Year : 2023

THREE ESSAYS IN INFORMATION FINANCE

Nielsen Mads Bibow Busborg

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE FINANCE

THREE ESSAYS IN INFORMATION FINANCE

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès Sciences Économiques, mention « Finance »

par

Mads Bibow Busborg NIELSEN

Directeur de thèse
Prof. Norman Schürhoff

Jury

Prof. Christian Zehnder, président
Prof. Boris Nikolov, expert interne
Prof. Jérôme Dugast, expert externe
Prof. Semyon Malamud, expert externe

LAUSANNE
2023

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE FINANCE

THREE ESSAYS IN INFORMATION FINANCE

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès Sciences Économiques, mention « Finance »

par

Mads Bibow Busborg NIELSEN

Directeur de thèse
Prof. Norman Schürhoff

Jury

Prof. Christian Zehnder, président
Prof. Boris Nikolov, expert interne
Prof. Jérôme Dugast, expert externe
Prof. Semyon Malamud, expert externe

LAUSANNE
2023

UNIL | Université de Lausanne
HEC Lausanne
Le Décanat
Bâtiment Internef
CH-1015 Lausanne

I M P R I M A T U R

Sans se prononcer sur les opinions de l'auteur, la Faculté des Hautes Etudes Commerciales de l'Université de Lausanne autorise l'impression de la thèse de Monsieur Mads Bibow Busborg NIELSEN, titulaire d'un bachelor en International Business de Copenhagen Business School, et d'un master en économie et finance de Copenhagen Business School, en vue de l'obtention du grade de docteur ès Sciences économiques, mention « finance ».

La thèse est intitulée :

THREE ESSAYS IN INFORMATION FINANCE

Lausanne, le 26 juin 2023

La Doyenne

Marianne SCHMID MAST

Members of the thesis committee

Prof. Norman Schürhoff
Professor of Finance, University of Lausanne
Senior Chair, Swiss Finance Insitute
Thesis supervisor

Prof. Boris Nikolov
Professor of Finance, University of Lausanne
Junior Chair, Swiss Finance Insitute
Internal member of the thesis committee

Prof. Jérôme Dugast
Professor of Finance, Université Paris Dauphine & PSL
External member of the thesis committee

Prof. Semyon Malamud
Professor of Finance, École Polytechnique Fédérale de Lausanne
Senior Chair, Swiss Finance Institute
External member of the thesis committee

University of Lausanne
Faculty of Business and Economics

PhD in Economics,
Subject area Finance

I hereby certify that I have examined the doctoral thesis of

Mads Bibow Busborg NIELSEN

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: Norman Schürhoff__________ Date: ____May 13, 2023____

Prof. Norman SCHÜRHOFF
Thesis supervisor

University of Lausanne
Faculty of Business and Economics

PhD in Economics,
Subject area Finance

I hereby certify that I have examined the doctoral thesis of

Mads Bibow Busborg NIELSEN

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: ______________________________ Date: 13/5/2023 __________________

Prof. Boris NIKOLOV
Internal member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

PhD in Economics,
Subject area Finance

I hereby certify that I have examined the doctoral thesis of

Mads Bibow Busborg NIELSEN

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: ____________________________ Date: ___13/5/2023___________

Prof. Jérôme DUGAST
External member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

PhD in Economics,
Subject area Finance

I hereby certify that I have examined the doctoral thesis of

Mads Bibow Busborg NIELSEN

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: Semyon Malamud _______________ Date: 12.05.2023 ____________

Prof. Semyon MALAMUD
External member of the doctoral committee

Acknowledgements

I would like to thank my supervisor Norman Schürhoff for many good discussions over the years, for both challenging and encouraging me.

I would like to thank the members of my thesis committee, Jérôme Dugast, Semyon Malamud, and Boris Nikolov, for their valuable feedback on the thesis and even before that for suggestions for my job market paper (chapter 1).

I would like to thank my fellow PhD students of SFI Léman for a motivating work environment mixed with a healthy amount of fun. For my own cohort, specifically, thank you for "going in the tunnel" with me.

I would like to thank my friends back home (and elsewhere) for never forgetting to ask when I will be done studying but mostly for staying interested in how I was doing even when it got difficult to follow what I was doing.

I would like to thank my brothers for always welcoming me home as if I had only just left and my parents for so many things, but to keep it short: their support and visits over the years, virtual as well as in person.

I would like to thank my co-author (chapter 2), co-locataire, and now co-parent, my wife Suzanne for always making everything better.

Finally, I would like to thank my daughter Ada for the inspirational feat of demonstrating how much you can actually get done in 8 months if you focus on practical things such as playing and eating.

Summary

This thesis focuses on the role of information and beliefs in finance. In particular, it explores how to interpret data generated as an outcome of market equilibrium. The era of digitization has made data and analytical tools increasingly available. Consequently, it is essential to critically consider how data is generated in specific contexts to avoid misinterpreting results obtained from generalized statistical or algorithmic methods. Data sources and information technology are key concepts of the first chapter. Investors face the challenge of approximating the high-dimensional prediction function for the pay-off of a risky asset. Optimal decision-making involves choosing a biased estimator to lower variance, as well as excluding data sources to manage complexity. Historical returns appear predictable due to improved technology not available at the time. The second chapter formalizes the problem of regulating dividend payments in the banking sector, where asymmetric information leads to strategic deviations from the optimal pay-out policy. Specifically, the information concerns an imminent shock that threatens only a subset of banks, and the model explores the link between the size and scope of this shock and the opportunities and pitfalls associated with regulatory measures. The third chapter investigates sentiment analysis of social media posts as a novel source of data for understanding investor beliefs. However, it is important to recognize that communication on social media is itself an equilibrium outcome. Interactions on social media between investors with differing opinions allows for an equilibrium with information sharing. This equilibrium aligns better with empirical patterns observed in the literature than an alternative scenario where sentiment solely reflects misguided beliefs.

Résumé

Cette thèse porte sur le rôle de l'information et des croyances en finance. En particulier, elle explore comment interpréter les données produites par l'équilibre du marché. À l'ère de la numérisation, il est essentiel d'examiner de manière critique les contextes spécifiques dans lesquels celles-ci sont générées, afin d'éviter toute mauvaise interprétation des résultats obtenus à partir de méthodes statistiques ou algorithmiques généralisées. Dans le premier chapitre, les investisseurs sont confrontés au défi d'approximer la fonction de prédiction de grande dimension du rendement d'un actif risqué, en utilisant deux concepts clés : les sources de données et les technologies de l'information. La prise de décision optimale consiste à choisir un estimateur biaisé pour réduire la variance, et à exclure certaines sources de données pour gérer la complexité. Les rendements historiques semblent être prédictibles avec l'utilisation de technologies non disponibles à l'époque. Le deuxième chapitre formalise le problème de la régulation du versement des dividendes bancaires en présence d'informations asymétriques. Les informations concernent un choc imminent ne menaçant qu'une partie des banques. Le modèle explore le lien entre la taille et la portée du choc, et les opportunités et les pièges des mesures réglementaires. Le troisième chapitre examine l'analyse des sentiments des publications sur les réseaux sociaux en tant que nouvelle source de données, reconnaissant que la communication sur les réseaux sociaux est elle-même un résultat de l'équilibre. Les interactions sur les réseaux entre investisseurs ayant des opinions divergentes permettent un équilibre avec partage d'informations, ce qui correspond aux schémas empiriques observés dans la littérature.

Contents

Asset pricing with complexity

Machine learning methods for big data trade off bias for precision in prediction. To understand the implications for financial markets, I formulate a trading model with a prediction technology where investors optimally choose a biased estimator. The model identifies a novel cost of complexity that arises endogenously. This effect makes it optimal to ignore costless signals and introduces in- and out-of-sample return predictability that is not driven by priced risk or behavioral biases. Empirically, the model can explain patterns of vanishing predictability of the equity risk premium. The model calibration is consistent with a technological shift following the rise of private computers and the invention of the internet. When allowing for heterogeneity in information between agents, complexity drives a wedge between the private and social value of data and lowers price informativeness. Estimation errors generate short-term price reversals similar to liquidity demand.

