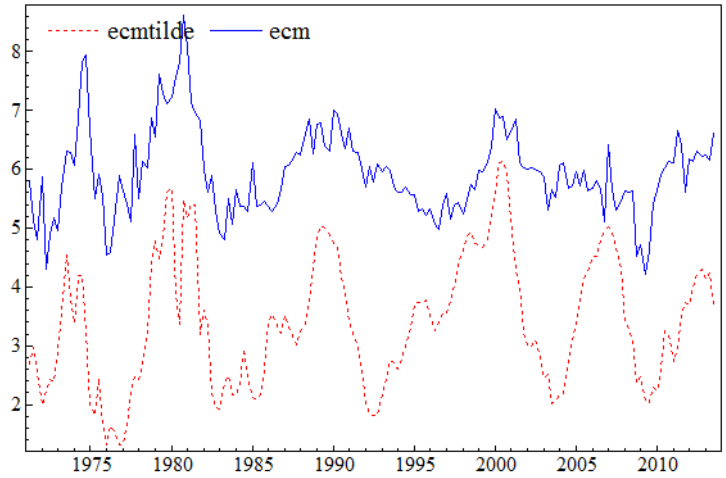
Vector normality is rejected based on $\chi^2(6) = 107.64$ and vector homoscedasticity is rejected based on $F(288, 700) = 1.4692$.

Figure B.2: Comparison of alternative Equilibrium Correction Mechanisms



The chosen ECM is robust to alternative lag specifications and estimating it using a different time period.

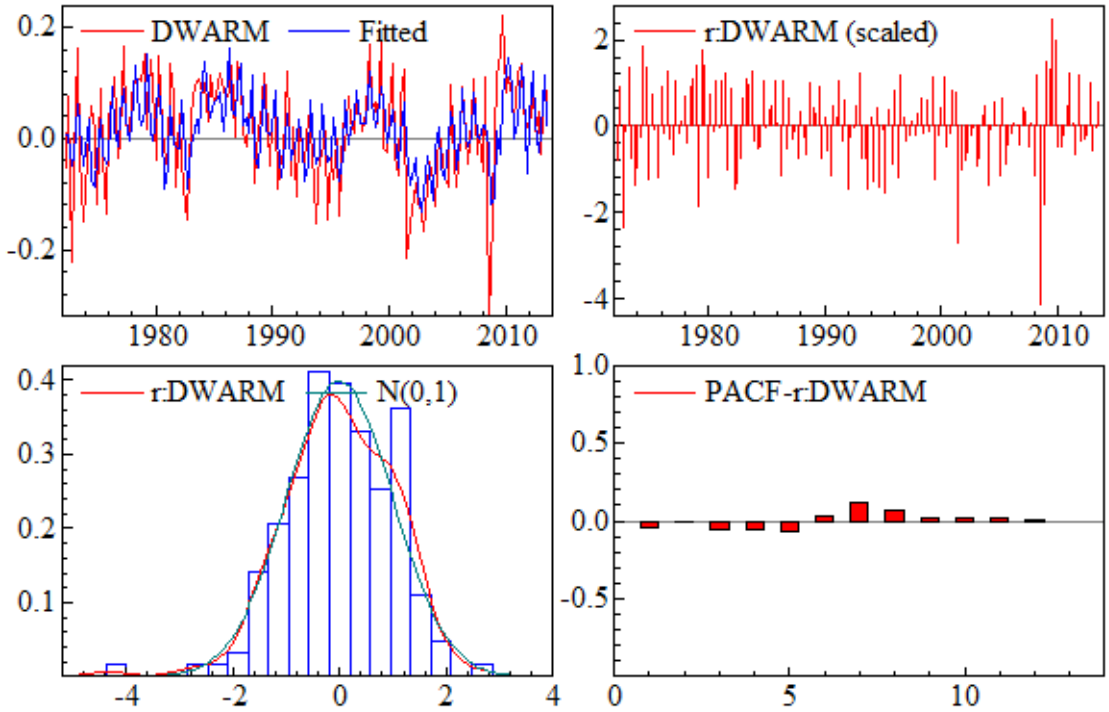
Figure B.3: Comparison of alternative Equilibrium Correction Mechanisms



Two vectors are correlated.

B.11 Specification tests

Figure B.4: Residuals of the single equation model



There is no serial correlation in the residuals, while non-Normality is negligible and heteroscedasticity is limited.

Bibliography

Abbas, S. M. A., L. Blattner, M. D. Broeck, A. El-Ganainy, and M. Hu (2014). Sovereign Debt Composition in Advanced Economies: A Historical Perspective. *IMF Working Papers* (162).