1.1 Introduction

Big data is revolutionizing the finance industry (Goldstein, Spatt, and Ye, 2021) and is the key input to a new information economy (Farboodi and Veldkamp, 2021). Machine learning is recognized as a technology that unleashes the potential of Big Data, but empirical research suggests that there is more to the story (Gu, Kelly, and Xiu, 2020). Applied to return prediction, methods that rely on dimensionality reduction outperform ordinary least squares regression by successfully employing more predictors. However, they are outperformed by neural networks taking advantage of non-linear function approximation. A sophisticated approximation is useful when the true functional form of the underlying prediction problem is unknown. In this paper, I introduce a tractable formulation of such a prediction problem and derive its implications for financial markets.

Investors trade an asset with a pay-off for which some statistical properties are unknown and must be estimated from data. A multitude of data sources is available for designing the statistical model. All data sources would improve investors' predictions of the pay-off if they knew the true model. However, estimating the statistical model generates a loss of predictive performance compared to the baseline, a cost of complexity. Investors limit the cost of complexity by choosing an estimator that optimally trades off bias and variance, given the difficulty of the estimation problem and their level of sophistication, their estimation technology. Estimation technology covers algorithms and data quality/quantity as well as heuristics and experience. An unbiased estimator is generally not optimal due to its high variance. The immediate effect of improving estimation technology may be a larger bias if the trade-off with variance is attractive. Even with the optimal estimator, the cost of complexity of including certain data sources can be so high that it outweighs the benefits, and investors improve their prediction of the pay-off by excluding these data sources.

In a representative agent model, econometricians analyzing the generated market data will find predictability in returns even if they perform their analysis out of sample unless their optimal bias coincides with investors' bias. Furthermore, a lower cost of complexity due to better estimation technology can lead econometricians to consider data sources that investors originally ignored. While this is actual predictability not exploited by investors, it is not due to risk premia or irrational mispricing but the complexity of extending the statistical model to include those data sources.[1] In both cases, the predictability generated by the necessity of function approximation is fundamentally different from the in-sample predictability in returns related to parameter uncertainty (Lewellen and Shanken, 2002; Martin and Nagel, 2021). With parameter uncertainty, in-sample predictions of returns are biased, but conditioning only on information available at the time of trading (out-of-sample testing) removes the bias. With function approximation, advances in estimation technology, whether those are new tools like machine learning or better heuristics, become a source of predictability that persists out-of-sample as well.

Embedding the prediction problem in the workhorse asymmetric information model of Grossman and Stiglitz (1980), with informed and uninformed investors, the cost of complexity may be so high that no one would decide to become an informed investor even if data was available for free. The condition for informed predictions to outperform uninformed predictions is related to the condition for whether or not to ignore new data sources, but they are not identical. It is possible for informed predictions to deteriorate due to higher dimensional[2] data without informed investing being given up. I explicitly show how information asymmetry is another source of predictability that does not disappear out-of-sample. Earlier works by O'Hara (2003) and Biais, Bossaerts, and Spatt (2010) analyze asset pricing implications of models similar to the baseline model without the estimation problem, but rather than discussing return predictability per se, they focus on implications for CAPM. In the full model, the interaction between asymmetric information and function approximation produces out-of-sample 'echoes' of in-sample results even if econometricians manage to match the optimal bias and active information set of investors. This problem is particularly pronounced in markets prone to large supply shocks, high levels of noise trading. This is a concern for empirical work since trading noise might appear to be behavioral bias as these markets are populated by investors who can be considered particularly susceptible to such biases, i.e., specific stocks with large exposure to retail investors. The interaction, however, also represents an opportunity for empirical analysis since cross-sectional variation in noise trading affects the predictability arising from investors' optimal bias but not ignored data sources. The channel is price responsiveness, how strongly prices react to information and supply shocks. Generally, shocks that enter the price through investors' predictions will vary with price responsiveness, and others, e.g., priced risk, do not. In either case, additional variation is required to distinguish between such sources of predictability and those generated by the prediction problem.

The model provides additional predictions for price informativeness, price pressure and reversals, trading volume, and fund performance. Objective price informativeness is subject to a bias-variance trade-off, which is not optimized by the solution to the investors' prediction problem. Investors might ignore new data sources that improve price informativeness. Improving estimation technology closes the gap between the private and social value data. Price pressure is generated both by supply shocks and estimation noise of the prediction problem, making short-term price reversals more likely, other things being equal. In contrast to an established tradition of analyzing price pressure (Campbell, Grossman, and Wang, 1993; Hendershott and Menkveld, 2014), conditioning on price variance is more effective than trading volume for distinguishing between the two sources. Fund performance is subject

[1] For a real-world example, consider the application of natural language processing to newspaper articles and annual reports. Both sources of information have been available and presumably used by investors. However, the ability to search for patterns across thousands and thousands of publications is unique to the statistical/algorithmic approach to textual analysis.

[2] Such a deterioration can occur if subsets of the data cannot be ignored selectively.

to an unanticipated transfer between investors when comparing investors' ex-ante expectations of profits to the ex-post average of realized profits. The sign of this transfer depends on whether investors over- or underestimate the covariance between the pay-off and informed investors' prediction and appears predictable in retrospect.

In an empirical application of the model, I analyze two patterns of predictability from the literature on predicting the equity risk premium following (Welch and Goyal, 2008). The first pattern emerges *across* studies. A group of predictive variables outperforms the historical mean in the earlier part of the sample, followed by under-performance in the later part. The turn-around falls in the early 1990s. It follows the rise of the private computer in the 1980s and coincides with the early years of the internet. The second pattern emerges *between* studies and is that later papers present estimation approaches that outperform earlier papers (Campbell and Thompson, 2008; Rapach, Strauss, and Zhou, 2010; Neely, Rapach, Tu, and Zhou, 2014; Buncic and Tischhauser, 2017; Hammerschmid and Lohre, 2018). The second pattern does to a certain extent coincide with the introduction of additional data sources. However, I show that the pattern can be replicated by comparing ordinary least squares to regularized linear models (Ridge regression and LASSO). An improvement in investors' estimation technology over the run of the data series can explain the first pattern, and improved estimation technology employed in later studies can explain the second pattern. In the model, technology has a one-dimensional and, as such, ordered representation.

Calibrating the representative agent model to the data to replicate the first pattern, the change in predictability requires a large shift in technology, with the relevant parameter tripled from the earlier to the later period. The calibration is consistent with a substantial improvement in investors' estimation technology. Furthermore, the calibration demonstrates the importance of modeling optimal bias, i.e., a bias traded off for lower variance. The fit that generates the pattern of predictive out-performance followed by under-performance does not produce a decrease in the bias of investors' estimator but rather an increase.

1.1.1 Related literature

The central prediction problem in this paper is motivated by results in the empirical literature on applying machine learning to asset pricing (Gu et al., 2020; Ma, 2021). The formalization abstracts the practical approach of estimating an information structure with factor loadings and factors as distinct sub-problems formulated for linear sub-problems in Kelly, Pruitt, and Su (2017) and extended to non-linear sub-problems in Gagliardini and Ma (2019) and Gu, Kelly, and Xiu (2021) (in particular Figure 2 of that paper is a clear representation). An emerging literature on the virtue of complex models (Kelly, Malamud, and Zhou, 2022; Didisheim, Ke, Kelly, and Malamud, 2023) establish theoretically and empirically how over-parametrized models achieve good performance in return prediction. Through a rigorous application of random matrix theory the existence of a complexity wedge between a feasible and an ideal model is established in a partial equilibrium setting. The virtue of complex models suggests that increasingly complex models can outperform their antecedents and provides a motivation for studying how technological improvement of prediction methods affects a general equilibrium setting as done in the this paper.

In the analysis of return predictability as econometricians' prediction problem, this paper builds on the literature on learning in financial markets, specifically learning about parameters (Lewellen and Shanken, 2002; Pastor and Veronesi, 2009) and its extension to the high-dimensional regime of big data (Martin and Nagel, 2021). The step from learning to machine learning introduced in this paper is achieved by necessitating function approximation and, as such, introducing a bias-variance trade-off in investors' prediction problem.

The model extends classic models of information aggregation (Grossman and Stiglitz, 1980; Hellwig, 1980; Kyle, 1985) where one signal is sufficient to model. That signal might be the outcome of a complicated process of following news, analyzing company and industry fundamentals, or having private information about a firm, but the sources themselves are not important. The cost of complexity depends on the information structure, which breaks the irrelevance of the individual sources. The setting is also distinct from the multi-asset setup of Admati (1985) in which the relevance of the information structure comes from the multitude of assets rather than information sources.

In the literature on costly information acquisition (Van Nieuwerburgh and Veldkamp, 2010), a feedback effect between trading and learning decisions makes the covariance structure of pay-offs and signals relevant. The cost functions in these specifications are rather flexible and have been modeled on the spectrum from rational inattention, see Sims (2003), to the entire process of cleaning, evaluating, and processing data, see Dugast and Foucault (2020). The cost of complexity in my model is conceptually different from these exogenous cost structures in that it is an integrated part of the estimation problem and is inherently a cost measured in predictive performance. In terms of implications, the function approximation prediction problem is distinct from costly information acquisition in that it introduces an optimal bias. Furthermore, the decision to include a data source or not is not strategic, as in anticipating the feedback between trading and learning, but purely based on the statistical properties of the information. Allowing this decision to be strategic in a model about constrained but optimal prediction would correspond to allowing investors to distort their beliefs against their better knowledge, more like the optimal self-deception of Brunnermeier and Parker (2005) than rational inattention. Ultimately, the burden put on investors' ability to comprehend the full information counterfactual is smaller in my model as the trading and prediction problems are separated. In this way, the model is also distinct from the bounded rationality models of complexity of Gabaix (2014), and Molavi, Tahbaz-Salehi, and Vedolin (2021).

A pertinent question following the rise of big data has been whether it has made prices more informative. Early results in Bai, Philippon, and Savov (2016) suggest that this is the case, whereas later findings suggest that a subset of firms drives earlier results (Farboodi, Matray, Veldkamp, and Venkateswaran, 2020) and that price informativeness of other subgroups have been constant or might even have declined. This discussion has taken place in the equities market space, and the explanation proposed by Farboodi et al. (2020) focuses on the value of data about large firms versus small ones. In the presence of a cost of complexity, more data can lead to worse predictions if investors cannot separate the data into distinct sources and ignore some of them. If investors can separate the data, they might still ignore new data, and, as a result, price informativeness is unchanged. Advances in estimation technology asymptotically close the gap, but the pace of convergence at lower levels depends on the information structure. Therefore, it varies across assets even if the estimation technology is standardized.

In parallel to the discussion of price informativeness, the discussion of return predictability has, in the last ten years, seen the declaration of the 'factor zoo' (Cochrane, 2011), a replication crisis (Harvey, 2017; Hou, Xue, and Zhang, 2020), and a potential rebuttal by reference to a Bayesian baseline (Jensen, Kelly, and Pedersen, 2021). It might not only be econometricians who are affected by the high-dimensional inference problems but also investors (Martin and Nagel, 2021), and investors' solution to their prediction problem feeds into econometricians' empirical analysis. Changes in estimation technology can generate return predictability even out-of-sample, and additional sources of variation are necessary to distinguish it from risk premia and/or anomalies.

I proceed as follows. In Section 1.2, I show the inference problem in the context of a representative agent model and how it generates optimal bias and a cost of complexity. In Section 1.3, I embed the inference problem in an asymmetric information model with heterogeneity across agents. In Section 1.4, I discuss model predictions and the value of data analytically and numerically. Section 1.5 covers the empirical aplication to patterns in predictability of the equity risk premium and Section 1.6

implications for further empirical work before I conclude with Section 1.7.

1.2 Representative agent

The inference problem of the representative agent (or of the informed investors in my extension to a Grossman and Stiglitz (1980) setting in Section 1.3) is my key information friction. It presents a principled deviation from the dogma of rational expectations by only allowing partial knowledge of the true underlying information structure. The remaining part must be estimated, and I assume that this is done by choosing an estimator that optimizes the quality of the prediction of a risky pay-off, with mean squared error as the measure of quality.

To highlight the impact of the main information friction, assume investors are symmetric and have demand that is linear in the difference between their prediction and the price.[3] Additionally, I assume that agents first fix believes before making trading decisions, that is, they first build the best possible inference model and then trade on it, rather than performing a joint optimization. From the perspective of a mental model, this is coherent with an agent not changing their beliefs based on which asset they are trading. Similarly, from the perspective of a machine learning model, this is coherent with using the same predictive model for trading various assets and performing the portfolio choice step separately from the predictive step.

Trading one risky asset and one risk-less asset in elastic supply in one period and consuming their wealth in the second (see Appendix A.2.1), market clearing amongst symmetric agents require price to equal prediction. Denoting the pay-off by y and investors prediction of $\hat{y}$ that is $p = \hat{y}$. In Section 1.3, I embed the inference problem described below in the asymmetric information work-horse model of Grossman and Stiglitz (1980), introducing heterogeneity in agents.

1.2.1 Inference problem

For clarity, I impose a factor structure on the pay-off $y = \boldsymbol{\beta}^\top \boldsymbol{q}$ and allow for high dimensionality through factors $\boldsymbol{q}$ with signals $\boldsymbol{s}$ that are well-behaved. Meanwhile true factor loadings $\boldsymbol{\beta}$ are constant (and finite) but must be estimated from noisy data by estimator $\hat{\boldsymbol{\beta}}$, which, due to the noise, is a random variable. The choice of estimator is the key problem in the model.

Assumption 1 (Estimator choice)**.** *Investors minimize the mean squared error of their predictor $\hat{y}$ by trading off the bias and variance of the estimator $\hat{\boldsymbol{\beta}}$. The elements of the vectors of biases is assumed to be finite and the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ full-rank. True $\boldsymbol{\beta}$ is constant and element-wise finite.*

Intuitively, Assumption 1 says that investors care about making the best possible prediction, in contrast to putting more weight on unbiasedness for instance. Furthermore, while turning estimator choice into an optimization problem is inspired by the approach to predictions of machine learning the trade-off could be considered a prior selection mechanism. For tractability, an additional assumption on the independence of the noise in the estimator is important.

Assumption 2 (Independent noise)**.** *The noise in the data of estimator $\hat{\boldsymbol{\beta}}$ is independent of factors $\boldsymbol{q}$ and signals $\boldsymbol{s}$.*

The noise in the estimation data makes $\hat{\boldsymbol{\beta}}$ a random variable with unconditional expectation $E\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\mu}_\beta$

[3]Risk neutrality or the demand function described in Appendix A.2.3 or Appendix A.2.4 are all valid choices.

and variance-covariance matrix $Var\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\sigma}_\beta^\top \boldsymbol{R}_\beta \boldsymbol{\sigma}_\beta$ where $\boldsymbol{R}_\beta$ is the correlation matrix of $\hat{\boldsymbol{\beta}}$ and I drop the hat on subscripts to avoid clutter.

Factors and signals are well-behaved, assuming joint normality of non-constant, non-redundant variables. I use the notation $\boldsymbol{\Gamma} := \boldsymbol{R}_{qs}\boldsymbol{R}_s^{-1}\boldsymbol{R}_{qs}^\top$ where matrices $\boldsymbol{R}_{qs}, \boldsymbol{R}_s, \boldsymbol{R}_{sq}$ are correlation matrices. Additionally, I denote a diagonal matrix by $\boldsymbol{D}$.