Aboody, D. and B. Lev (2000). Information Asymmetry, R & D, and Insider Gains. *The Journal of Finance 55*(6), 2747–2766.

Adams, R. B. and D. Ferreira (2009). Women in the boardroom and their impact on governance and performance. *Journal of Financial Economics 94*(2), 291–309.

Adams, R. B. and V. Ragunathan (2017). Lehman Sisters.

Adrian, T., R. K. Crump, and E. Moench (2013). Pricing the term structure with linear regressions. *Journal of Financial Economics 110*(April), 110–138.

Agarwal, R. and M. Gort (2002). Firm and Product Life Cycles and Firm Survival. *American Economic Review 92*(2), 184–190.

Agarwal, V., W. Jiang, Y. Tang, and B. Yang (2013). Uncovering Hedge Fund Skill from the Portfolio Holdings They Hide. *The Journal of Finance 68*(2), 739–783.

Agrawal, A., C. Catalini, and A. Goldfarb (2014). Some Simple Economics of Crowdfunding. In J. Lerner and S. Stern (Eds.), *Innovation Policy and the Economy*, Volume 14, pp. 63–97. University of Chicago Press Volume.

Agrawal, A., C. Catalini, and A. Goldfarb (2015). Crowdfunding: Geography, Social Networks, and the Timing of Investment Decisions. *Journal of Economics and Management Strategy 24*(2), 253–274.

Ahlers, G. K., D. Cumming, C. Günther, and D. Schweizer (2015). Signaling in Equity Crowdfunding. *Entrepreneurship: Theory and Practice 39*(4), 955–980.

Allen, W. (2013). Quantitative monetary policy and government debt management in Britain since 1919. *Oxford Review of Economic Policy 28*(4), 804–836.

Alloway, T. (2011, apr). The importance of debt maturity.

Andrei, D. and J. Cujean (2017). Information percolation, momentum and reversal. *Journal of Financial Economics 123*(3), 617–645.

Angeletos, G.-m. (2002). Fiscal Policy with Noncontingent Debt and the Optimal Maturity Structure. *The Quarterly Journal of Economics* (August), 1105–1131.

Åstebro, T., M. Fernández, S. Lovo, and N. Vulkan (2017). Herding in Equity Crowdfunding.

Atkinson, S. M., S. B. Baird, and M. B. Frye (2003). Do female mutual fund managers manage differently? *Journal of Financial Research 26*(1), 1–18.

Baker, M., R. Greenwood, and J. Wurgler (2002). The maturity of debt issues and predictable variation in bond returns. *Journal of Financial Economics 70*(2), 261–291.

Banerjee, S. (2011). Learning from prices and the dispersion in beliefs. *Review of Financial Studies 24*(9), 3025–3068.

Banerjee, S., R. Kaniel, and I. Kremer (2009). Price drift as an outcome of differences in higher-order beliefs. *Review of Financial Studies 22*(9), 3707–3734.

Banerjee, S. and I. Kremer (2010). Disagreement and learning: Dynamic patterns of trade. *Journal of Finance 65*(4), 1269–1302.

Barber, B. M. and T. Odean (2001). Boys Will be Boys: Gender, Overconfidence, and Common Stock Investment. *The Quarterly Journal of Economics 116*(1), 261–292.

Barber, B. M., T. Odean, B. M. Barber, and T. Odean (2008). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *The Review of Financial Studies 21*(2), 785–818.

Baron, R. M. and D. a. Kenny (1986). The Moderator-Mediator Variable Distinction in Social The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology 51*(6), 1173–1182.

Barro, R. (1979). On the Determination of the Public Debt. *Journal of Political Economy 87*(5), 940–971.

Barro, R. (2002). Optimal Management of Indexed and Nominal Debt. In F. Lefort and K. Schmidt-Hebbel (Eds.), *Indexation, Inflation, and Monetary Policy*, pp. 135–150. Central Bank of Chile.

Barro, R. J. (1974). Are Government Bonds Net Wealth? *Journal of Political Economy 82*(6), 1095–1117.

Basak, S. (2005). Asset pricing with heterogeneous beliefs. *Journal of Banking and Finance 29*(11), 2849–2881.

Bates, J. M. and C. W. J. Granger (1969). The Combination of Forecasts. *Journal of the Operational Research Society 20*(4), 451–468.

Becker, G. S. (1992). Fertility and the Economy. *Journal of Population Economics 5*(3), 185–201.

Belleflamme, P., T. Lambert, and A. Schwienbacher (2014). Crowdfunding: Tapping the right crowd. *Journal of Business Venturing 29*(5), 585–609.

Bernstein, S., A. Korteweg, and K. Laws (2017). Attracting Early-Stage Investors: Evidence from a Randomized Field Experiment. *Journal of Finance 72*(2), 509–538.

Bird, M. (2016, sep). British Pension Funds British pension Funds Hunger for Superlong Bonds.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*(1), 115–143.

Bohn, H. (1990). Tax Smoothing with Financial Instruments. *American Economic Review 80*(5), 1217–1230.

Bolder, D. J. (2008). The Canadian Debt-Strategy Model. *Bank of Canada Review* (3), 3–16.

Bottazzi, L., M. Da Rin, and T. Hellmann (2008). Who are the active investors? Evidence from venture capital. *Journal of Financial Economics 89*(3), 488–512.

Bradshaw, M. T. (2011). Analysts' Forecasts: What Do We Know after Decades of Work?

Broeck, M. D. and A. Guscina (2011). Government Debt Issuance in the Euro Area: The Impact of the Financial Crisis. *IMF Working Papers* (21).

Budescu, D. V. and E. Chen (2015). Identifying Expertise to Extract the Wisdom of Crowds. *Management Science 61*(2), 267–280.

Buera, F. and J. P. Nicolini (2004). Optimal maturity of government debt without state contingent bonds. *Journal of Monetary Economics 51*(3), 531–554.

Buraschi, A. and A. Jiltsov (2006). Model uncertainty and option markets with heterogeneous beliefs. *Journal of Finance 61*(6), 2841–2897.

Buraschi, A., I. Piatti, and P. Whelan (2017). Expected Term Structures.

Cardona, J. C., R. Stanojevic, and R. Cuevas (2013). On Weather and Internet Traffic Demand. In M. Roughan and R. Chang (Eds.), *Passive and Active Measurement*, Volume 7799, pp. 260–263. Hong Kong: Springer.

Chadha, J., P. Turner, and F. Zampolli (2013). The Interest Rate Effects of Government Debt Maturity. *BIS Working Paper* (415).

Chalmers, J., A. Kaul, and B. Phillips (2013). The wisdom of crowds: Mutual fund investors' aggregate asset allocation decisions. *Journal of Banking and Finance 37*(9), 3318–3333.

Chemla, G. and K. Tinn (2017). Learning Through Crowdfunding.

Chen, H., V. Curdia, and A. Ferrero (2012). The Macroeconimic Effects of Large Scale Asset Purchase Programmes. *The Economic Journal 122*(September 2011), F289–F315.

Chen, L., Z. Huang, and D. Liu (2016). Pure and hybrid crowds in crowdfunding markets. *Financial Innovation 2*(1), 19.

Cohen, L. and A. Frazzini (2008). Economic links and predictable returns. *The Journal of Finance 63*(4), 1977–2011.

Cohen, L., A. Frazzini, and C. Malloy (2010). Sell-Side School Ties. *The Journal of Finance 65*(4), 1409–1437.

Cohen, L. and D. Lou (2012). Complicated F irms. *Journal of Financial Economics 104*(2), 383–400.

Cohen, L., D. Lou, and C. Malloy (2016). Cloaked Trading. *Journal of Investment Consulting 17*(2), 69–80.

Coibion, O. and Y. Gorodnichenko (2015). Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts. *American Economic Review 105*(8), 2644–2678.

Coleman, S. and A. Robb (2009). A comparison of new firm financing by gender: Evidence from the Kauffman Firm Survey data. *Small Business Economics 33*(4), 397–411.

Colombo, M. G. and L. Grilli (2005). Founders' human capital and the growth of new technology-based firms: A competence-based view. *Research Policy 34*(6), 795–816.

Cremers, K. J. and A. Petajisto (2009). How Active Is Your Fund Manager A New Measure That Predicts Performance. *Review of Financial Studies 22*(9), 3329–3365.

Croson, R. and U. Gneezy (2009). Gender Differences in Preferences. *Journal of Economic Literature 47*(2), 448–474.

CrowdCube (2018). How should I set my valuation?

Culbertson, J. (1957). The term structure of interest rates. *Quarterly Journal of Economics CXVII*(4), 735–775.

Da Rin, M., T. F. Hellmann, and M. Puri (2013). A survey of venture capital research.pdf. In *Handbook of the Economics of Finance*, Volume 2A, pp. 573–648.

D'Amico, S. and T. B. King (2013). Flow and stock effects of large-scale treasury purchases: Evidence on the importance of local supply. *Journal of Financial Economics 108*(2), 425–448.

De Bondt, W. F. M. and R. H. Thaler (1990). Do Security Analysts Overreact? *The American Economic Review 80*(2).

Debortoli, D., R. Nunes, and P. Yared (2014). Optimal Government Debt Maturity under Limited Commitment.