Assumption 3 (Gaussian factor expectations)**.** *Factors and signals follows a multi-variate normal distribution, such that the conditional expectation of factors given signals is*

$$\boldsymbol{\zeta} := E[\boldsymbol{q}|\boldsymbol{s}] = \boldsymbol{\mu}_q + \boldsymbol{\Sigma}_{qs}\boldsymbol{\Sigma}_s^{-1}(\boldsymbol{s} - \boldsymbol{\mu}_s), \text{ where } \boldsymbol{\zeta} \sim \mathcal{N}\left(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_\zeta\right),$$

with element-wise finite expectations $\boldsymbol{\mu}_q$ and full-rank variance-covariance matrix

$$\boldsymbol{\Sigma}_\zeta = \boldsymbol{\Sigma}_{qs}\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Sigma}_{qs}^\top = \boldsymbol{D}_{\sigma_q}\boldsymbol{R}_{qs}\boldsymbol{D}_{\sigma_s}\boldsymbol{D}_{\sigma_s}^{-1}\boldsymbol{R}_s^{-1}\boldsymbol{D}_{\sigma_s}^{-1}\boldsymbol{D}_{\sigma_s}\boldsymbol{R}_{qs}\boldsymbol{D}_{\sigma_q} = \boldsymbol{D}_{\sigma_q}\boldsymbol{\Gamma}\boldsymbol{D}_{\sigma_q}.$$

To derive the bias-variance trade-off of Assumption 1, I first decompose the mean squared error of predictor $\hat{y}$. Some notation is helpful here. The second moment matrix of conditional expectation $\boldsymbol{\zeta}$ and unconditional bias of estimator $\hat{\boldsymbol{\beta}}$ are respectively

$$\boldsymbol{\Omega}_\zeta := E\left[\boldsymbol{\zeta}\boldsymbol{\zeta}^\top\right] = \boldsymbol{\mu}_q\boldsymbol{\mu}_q^\top + \boldsymbol{\Sigma}_\zeta, \quad \text{and} \quad \boldsymbol{\varepsilon}_\beta := E\left[\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta} - \boldsymbol{\mu}_\beta.$$

Lemma 1 (Mean squared error decomposition)**.** *The mean squared error of predictor $\hat{y}$ can be decomposed into three terms*

$$\begin{aligned} E\left[(y-\hat{y})^2\right] &= \left(E[y] - E[\hat{y}]\right)^2 + Var[y] + Var\left[\hat{y}\right] - 2Cov\left[y, \hat{y}\right] \\ &= \underbrace{\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\Omega}_\zeta \boldsymbol{\varepsilon}_\beta}_{\textit{bias-squared term}} + \underbrace{\boldsymbol{\sigma}_\beta^\top(\boldsymbol{R}_\beta \odot \boldsymbol{\Omega}_\zeta)\boldsymbol{\sigma}_\beta}_{\textit{variance term}} + \underbrace{Var[y|\boldsymbol{s},\boldsymbol{\beta}]}_{\textit{irreducible error}} \end{aligned}$$

Proof. Given Assumption 2, the variance of the predictor is the variance of two independent random vectors, which can be written using the Hadamard product $\odot$ (see Appendix A.1.1)

$$Var[\hat{y}] = \boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta - \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\beta} + 2\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta(\boldsymbol{\beta} - \boldsymbol{\varepsilon}_\beta) + \boldsymbol{\sigma}_\beta^\top(\boldsymbol{R}_\beta \odot \boldsymbol{\Omega}_\zeta)\boldsymbol{\sigma}_\beta, \tag{1.1}$$

and $\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta(\boldsymbol{\beta} - \boldsymbol{\varepsilon}_\beta) = Cov\left[y, \hat{y}\right]$ which cancels out with the negative covariance terms of the mean squared error. The squared bias of predictor $\hat{y}$ can be written as a quadratic form of the bias of the estimator and factor means $\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\mu}_q \boldsymbol{\mu}_q^\top \boldsymbol{\varepsilon}_\beta$ and collected in $\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\Omega}_\zeta \boldsymbol{\varepsilon}_\beta$ with $\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta$ of $Var[\hat{y}]$ in equation (1.1). The remaining two terms can be collected in the conditional variance under the true model

$$\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_q \boldsymbol{\beta} - \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\beta} = Var[y|\boldsymbol{\beta}] - Var[\hat{y}|\boldsymbol{\beta}] = Var[y|\boldsymbol{s},\boldsymbol{\beta}]$$

Existence of the moments follows from Assumption 1 and Assumption 3. □

The labeling of the three terms in Lemma 1 matches the bias-variance decomposition as it is usually defined (see Hastie, Tibshirani, and Friedman (2009) chapter 7.3), and the variance under the true model is the irreducible noise from the perspective of choosing the estimator $\hat{\boldsymbol{\beta}}$. Therefore, the the bias-variance trade-off for a given information set only applies to the first two terms. This is in contrast to considerations on expanding the information set as I will return to in Section 1.2.4. Notice that the decomposition of Lemma 1 does not depend on the asumptions on the specific functional form of the bias and variance of the estimator except for the finiteness and full-rank condition of Assumption 1.

1.2.2 Bias-variance trade-off

More structure is necessary to formulate a meaningful minimization of the first two terms in Lemma 1. One piece is the constraint that it is a trade-off between bias and variance. Imposing the constraint directly on the bias and variance terms neglects the factor structure of the problem. Instead, I specify element-wise symmetric functions for bias $\boldsymbol{\varepsilon}_\beta$ and volatility $\boldsymbol{\sigma}_\beta$ and extend Assumption 2 to assume no correlation *between* factor loadings as well.[4] Bias and volatility functions are linked through a vector of controls $\boldsymbol{c}$. The controls are an abstraction that captures the choices involved in choosing an estimator. They incorporate both the high-level decision of which estimator (e.g. ordinary least square, LASSO, neural net etc.) but also the details such as how to clean the data and tune any hyper-parameters the estimator may have.

Assumption 4. *Bias and volatility are element-wise symmetric functions $\varepsilon_{\beta i} = f_\varepsilon(c_i)$ and $\sigma_{\beta i} = f_\sigma(c_i)$, such that $\partial\varepsilon_{\beta j}/\partial c_i = \partial\sigma_{\beta j}/\partial c_i = 0\ \forall\, j \neq i$ and factor loadings are uncorrelated, i.e. $\boldsymbol{R}_\beta = \boldsymbol{I}$.*

In this way, the trade-off constraint can be imposed cleanly as a set of pairwise restrictions $f'_\varepsilon(c_i) f'_\sigma(c_i) < 0$, and the interactions between factors loadings arise from the minimization rather than being imposed on it. The structure of these interactions is entirely determined by the factor structured encoded in $\boldsymbol{\Omega}_\zeta$ as the weighing matrix of the variance term simplifies to $\boldsymbol{I} \odot \boldsymbol{\Omega}_\zeta = \boldsymbol{D}_{\Omega_\zeta}$. To motivate Assumption 4, recall that the randomness in the estimator of the factor loadings follows from the noise in the data and (potentially) the estimation method. Without a specific type of interaction between the two in mind, it seems prudent to limit the impact on the solution of structure imposed on the noise of the problem. However, maintaining the assumption of a pairwise trade-off, interactions can be introduced by specifying a correlation matrix $\boldsymbol{R}_\beta$ and simply replacing the diagonal matrix $\boldsymbol{D}_\zeta$ by $\boldsymbol{R}_\beta \odot \boldsymbol{\Omega}_\zeta$ in the following derivations.

Applying Assumption 4 to the first two terms of Lemma 1 and requiring that volatility is non-negative, the bias-variance trade-off as a constrained minimization is

$$\min_{\boldsymbol{c}} \Theta := \boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\Omega}_\zeta \boldsymbol{\varepsilon}_\beta + \boldsymbol{\sigma}_\beta^\top \boldsymbol{D}_{\Omega_\zeta} \boldsymbol{\sigma}_\beta \quad \text{subject to} \quad f'_\varepsilon(c_i) f'_\sigma(c_i) < 0, f_\sigma(c_i) \geq 0\ \forall\, c_i \in \boldsymbol{c}. \tag{1.2}$$

The strict inequality of the constraint on the product of first derivatives requires that both are non-zero. To derive a closed form solution, I impose additional restrictions on the functional form of bias and volatility.