Dellavigna, S., J. M. Pollet, J. Campbell, D. Card, R. Chetty, J. Choi, K. Daniel, Y. Ertimur, J. Graham, D. Hirshleifer, W. Jiang, L. Katz, D. Laibson, O. Lamont, U. Malmendier, M. Nondorf, A. Pollet, A. Poteshman, T. Persson, M. Rabin, J. Stein, X.-J. Zhang, J. Chan, E. Fleekop, R. Kim, C. Li, M. Mai, R. Son, M. Stone, and T. Yee (2008). Investor Inattention and Friday Earnings Announcements. *The Journal of Finance 64*, 709–749.

Diamond, D. (1984). Financial intermediation and delegated monitoring. *Review of Economic Studies 51*(3), 393–414.

Diamond, D. W. and R. E. Verrecchia (1981). Information aggregation in a noisy rational expectations economy. *Journal of Financial Economics 9*(3), 221–235.

Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. G. Marquéz, B. Gruber, B. Lafourcade, P. J. Leitão, T. Münkemüller, C. McClean, P. E. Osborne, B. Reineking, B. Schröder, a. K. Skidmore, D. Zurell, and S. Lautenbach (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography 36*(1), 27–46.

Dottori, D. and M. Manna (2016). Strategy and tactics in public debt management. *Journal of Policy Modeling 38*(1), 1–25.

Duffie, B. D. and G. Manso (2007). Information Percolation in Large Markets. *The American Economic Review 97*(2), 203–209.

Duffie, D., S. Malamud, and G. Manso (2009). Information Percolation With Equilibrium Search Dynamics. *Econometrica 77*(5), 1513–1574.

Easterwood, J. C. and S. R. Nutt (1999). Inefficiency in analysts' earnings forecasts: Systematic misreaction or systematic optimism? *Journal of Finance 54*(5), 1777–1797.

Eckel, C. C. and P. J. Grossman (2008). Men, Women and Risk Aversion: Experimental Evidence. In *Handbook of Experimental Economics Results*, Volume 1, pp. 1061–1073.

Eichengreen, B. and R. Hausmann (1999). Exchange rates and financial fragility. *NBER Working Paper Series* (7418).

Ellison, M. and A. Tischbirek (2014). Unconventional government debt purchases as a supplement to conventional monetary policy. *Journal of Economic Dynamics and Control 43*, 199–217.

Ellman, M. and S. Hurkens (2015). Optimal Crowdfunding Design.

Estrin, S., D. Gozman, and S. Khavul (2017). Equity Crowdfunding and Early Stage Entrepreneurial Finance: Damaging or Disruptive?

Estrin, S. and S. Khavul (2015). Equity Crowdfunding and the socialization of entrepreneurial finance.

Ewens, M. and R. R. Townsend (2018). Are Early Stage Investors Biased Against Women?

Fairlie, R. W. and A. M. Robb (2009). Gender differences in business performance: Evidence from the characteristics of business owners survey. *Small Business Economics 33*(4), 375–395.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance 25*(2).

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Faraglia, E., A. Marcet, and A. Scott (2010). In search of a theory of debt management. *Journal of Monetary Economics 57*(7), 821–836.

Fischer, S. (1983). Welfare aspects of government issue of indexed bonds. In R. Dornvusch and M. H. Simonsen (Eds.), *Inflation, Debt and Indexation*. Cambridge, MA: MIT Press.

Francis, J., D. Philbrick, J. Francis, and D. Philbrickt (1993). Analysts' Decisions As Products of a Multi-Task Environment. *Journal of Accounting Research 31*(2), 216–230.

Frazzini, A., R. Israel, and T. J. Moskowitz (2012). Trading costs of asset pricing anomalies.

Fried, D. and D. Givoly (1982). Financial analysts forecasts of earnings: A better surrogate for earnings expectations. *Journal of Accounting and Economics 4*(2), 85–107.

Friedman, M. (1960). *A program for monetary stability*, Volume 541. New York: Fordham University Press.

Froot, K. A. (1989). New Hope for the Expectations Hypothesis of the Term Structure of Interest Rates. *The Journal of Finance 44*(2), 283–305.

Gagnon, J., M. Raskin, J. Remache, and B. Sack (2011). The financial market effects of the federal reserve's large-scale asset purchases. *International Journal of Central Banking 7*(1), 3–43.

Gale, D. (1990). The efficient design of public debt. In R. Dornbusch and M. Draghi (Eds.), *Public Debt Management: Theory and History*. Cambridge: Cambridge University Press.

Galton, F. (1907). Vox Populi (The Wisdom of Crowds). *Nature 75*, 450–451.

Genre, V., G. Kenny, A. Meyler, and A. Timmermann (2013). Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting 29*(1), 108–121.