Assumption 5. *Bias is a linear and volatility an affine function of control given by*

$$f_\varepsilon(c_i) = k_\varepsilon c_i, \quad f_\sigma(c_i) = k_\sigma c_i + k_{\sigma 0}, \quad \textit{subject to} \quad k_\varepsilon k_\sigma > -\infty, k_{\sigma 0} \in (0, \infty).$$

Assumption 5 is more technical than Assumption 4, and its motivation is to limit the problem defined in (1.2) to a class of minimizations with unique minima, and make it easier to parse equilibrium outcomes[5]. The assumption of affinity of the volatility function is not necessary for uniqueness and has the drawback that while a solution to an unrestricted solution to optimization (1.2) exists (see Proposition 1) it might not be feasible as it could violate the non-negativity constraint. However, when that solution is feasible, it is available in closed form and feasibility only depends on the factor structure, not the parameters of the bias and volatility functions. Bias can be extended to an affine form without fundamentally altering the form of optimal controls and minimized bias-variance

[4] Estimating factors and factor loadings are two sub-problems of the prediction problem. The assumptions made about the factors and signals of the risky pay-off in Assumption 3 are separate from the specification of the factor loadings.

[5] An intuitive direction for future work is to apply the analysis to a specific estimator. The ridge estimator with its closed form solution is a natural candidate due to its tractability.

(see Appendix A.1.2) but it comes at the cost of more involved expressions. In contrast, restricting both functions to a linear form leaves controls of all zeros as the only solution to the minimization. Restricting bias rather than variance makes it possible to study the impact of forcing the estimator to be unbiased by setting controls to zero $\boldsymbol{c} = 0$ for an estimator with variance $\boldsymbol{\sigma}_\beta = k_{\sigma 0}\mathbf{1}$.

1.2.3 Optimal bias

I denote the optimal controls by a star $\boldsymbol{c}^* = \arg\max_{\boldsymbol{c}} \Theta$ and denote the minimized objective as $\chi := \Theta|_{\boldsymbol{c}=\boldsymbol{c}^*}$. In Section 1.2.4, I show formally that χ is the cost of complexity and it increases in the number of signals n_s. Additionally, to describe the solution to optimization (1.2) it is convenient to define the ratio of the slope parameters of the bias and volatility functions as $k_c := k_\sigma / k_\varepsilon$. I will refer to the square of the ratio of slope parameters k_c^2 as the estimation technology parameter because the cost of complexity χ is everywhere decreasing in it (see Equation 1.3). In contrast, I will interpret the constant of the volatility function $k_{\sigma 0}$ as the difficulty of the estimation because it determines the cost of complexity under the inefficient but unbiased estimator $\Theta|_{\boldsymbol{c}=0}$ that is only optimal for $k_c^2 = 0$. Increasing k_c^2 changes the trade-off whereas increasing $k_{\sigma 0}$ simply scales up bias, volatility, and the cost of complexity.

Proposition 1 (Bias-variance trade-off solution)**.** *Under Assumption 4 and Assumption 5, the unconstrained solution to optimization (1.2) and cost of complexity are*

$$\boldsymbol{c}^* = -k_\sigma^{-1} k_{\sigma 0} \left\{ \boldsymbol{I} - \boldsymbol{D}_{\Omega_\zeta}^{-1} \boldsymbol{X}^{-1} \right\} \mathbf{1}, \quad \chi = \Theta|_{\boldsymbol{c}=\boldsymbol{c}^*} = k_{\sigma 0}^2 \mathbf{1}^\top \boldsymbol{X}^{-1} \mathbf{1}, \ \textit{where}\ \boldsymbol{X} = k_c^2 \boldsymbol{\Omega}_\zeta^{-1} + \boldsymbol{D}_{\Omega_\zeta}^{-1}$$

This solution to the unconstrained minimization problem always exists and it is unique.

Proof. For the complete algebraic manipulation see Appendix A.1.2. Existence follows from the positive definiteness of $\boldsymbol{X}$, and uniqueness from the positive definiteness of the Hessian matrix of the objective $k_\varepsilon^2 \boldsymbol{\Omega}_\zeta + k_\sigma^2 \boldsymbol{D}_{\Omega_\zeta}$. Since both are sums of positive definite matrices they are also positive definite. □

Corollary 1.1 (Optimal bias and volatility)**.** *Bias and volatility only depend on slope parameters k_ε and k_σ through their ratio k_c*

$$\boldsymbol{\varepsilon}_\beta|_{\boldsymbol{c}=\boldsymbol{c}^*} = -k_c^{-1} k_{\sigma 0} \left\{ \boldsymbol{I} - \boldsymbol{D}_\Omega^{-1} \boldsymbol{X}^{-1} \right\} \mathbf{1} \geq \mathbf{0}, \quad \boldsymbol{\sigma}_\beta|_{\boldsymbol{c}=\boldsymbol{c}^*} = k_{\sigma 0} \boldsymbol{D}_\Omega^{-1} \boldsymbol{X}^{-1} \mathbf{1}.$$

Proof. For the proof of the inequality see Appendix A.1.2. □

Proposition 1 leaves open the question of the feasibility of the solution described. In Assumption 6, I present the condition for feasibility of the solution in Proposition 1. Technological developments captured by changes in the technology parameters k_c and $k_{\sigma 0}$ do not affect the status of the feasibility condition because it does not depend on them.

Assumption 6 (Bias-variance trade-off feasibility)**.** *I assume that the following element-wise vector inequality holds*

$$\boldsymbol{\sigma}_\beta|_{\boldsymbol{c}=\boldsymbol{c}^*} > \mathbf{0} \iff \boldsymbol{\Omega}_\zeta \mathbf{1} > \mathbf{0}.$$

See Appendix A.1.2 for the algebraic manipulation of $\boldsymbol{\sigma}_\beta|_{\boldsymbol{c}=\boldsymbol{c}^*}$ that demonstrates equivalence of the two inequalities. Informally, since the on-diagonal entries of $\boldsymbol{\Omega}_\zeta$ are positive, Assumption 6 restricts the set of factor structures with a feasible solution to structures with “not too” negative average cross-second moments of conditional expectations of factors given signals.

For the interpretation of the square of the ratio of slope parameters k_c^2 as a measure of estimation technological development, notice that the cost of complexity is indeed decreasing in it

$$\frac{\partial \chi}{\partial k_c^2} = -k_{\sigma 0}^2 \mathbf{1}^\top \boldsymbol{X}^{-1} \boldsymbol{\Omega}_\zeta^{-1} \boldsymbol{X}^{-1} \mathbf{1} \leq 0, \tag{1.3}$$

but in a non-linear fashion, which means that it changes the trade-off between bias and variance rather than simply scaling them up or down. This improvement could represent new techniques or better input data for the estimation of factor loadings $\hat{\boldsymbol{\beta}}$ since both are abstracted into the properties of the estimator in Assumption 1 and Assumption 2.

1.2.4 Cost of complexity

To demonstrate that the minimized objective χ is the cost of complexity and the variance under true model $Var[y|\boldsymbol{\beta}, \boldsymbol{s}_I]$ is the balancing cost of simplicity, I first provide a recursive formulation of the second moment matrix of conditional expectations $\boldsymbol{\Omega}_\zeta$ based on block matrix inversion and multiplication. The recursion is over the number of signals n_s, and concerns the decision of including the n_śth signal in the vector of signals $\boldsymbol{s}$. The main step is to operate on the matrix of correlations $\boldsymbol{\Gamma}_{n_s} = \boldsymbol{R}_{qs,n_s} \boldsymbol{R}_{s,n_s}^{-1} \boldsymbol{R}_{qs,n_s}^\top$, and in particular the inverse signal correlation matrix $\boldsymbol{R}_{s,n_s}^{-1}$. Explicit derivations are in the Appendix A.1.3, but it is necessary to introduce some notation here. Denote the correlation between the new signal n_s and the extant signals $\boldsymbol{s}_{n_s-1}$ by $\boldsymbol{\rho}_{s,n_s} := Corr[\boldsymbol{s}_{n_s-1}, s_{n_s}]$, the correlation between the new signal and factors $\boldsymbol{q}$ by $\boldsymbol{\rho}_{q_i s,n_s}^\top = (\boldsymbol{\rho}_{q_i s,n_s-1} \;\; \rho_{q_i s_{n_s}})$, and define the correlation correction $\rho_{s,n_s|n_s-1} := 1 - \boldsymbol{\rho}_{s,n_s}^\top \boldsymbol{R}_{s,n_s}^{-1} \boldsymbol{\rho}_{s,n_s}$. With this notation, it is possible to define the vector $\boldsymbol{\phi}_{n_s}$ with elements $\phi_{i,n_s} := \boldsymbol{\rho}_{q_i s,n-1}^\top \boldsymbol{R}_{s,n_s-1}^{-1} \boldsymbol{\rho}_{s,n} - \rho_{q_i s_{n_s}}$. The recursive formulation of the matrix of correlations is then $\boldsymbol{\Gamma}_{n_s} = \boldsymbol{\Gamma}_{n_s-1} + \rho_{s,n_s|n_s-1}^{-1} \boldsymbol{\phi}_{n_s} \boldsymbol{\phi}_{n_s}^\top$ and it follows immediately that the difference $\boldsymbol{\Gamma}_{n_s} - \boldsymbol{\Gamma}_{n_s-1} = \rho_{s,n_s|n_s-1}^{-1} \boldsymbol{\phi}_{n_s} \boldsymbol{\phi}_{n_s}^\top$ is positive semi-definite. Finally, this means that the second moment matrix of conditional expectations has the recursive formulation