Gilchrist, D. S. and E. G. Sands (2016). Something to Talk About: Social Spillovers in Movie Consumption. *Journal of Political Economy 124*(5), 1339–1382.

Gimmon, E. and J. Levie (2010). Founder's human capital, external investment, and the survival of new high-technology ventures. *Research Policy 39*(9), 1214–1226.

Gompers, P., A. Kovner, J. Lerner, and D. Scharfstein (2010). Performance persistence in entrepreneurship. *Journal of Financial Economics 96*(1), 18–32.

Gompers, P. A. and S. Q. Wang (2017a). And the Children Shall Lead: Gender Diversity and Performance in Venture Capital.

Gompers, P. A. and S. Q. Wang (2017b). Diversity in innovation.

Gonzalez-Fernandez, M. and C. Gonzalez-Velasco (2014). Effects of Socio-Demographic Factors on the Maturity Structure of Sovereign Debt.

Goudswaard, K. (1990). Determinants of Public Debt Maturities. *De Economist 138*(1), 33–46.

Greenwood, R., S. Hanson, and J. Stein (2010). A GapFilling Theory of Corporate Debt Maturity Choice. *The Journal of Finance LXV*(3).

Greenwood, R. and D. Vayanos (2010). Price Pressure in the Government Bond Market. *American Economic Review 100*(2), 585–590.

Greenwood, R. and D. Vayanos (2014). Bond supply and excess bond returns. *Review of Financial Studies 27*(3), 663–713.

Grossman, S. J. (1981). An Introduction to the Theory of Rational Expectations Under Asymmetric Information. *Review of Economic Studies 48*(4), 541–559.

Grossman, S. J. and J. E. Stiglitz (1980). On the Impossibility of Informationally Efficient Markets. *The American Economic Review 74*(3), 433–444.

Guibaud, S., Y. Nosbusch, and D. Vayanos (2013). Bond market clienteles, the yield curve, and the optimal maturity structure of government debt. *Review of Financial Studies 26*(8), 1913–1961.

Halmari, E., E. Sharman, H. Whorwood, J. Ford, T. Nyamadzawo, E. Simoes, D. Osei, and T. Sheils (2017). The Deal: Equity investment in the UK 2017. Technical report, Beauhurst, London.

Harrison, J. M. and D. M. Kreps (1978). Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations. *The Quarterly Journal of Economics 92*(2), 323–336.

Havrylchyk, O. and M. Verdier (2018). The Financial Intermediation Role of the P2P Lending Platforms. *Comparative Economic Studies 60*(1), 115–130.

Hayek, F. A. (1945). The use of knowledge in society. *The American Economic Review 35*(4), 519–530.

Hayes, A. F. (2014). The Simple Mediation Model. In *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*, pp. 85–122.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica 47*(1), 153–161.

Hellmann, T., A. Admati, B. Barnett, C. Barry, J. Brander, B. Cassiman, D. Duffie, P. Gompers, S. Kaplan, J. Lerner, R. Mcdonald, P. Pfleiderer, W. Sahlman, A. Shepard, and J. Van (2000). Market and Financing Strategy: The Role of Venture Capital. *The Review of Financial Studies 13*(4), 959–984.

Hellmann, T. and M. Puri (2002). Venture Capital and the Professionalization of Start-up Firms: Empirical Evidence. *The Journal of Finance 57*(1), 169–197.

Hervé, F. (2017). Determinants of Individual Investment Decisions in Investment-Based Crowdfunding.

Hilary, G. and C. Hsu (2013). Analyst Forecast Consistency. *Journal of Finance 68*(1), 271–297.

Hong, H. and J. D. Kubik (2003). Analyzing the Analysts: Career Concerns and Biased Earnings Forecasts. *The Journal of Finance 58*(1), 313–351.

Hong, H., J. D. Kubik, and J. C. Stein (2005). Thy Neighbor's Portfolio: Word-of-Mouth Effects in the Holdings and Trades of Money Managers. *The Journal of Finance 60*(6), 2801–2824.

Hong, H. and J. C. Stein (1999). A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets. *The Journal of Finance 54*(6), 2143–2184.

Hong, L. and S. Page (2009). Interpreted and generated signals. *Journal of Economic Theory 144*(5), 2174–2196.

Hornuf, L. and A. Schwienbacher (2016). Crowdinvesting: angel investing for the masses? *Handbook of Research on Business Angels* (September 2016), 381–398.

Hsu, D. H. (2007). Experienced entrepreneurial founders, organizational capital, and venture capital funding. *Research Policy 36*(5), 722–741.

Huang, J. and D. J. Kisgen (2013). Gender and corporate finance: Are male executives overconfident relative to female executives? *Journal of Financial Economics 108*(3), 822–839.