$$\begin{aligned}
\boldsymbol{\Omega}_{\zeta,n_s} &= \boldsymbol{\mu}_q \boldsymbol{\mu}_q^\top + \boldsymbol{D}_{\sigma_q} \boldsymbol{\Gamma}_{n_s-1} \boldsymbol{D}_{\sigma_q} + \rho_{s,n_s|n_s-1}^{-1} \boldsymbol{D}_{\sigma_q} \boldsymbol{\phi}_{n_s} \boldsymbol{\phi}_{n_s}^\top \boldsymbol{D}_{\sigma_q} \\
&= \boldsymbol{\Omega}_{\zeta,n_s-1} + \rho_{s,n_s|n_s-1}^{-1} \boldsymbol{D}_{\sigma_q} \boldsymbol{\phi}_{n_s} \boldsymbol{\phi}_{n_s}^\top \boldsymbol{D}_{\sigma_q},
\end{aligned} \tag{1.4}$$

where the last term also can be written as an outer product

$$\boldsymbol{D}_{\sigma_q} \boldsymbol{\phi}_{n_s} \boldsymbol{\phi}_{n_s}^\top \boldsymbol{D}_{\sigma_q} = \left(\boldsymbol{\sigma}_q \odot \boldsymbol{\phi}_{n_s}\right) \left(\boldsymbol{\sigma}_q \odot \boldsymbol{\phi}_{n_s}\right)^\top.$$

Proposition 2 (Cost of complexity vs simplicity)**.** *Signs of the increments in the minimized bias-variance trade-off objective and the conditional variance under the true model based on including a signal are*

$$\begin{aligned}
\chi_{n_s} - \chi_{n_s-1} &= k_{\sigma 0} \mathbf{1}^\top \left\{ \boldsymbol{X}_{n_s}^{-1} - \boldsymbol{X}_{n_s-1}^{-1} \right\} \mathbf{1} \geq 0, \\
Var[y|\boldsymbol{\beta}, \boldsymbol{s}_{n_s}] - Var[y|\boldsymbol{\beta}, \boldsymbol{s}_{I,n_s-1}] &= \boldsymbol{\beta}^\top \boldsymbol{D}_{\sigma_q} (\boldsymbol{\Gamma}_{n_s-1} - \boldsymbol{\Gamma}_{n_s}) \boldsymbol{D}_{\sigma_q} \boldsymbol{\beta} \leq 0.
\end{aligned}$$

Proof. The second inequality follows from the observation made in the main-text that $\boldsymbol{\Gamma}_{n_s} - \boldsymbol{\Gamma}_{n_s-1}$ is positive semi-definite. By properties of symmetric positive definite matrices the difference $\boldsymbol{X}_{n_s}^{-1} - \boldsymbol{X}_{n_s-1}^{-1}$ is positive semi-definite if $\boldsymbol{X}_{n_s-1} - \boldsymbol{X}_{n_s}$ is positive semi-definite. This is shown to be the case in Appendix A.1.4. □

An illustration of Proposition 2 can be found in Figure 1.2, which covers both the case where the cost of complexity dominates and the case where the benefit of reducing variance under the true model is greater. In Appendix A.1.5 and Appendix A.1.6, I show how to extend this analysis to an arbitrary group of additional signals. By induction, the results in Proposition 2 must hold for groups of signals, but I demonstrate that a convenient form similar to 1.4 exist for a group of signals and, indeed, confirm Proposition 2 for this more general case.

Intuitively what Proposition 2 shows is that adding another signal always (weakly) increases the cost of complexity and lowers the conditional variance under the true model. This is key for interpreting χ as the cost of complexity because it increases when the model is expanded along new dimensions[6]. It is also worth noticing, that the cost of complexity only increases in signals that are fundamentally informative about the factors or the already included signals. This follows from the recursive formulation of the second moment matrix in equation (1.4) and the definition of the vector $\boldsymbol{\phi}_{1,n_s}$. For a signal uncorrelated with factors $\boldsymbol{q}$ and already included signals $\boldsymbol{s}_{I,n_s-1}$, the vector vector $\boldsymbol{\psi}_{n_s}$ is zero and there is no difference between second moment matrix $\boldsymbol{\Omega}_{\zeta,n_s}$ and $\boldsymbol{\Omega}_{\zeta,n_s-1}$, which propagates to $\boldsymbol{X}_{n_s}$ and $\boldsymbol{X}_{n_s-1}$. This aligns well with the cost of simplicity being the conditional variance under the true model, which clearly is not decreased by conditioning on an irrelevant signal.

1.3 Heterogeneous agents

In this section, I formulate an extension of the model on the basis of the work-horse asymmetric information model of Grossman and Stiglitz (1980). I make two adjustments to uninformed inference and demand that do not change the classic model but make a difference when informed investors solve the inference problem described in the previous Section 1.2.1.

Two homogeneous groups of investors, informed and uninformed denoted by $i \in \{I, U\}$, trade a risky asset with independent mean-zero stochastic supply z optimizing demand δ_i over the utility of ultimate profit (equivalent to final wealth, See Appendix A.2.1). A risk-free asset, which acts as a numeraire with a price and pay-off of one, is available in perfectly elastic supply. Investors are price-takers trading in demand-schedules akin to posting limit orders rather than market orders, see Kyle (1989).

1.3.1 Uninformed inference

Investors have common priors, which would lead all investors to make the same predictions if endowed with the same information. Investors of type i have information set $\mathscr{F}_i$ and a linear demand function of the form $\delta_i = \psi_i(\hat{y}_i - p)$ where $\hat{y}_i = E[y|\mathscr{F}_i]$ and, with uncertainty aversion α_i, $\psi_i = \{\alpha_i E[(y-\hat{y}_i)^2]\}^{-1}$. In Appendix A.2.3 and Appendix A.2.4, I present two foundations for this demand function, respectively a robust profit maximization objective, and CARA-utility with ambiguity aversion. For simplicity, assume that investors know the unconditional mean squared error.[7] In equilibrium, the market clears and the uninformed can extract the signal $s_U := p - \psi_I^{-1}\delta_U = \hat{y}_I - \psi_I^{-1} z$, where ψ_I is the scaling factor of informed demand. I assume that the uninformed investors' prediction of informed investors' prediction is the best linear approximation which I signify by adding a tilde to the expectation, i.e. $\tilde{E}[\cdot]$.

[6] An aspect that is not treated in the current version of the model, is what this means for the production of information. Presumably, investors would only invest in useful signals, however, the uncertainty ex ante could lead to date being collected and later ignored. Further work is necessary to exactly understand the impact on the production of data, but one prediction of the model would be how better technology can also improve the value of certain data sources since they can better be accommodated.