Huang, L. and H. Liu (2007). Rational Inattention and Portfolio Selection. *The Journal of Finance 62*(4), 1999–2040.

Huang, S., B.-H. Hwang, and D. Lou (2016). The Speed of Communication.

Ivkovic, Z. and N. Jegadeesh (2004). The timing and value of forecast and recommendation revisions. *Journal of Financial Economics 73*(3), 433–463.

Jegadeesh, N., J. Kim, S. D. Krische, and C. M. C. Lee (2004). Analyzing the Analysts: When Do Recommendations Add Value? *Journal of Finance 59*(3), 1083–1124.

Jegadeesh, N. and W. Kim (2010). Do analysts herd? An analysis of recommendations and market reactions. *Review of Financial Studies 23*(2), 901–937.

Jensen, M. (1978). Some Anamalous Evidence Regarding Market Efficiency. *Journal of Financial Economics 6*(2/3), 95–101.

Jensen, S. T., R. Pemantle, L. H. Ungar, and V. A. Satopää (2016). Supplementary Material: Partial Information Framework: Model-Based Aggregation of Estimates from Diverse Information Sources. pp. 1–9.

Johnson, M. A., R. M. Stevenson, and C. R. Letwin (2018). A woman's place is in the... startup! Crowdfunder judgments, implicit bias, and the stereotype content model. *Journal of Business Venturing* (November 2016), 1–19.

Joyce, M. A. S., A. Lasaosa, I. Stevens, and M. Tong (2011). The financial market impact of quantitative easing in the United Kingdom. *International Journal of Central Banking 7*(3), 113–161.

Juselius, K. (2006). *The Cointegrated VAR Model: methodology and applications*. Oxford: Oxford University Press.

Kane, E. J. (1981). Nested Tests of Alternative Term-Structure Theories.

Keim, D. B. and A. Madhavan (1998). The cost of institutional equity trades. *Financial Analysts Journal 54*(4), 50–69.

Keren, G. (1987). Facing uncertainty in the game of bridge: A calibration study. *Organizational Behavior and Human Decision Processes 39*(1), 98–114.

Kerr, W. R., J. Lerner, and A. Schoar (2014). The consequences of entrepreneurial finance: Evidence from angel financings. *Review of Financial Studies 27*(1), 20–55.

Kerr, W. R., R. Nanda, and M. Rhodes-Kropf (2014). Entrepreneurship as Experimentation. *Journal of Economic Perspectives 28*(3), 25–48.

Krishnamurthy, A. and A. Vissing-Jorgensen (2012). The Aggregate Demand for Treasury Debt. *Journal of Political Economy 120*(2), 233–267.

Krishnamurthy, A., A. Vissing-Jorgensen, J. Bao, O. Blanchard, G. Duffee, C. Evans, E. Faia, S. Gilchrist, R. Greenwood, M. Piazzesi, D. Romer, T. Philippon, T. Watanabe, and J. Wolfers (2011). The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy. *NBER Working Paper Series* (17555).

Kuttner, K. N. (2006). Can central banks target bond prices? *NBER Working Paper Series* (12454).

Lin, H.-w. and M. F. Mcnichols (1998). Underwriting relationships, analysts' earnings forecasts and investment recommendations. *Journal of Accounting and Economics 25*, 101–127.

Loh, R. K. and R. M. Stulz (2011). When are analyst recommendation changes influential? *Review of Financial Studies 24*(2), 593–627.

Longstaff, F. A. (2000). The term structure of very short-term rates: New evidence for the expectations hypothesis. *Journal of Financial Economics 58*(3), 397–415.

Lucas, R. E. (1973). Some International Evidence on Output-Inflation Tradeoffs. *American Economic Review 63*(3), 326–334.

Lucas, R. E. J. and N. L. Stokey (1983). Optial fiscal and monetary policy in an economy without capital. *Journal of Monetary Economics 12*, 55–93.

Lukkarinen, A., J. E. Teich, H. Wallenius, and J. Wallenius (2016). Success drivers of online equity crowdfunding campaigns. *Decision Support Systems 87*, 26–38.

Mackowiak, B. B. and M. Wiederholt (2009). Optimal Sticky Prices under Rational Inattention. *The American Economic Review 99*(3), 769–803.

Mankiw, G. N. and R. Reis (2002). Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve. *Quarterly Journal of Economics 117*(4), 1295–1328.

Manso, G. (2016). Experimentation and the Returns to Entrepreneurship. *Review of Financial Studies 29*(9), 2319–2340.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance 7*(1), 77–91.

Marois, R. and J. Ivanoff (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences 9*(6), 296–305.

Marom, D., A. Robb, and O. Sade (2016). Gender Dynamics in Crowdfunding (Kickstarter): Evidence on Entrepreneurs, Investors, Deals, and Taste-Based Discrimination.

McNichols, M. and P. C. O. Brien (1997). Self-Selection and Analyst Coverage. *Journal of Accounting Research 35*(May), 167–199.

Merton, R. C. (1987). A Simple Model of Capital Market Equilibrium with Incomplete Information. *The Journal of Finance 42*(3), 483–510.

Merton, R. C. and Z. Bodie (1995). a Conceptual Framework for Analyzing the Financial System. *Harvard Business School Press* (August), 1–42.

Michaely, R. and K. L. Womack (1999). Conflict of interest and the credibility of underwriter analyst recommendations. *Review of Financial Studies 12*(4), 653–686.

Mikhail, M. B., B. R. Walther, and R. H. Willis (1997). Do Security Analysts Improve Their Performance with Experience? *Journal of Accounting Research 35*(Studies on Experts and the Application of Expertise in Accounting, Auditing, and Tax), 131–157.

Mikhail, M. B., B. R. Walther, and R. H. Willis (2004). Do security analysts exhibit persistent differences in stock picking ability? *Journal of Financial Economics 74*(1), 67–91.

Mikhail, M. B., B. R. Walther, R. H. Willis, B. R. Walther, and R. H. Willis (1999). Does Forecast Accuracy Matter to Security Analysts ? *The Accounting Review 74*(2), 185–200.

Missale, A. (1999). *Public Debt Management*. Oxford: Oxford University Press.

Missale, A. (2012). Sovereign debt management and fiscal vulnerabilities. *BIS Paper* (65), 157–176.

Missale, A. and O. J. Blanchard (1994). The debt burden and debt maturity. *American Economic Review 84*(1), 309–319.

Modigliani, F. and R. Sutch (1966). Innovations in Interest Rate Policy. *American Economic Review 56*(1), 178–197.

Modigliani, F. and R. Sutch (1967). Debt Management and the Term Structure of Interest Rates: An Empirical Analysis of Recent Experience. *Journal of Political Economy 75*(4), 569–589.

Mohammadi, A. and K. Shafi (2016). How Wise Are Crowd? A Comparative Study of Crowd and Institutions in Peer-to-Business Online Lending Markets.

Mohammadi, A. and K. Shafi (2018). Gender differences in the contribution patterns of equity-crowdfunding investors. *Small Business Economics 50*(2), 275–287.

Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing 29*(1), 1–16.

Mollick, E. and R. Nanda (2016). Wisdom or Madness? Comparing Crowds with Expert Evaluation in Funding the Arts. *Management Science 62*(6), 1533–1553.

Moscoso del Prado Martín, F. (2011). Macroscopic thermodynamics of reaction times. *Journal of Mathematical Psychology 55*(4), 302–319.

Murphy, A. H. and R. L. Winkler (2010). Reliability of Subjective Probability Forecasts of Precipitation and Temperature. *Journal of the Royal Statistical Society, Series C (Applied Statistics) 26*(1), 41–47.

Nosbusch, Y. (2008). Interest Costs and the Optimal Maturity Structure of Government Debt. *The Economic Journal 118*(527), 477–498.

Pavlova, A. and R. Rigobon (2007). Asset Prices and Exchange Rates. *The Review of Financial Studies 20*(4), 1139–1180.

Portes, J. and P. Tucker (1995). Report of Debt Management Review. Technical report, Her Majesty's Treasury and the Bank of England, London.

Prelec, D., H. S. Seung, and J. McCoy (2017). A solution to the single-question crowd wisdom problem. *Nature 541*(7638), 532–535.

Raina, S. (2017). VC financing and the entrepreneurship gender gap. *Working Paper*.

Rajan, R. G. (2012). Presidential Address: The Corporation in Finance. *Journal of Finance 67*(4), 1173–1217.

Rajan, R. G. and L. Zingales (2001). The Firm as a Dedicated Hierarchy: a Theory Of the Origins and Growth of Firms. *The Quarterly Journal of Economics* (August), 805–851.

Rapach, D. and G. Zhou (2013). *Forecasting stock returns*, Volume 2. Elsevier B.V.

Rapach, D. E., J. K. Strauss, and G. Zhou (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies 23*(2), 821–862.

Ray, R. (2006). Prediction Markets and the Financial "Wisdom of Crowds". *The Journal of Behavioral Finance 7*(10), 58.

Robb, A. M. and S. Coleman (2010). Financing strategies of new technology-based firms: A comparison of women- and men-owned firms. *Journal of Technology Management & Innovation 5*(1), 1–18.