[7] This assumption can be understood as investors having access to methods such as simulation and cross-validation, to approximate the unconditional mean squared error, and that these methods are accurate enough to abstract away this approximation step and model the approximation as the true unconditional mean squared error. As such, what is abstracted away is the noise in the approximation step.

It is given by the projection

$$\hat{y}_U = \tilde{E}[\hat{y}_I|s_U] = (1-\lambda_U)E[\hat{y}_I] + \lambda_U s_U \text{ where } \lambda_U = \frac{Var[\hat{y}_I]}{Var[\hat{y}_I] + \psi_I^{-2}\sigma_z^2} < 1,\ \sigma_z^2 = Var[z]$$

The best linear approximation minimizes the mean squared error $E[(\hat{y}_I - \hat{y}_U)^2]$, which is consistent with the way the informed chose their predictor (see Section 1.2.1). In the baseline model where factor loadings are known (equivalent to Grossman and Stiglitz (1980)) this prediction corresponds to the expectation of the pay-off given s_U (see Appendix A.2.2). The conditional expectation is linear and, therefore, the best linear approximation is *the* best approximation. The formulation here allows for the informed investors noisy estimation of factor loadings that does not depend on a full specification of the distribution of noise. This is true in the baseline model with known factor loadings as well.

1.3.2 Uninformed mean squared error

The mean squared error of the uninformed investors is a convex combination of the mean squared error of the informed investors and the sum of the unconditional variance of the pay-off and the square of biases scaled by factor means

$$E[(y-\hat{y}_U)^2] = (1-\lambda_U)\left\{Var[y] + (\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\mu}_q)^2\right\} + \lambda_U E[(y-\hat{y}_I)^2].$$

Other things equal, a high bias compared to the total cost of complexity χ, which is an element of the informed mean squared error, tends to make the mean squared error of the uninformed higher than that of the informed. Meanwhile, the variance of the pay-off contribute to both both terms but in the informed mean squared error it is through the conditional variance. The conditional variance decomposes into $Var[y|\boldsymbol{\beta}, \boldsymbol{s}_I] = Var[y] - Var[E[y|\boldsymbol{\beta}, \boldsymbol{s}_I]]$. Compared to the cost of complexity, signals more informative under the true model, captured by the variance of the expectation with known factors, also tend to make the predictions of the informed investors better than the uninformed. In a baseline model with known factor loadings there is no cost of complexity and the informed mean squared error is always (weakly) lower than the uninformed. I formalize this in Proposition 3.

Proposition 3 (Informed predictions do not always outperform)**.** *The necessary and sufficient condition for out-performance of uninformed predictions by informed predictions is that sum of the bias squared and the variance of the conditional expectation of the pay-off under the objective measure is greater than the cost of complexity*

$$\begin{aligned} E[(y-\hat{y}_U)^2] > E[(y-\hat{y}_I)^2] &\iff \left\{Var[y] + (\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\mu}_q)^2\right\} > E[(y-\hat{y}_I)^2] \\ &\iff Var[E[y|\boldsymbol{\beta}, \boldsymbol{s}_I]] + (\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\mu}_q)^2 > \chi. \end{aligned}$$

Proof. See Appendix A.2.5. □

For investors who make their best effort to produce the best prediction possible, the condition can be considered a requirement for anyone to choose to be informed. Notice that the variance of the conditional expectation is the reduction in variance achieved by using a vector of signals under the true model, formally $Var[y|\boldsymbol{\beta}, \boldsymbol{s}_I] = Var[y] - Var[E[y|\boldsymbol{\beta}, \boldsymbol{s}_I]]$. The condition echoes the cost and benefits analysis of adding signals in Section 1.2.4 in that, modulo the squared bias term, the quality of the signals under the true model must be greater than the cost of complexity.

1.3.3 Price

I close the model extension by deriving the equilibrium price. With linear demand, and an uninformed prediction that is a convex combination, price is a convex combination as well

$$p = (1-\lambda_p)E[\hat{y}_I] + \lambda_p s_U \quad s.t. \quad \lambda_p = \frac{\psi_I + \psi_U \lambda_U}{\psi_I + \psi_U}.$$

Irrespective of the details of the predictions and demand scaling factors ψ_i, the weight on the signal is greater than in the uninformed prediction since

$$\lambda_p > \lambda_U \iff \psi_I + \psi_U \lambda_U > (\psi_I + \psi_U)\lambda_U \iff 1 > \lambda_U.$$

The functional form of price is the same as the expectation of the uninformed investors. If price is viewed as the market's prediction of the risky pay-off, the uninformed are less responsive to the information and supply shocks of s_U than the market since the market also reflect the positioning of informed investors and noise traders. This follows from the asymmetric information and form of demand rather than the inference problem.

1.4 Predictions

In this section, I highlight a number of predictions where the model (with an without the heterogeneous agents extension) deviate from the baseline model with known factor loadings. To support the analytical analysis, I perform a numerical analysis. For tractability, the numerical analysis is carried out in the minimal setting of two factors, two established signals, and two new signals. For reference, parameters can be found in Table 1.1, the central matrices in Table 1.2, key moments in Table 1.3, and market structure variables in Table 1.4. First, I impose more structure on the addition of new or more data and discuss what it means for the value of data.

1.4.1 Value of data

The specification of the inference problem in Section 1.2.1 introduces two types of data. The signals about factors and the data that estimates of the factor loadings are based on. Throughout the paper, I have lumped in the second type with the estimation technology, because they are both abstracted into the properties of the estimator, and are mathematically summarized by the parameters k_c and $k_{\sigma 0}$. Conceptually, improving estimation technology could mean getting better input data, and better data could mean bigger data. From this perspective, the value of data is a matter of assumption, if it increases k_c^2 it lowers the cost of complexity, if it increases $k_{\sigma 0}$ it raises the cost of complexity. The type of data that affects inference under the true model is the signals. By affecting both the cost of complexity and cost of simplicity the model puts more structure on this data, and its effects on equilibrium outcomes can be compared directly to the baseline model with perfect inference. I extend the study of the inclusion of a group of discrete signals as in Section 1.2.4 (or groups of signals in Appendix A.1.5), to the continuous case by analyzing a degree formulation of the problem where the second-moment matrix is given by

$$\boldsymbol{\Omega}_\zeta = \boldsymbol{\Omega}_{\zeta 0} + k_S \boldsymbol{D}_{\sigma_q} \boldsymbol{\Phi}_{n_s} \boldsymbol{R}^{-1}_{s,n_s+|n_s-} \boldsymbol{\Phi}^\top_{n_s} \boldsymbol{D}_{\sigma_q} := \boldsymbol{\Omega}_{\zeta 0} + k_S \boldsymbol{S}, \tag{1.5}$$

and the formulation of the additional signals matrix is taken from Appendix A.1.5. If the additional signals are independent of each other and the extant signals and k_S is an integer, it can be interpreted directly as a count of the number of identical signals. More generally, k_S scales the signal group up or down without changing in-between correlations or correlations with the base signals in $\boldsymbol{\Omega}_{\zeta 0}$, providing a way to have more or less of the information it represents. A limitation of this approach is that there is no built-in restriction on k_S that guarantees that the overall correlation structure is feasible. In applications, k_S must be kept at levels that do not generate impossibilities like a negative conditional variance under the true model. However, with this restriction in place, comparative statics with the parameters k_c^2, $k_{\sigma 0}$, and k_S is a useful exercise that captures different aspects of the model.