Sadka, R. and A. Scherbina (2007). Analyst Disagreement , Mispricing, and Liquidity. *The Journal of Finance 62*(5), 2367–2403.

Satopää, V. A., J. Baron, D. P. Foster, B. A. Mellers, P. E. Tetlock, and L. H. Ungar (2014). Combining multiple probability predictions using a simple logit model. *International Journal of Forecasting 30*(2), 344–356.

Satopää, V. A., S. T. Jensen, R. Pemantle, and L. H. Ungar (2016). Partial Information Framework: Aggregating Estimates from Diverse Information Sources.

Satopää, V. A., R. Pemantle, and L. H. Ungar (2015). Modeling Probability Forecasts via Information Diversity. *Journal of the American Statistical Association*, 1–29.

Schumpeter, J. (1934). *The theory of economic development*. Cambridge, MA: Harvard University Press.

SIFMA (2017). 2017 Fact Book. Technical report, Securities Industry and Financial Markets Association, New York.

Sims, B. C. A. (2006). Rational Inattention: Beyond the Linear-Quadratic Case. *American Economic Review 96*(2), 158–163.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics 50*(3), 665–690.

Sims, C. a. (2010). Rational Inattention and Monetary Economics. In *Handbook of Monetary Economics*, Volume 3, pp. 155–181.

Stein, J. C. (1996). Rational Capital Budgeting in an Irrational World. *Journal of Business 69*(4), 429–455.

Stickel, S. E. (1995). The Anatomy of the Performance of Buy and Sell Recommendations. *Financial Analysts Journal 51*(5), 25–39.

Strausz, R. (2017). A theory of crowdfunding: A mechanism design approach with demand uncertainty and moral hazard. *American Economic Review 107*(6), 1430–1476.

Surowiecki, J. (2005). *The Wisdom of Crowds*. New York: Anchor.

Timmermann, A. (2006). Forecast Combinations. In *Handbook of Economic Forecasting*, Volume 1, pp. 135–196.

Unger, J. M., A. Rauch, M. Frese, and N. Rosenbusch (2011). Human capital and entrepreneurial success: A meta-analytical review. *Journal of Business Venturing 26*(3), 341–358.

Van Nieuwerburgh, S. and L. Veldkamp (2009). Information immobility and the home bias puzzle. *Journal of Finance 64*(3), 1187–1215.

Varian, H. R. (1985). Divergence of Opinion in Complete Markets: A Note. *The Journal of Finance 40*(1), 309–317.

Varian, H. R. (1989). Differences of Opinions in Financial Markets. In *Financial Risk: Theory, Evidence and Implications*, pp. 3–37. Dodrecht: Springer.

Vayanos, D. and J.-L. Vila (2009). A preferred-habitat model of the term structure of interest rates. *NBER Working Paper Series* (15487).

Veldkamp, L. L. (2006). Information markets and the comovement of asset prices. *Review of Economic Studies 73*(3), 823–845.

Verardo, M. (2009). Heterogeneous beliefs and momentum profits. *Journal of Financial and Quantitative Analysis 44*(4), 795–822.

Verheul, I. and R. Thurik (2001). Start-Up Capital: "Does Gender Matter"? *Small Business Economics 16*(February 2001), 329–345.

Vismara, S. (2016a). Equity retention and social network theory in equity crowdfunding. *Small Business Economics 46*(4), 579–590.

Vismara, S. (2016b). Information Cascades Among Investors in Equity Crowdfunding. *Entrepreneurship: Theory and Practice* (March).

Vulkan, N., T. Åstebro, and M. F. Sierra (2016). Equity crowdfunding: A new phenomena. *Journal of Business Venturing Insights 5*, 37–49.

Wallace, N. (1981). Theorem for A Modigliani-Miller Operations. *American Economic Review 71*(3), 267–274.

Wallmeroth, J., P. Wirtz, and A. P. Groh (2018). Venture capital, angel financing, and crowdfunding of entrepreneurial ventures: A literature review. *Foundations and Trends in Entrepreneurship 14*(1), 1–129.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal 5*(2), 171–180.

Williams, J. T. (1977). Capital asset prices with heterogeneous beliefs. *Journal of Financial Economics 5*(2), 219–239.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics 126*(1), 25–51.

Womack, K. L. (1996). Do Brokerage Analysts' Recommendations Have Investment Value? *The Journal of Finance 51*(1), 137–167.

Woodford, M. (2009). Information-constrained state-dependent pricing. *Journal of Monetary Economics 56*(SUPPL.), S100–S124.

Woodridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.

Xu, T. (2017). Learning from the Crowd: The Feedback Value of Crowdfunding.

Zhang, X. F. (2006). Information Uncertainty and Stock Returns. *The Journal of Finance 61*(1), 105–137.