It is possible to say a bit more about the specification in (1.5) because the conditional variance under the true model is linear in k_S and the cost of complexity can be shown to be a rational function with an oblique asymptote (i.e. the asymptote is linear in k_S), see Appendix A.4.1. As the cost of complexity is increasing in k_S and the cost of simplicity decreasing by Proposition 2, the overall mean squared error is eventually linear in k_S, and the tendency comes down to comparing

$$\frac{\partial Var[y|\boldsymbol{\beta}, \boldsymbol{s}_I]}{\partial k_S} = -\boldsymbol{\beta}^\top \boldsymbol{S}\boldsymbol{\beta}, \text{ to } \lim_{k_S \to \infty} \frac{\partial \chi(k_S)}{\partial k_S} = k_{\sigma 0}^2 \mathbf{1}^\top (k_c^2 \boldsymbol{S}^{-1} + \boldsymbol{D}_S^{-1})^{-1} \mathbf{1}.$$

Because $\chi(k_S)$ is everywhere increasing, it approaches the asymptote from below and it follows that $\frac{\partial \chi(k_S)}{\partial k_S}$ is decreasing. Over a given range of k_S, it is, therefore, possible for the mean squared error $E[(y - \hat{y}_I)^2]$ to take one of three shapes. If the derivative of the asymptote dominates the cost of simplicity, it is monotonically increasing. When the reverse is true, the mean squared error can be hump-shaped as the $\frac{\partial \chi(k_S)}{\partial k_S}$ is decreasing or monotonically decreasing. One shape the specification cannot generate is a U-shape where more data is initially beneficial and then eventually becomes a liability.[8] Treating the full second-moment matrix in (1.5) as one, it is possible that an initial reduction in mean squared error $\boldsymbol{\Omega}_{\zeta 0}$ is gradually undone by a higher k_S until it is eventually better to ignore the full vector of combined signals if it is not possible to separate it. This way the specification can simulate what is in Dugast and Foucault (2020) described as the needle in a haystack problem of big data that is that it becomes harder to find the good signals when there are many to search through.

1.4.2 Bias and volatility

In addition to covering the optimal bias and volatility across the technology parameters k_c^2 and $k_{\sigma 0}$ as well as the new data parameter k_S, Figure 1.1 includes a graphical representation of the restriction in Proposition 3 that informed predictions outperform uninformed predictions. For baseline levels $k_c^2 = 1$ and $k_S = 0.5$ the cut-off is at a bit above $k_{\sigma 0} = 0.6$, which is chosen as a harder estimation baseline compared to $k_{\sigma 0} = 0.3$. In Figure 1.4 the significance of these two levels of difficulty is demonstrated. Under the easy estimation, a stronger new data source input (a higher k_S) results in an overall lower mean squared error of the informed investors' predictor. In contrast, the cost of complexity dominates for the harder estimation problem. There is a basic tension between the constraint of Proposition 3 and predictive deterioration with a stronger new data signal because the former requires the cost of complexity to be bounded and the latter requires it to rise faster than conditional variance falls under the true model. For the special case of a diagonal second moment matrix $\boldsymbol{\Omega}_\zeta = \boldsymbol{D}_{\Omega_\zeta}$, which implies zero mean factors $\boldsymbol{\mu}_q = \mathbf{0}$ such that $\boldsymbol{D}_{\Omega_\zeta} = \boldsymbol{D}_{\Sigma_\zeta}$, and symmetric true factor loadings $\boldsymbol{\beta} = \bar{\beta}\mathbf{1}$ the juxtaposition is particularly clear since the constraint is

$$\begin{aligned} Var\left[E[y|\boldsymbol{\beta}, \boldsymbol{s}_I]\right] + (\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\mu}_q)^2 = \bar{\beta}^2 \mathbf{1}^\top \boldsymbol{D}_{\Sigma_\zeta} \mathbf{1} + 0 > k_{\sigma 0}^2 (1 + k_c^2)^{-1} \mathbf{1}^\top \boldsymbol{D}_{\Sigma_\zeta} \mathbf{1} = \chi \\ \iff \bar{\beta}^2 (1 + k_c^2) > k_{\sigma 0}^2, \end{aligned}$$

[8] By letting correlations decrease in k_S, it is possible to ensure that the lower bound of zero for conditional variance is respected, and with heterogeneity, in the effect, such a U-shape may appear, but in doing so one sacrifices tractability, and I leave the investigation of this extension for future work.

and the asymptotic condition for a positive derivative of the mean squared error with respect to new data parameter k_S, see Appendix A.4.1,

$$\lim_{k_S \to \infty} \frac{\partial \chi}{\partial k_S} = k_{\sigma 0}^2 (1 + k_c^2)^{-1} \mathbf{1}^\top \boldsymbol{D}_S \mathbf{1} > \bar{\beta}^2 \mathbf{1}^\top \boldsymbol{D}_S \mathbf{1} = \frac{\partial Var\left[y | \boldsymbol{\beta}, \boldsymbol{s}_I\right]}{\partial k_S}$$
$$\iff k_{\sigma 0}^2 > \bar{\beta}^2 (1 + k_c^2).$$

While the system of inequalities is only guaranteed to represent a contradiction asymptotically, numerical analysis suggests the intuition that a richer factor structure is necessary to accommodate these conflicting forces. Similarly to panel 1.1(d) in Figure 1.1, the restriction of Proposition 3 can be applied to ranges of technology parameter k_c^2 and new data parameter k_S. In the low difficulty case $k_{\sigma 0} = 0.3$ it does not restrict positive values of the two, whereas k_S is limited to be below 0.9 and k_c^2 above 0.8 for high difficulty $k_{\sigma 0} = 0.6$ and plots are adjusted accordingly.

In the first of the three remaining panels of Figure 1.1, panel 1.1(a), it is possible to see how initial increases in technology parameter k_c^2 introduces a bias of the estimator while lowering its volatility before eventually decreasing both. Despite the asymmetry of the chosen factor structure, the difference between factors is negligible compared to the difference between moments across both technology k_c^2 and difficulty $k_{\sigma 0}$ in panel 1.1(c). The comparative statics across new data parameter k_S in panel 1.1(b) represents more heterogeneity but also a striking symmetry whereby the bias and volatility of each factor visually mirrors one another. Due to the apparent mirroring, it is not obvious from this plot that the cost of complexity is increasing in k_S as stated in Proposition 2, however, the plots in Figure 1.2 show that it is indeed the case.

1.4.3 Return predictability

It is the introduction of bias that creates the ex-post predictability of realized price changes r, and bias is optimally chosen by investors to improve the precisions of their predictions. This is in contrast to models with parameter uncertainty (Lewellen and Shanken, 2002; Martin and Nagel, 2021), where bias is with respect to a rational expectations baseline where parameters of the model are known, while those parameters are random variables to agents who have to learn about their realization. Distinction betweeen in-sample and out-of-sample correspond to econometricians testing under the objective measure versus testing under the measure that investors use which I denote by $\boldsymbol{c}^*$, in a reference to the solution to the bias-variance minimization problem of equation (1.2). Econometricians make predictions through linear projections, which are tantamount to regressions, but estimation includes an estimator choice as in Section 1.2.1.

In the symmetric agents model of Section 1.2, my results are only the same as Martin and Nagel (2021), i.e. anomalies seem to exist in-sample, but disappear out of sample, when the quality of estimation technology, parametrized by k_c^2, that investors and econometrician use is the same, or econometrician match the optimal bias and active information set of investors, choice of data sources to include and ignore. By varying k_c^2 it is possible to analyse probable scenarios where econometricians do out of sampling testing, but have access to superior information estimation technology.[9] In this case, a different bias is optimal and I show in Section 1.4.3 that this is a source of predictability. Furthermore, due to the difference in cost of complexity, data sources that are freely available might optimally be ignored by investors using inferior estimation technology but used by econometricians.

The heterogeneous agents model of Section 1.3 introduces a second source of bias through the learning from prices. It arises from the weight uninformed investors put on their prior.[10] A specific concern for

[9] The objective measure corresponds to the limit $k_c^2 \to \infty$, the best possible technology, see Appendix A.1.8.

[10] Conceptually, econometricians cannot mimic the beliefs of a representative agent, which is a weighted average of the informed and uninformed beliefs (Biais et al., 2010).

the combination of these two sources of bias is that the out-of-sample testing required to correct the former does not correct the latter. Therefore, econometricians might find an echo of in-sample results in out-of-sample tests.

Returns

Returns can be decomposed into three components that relate to, respectively, the two groups of investors and stochastic supply

$$r = y - p = (1-\lambda_p)(\underbrace{y - E[\hat{y}_I]}_{\text{uninformed}}) + \lambda_p(\underbrace{y - \hat{y}_I}_{\text{informed}}) + \lambda_p \underbrace{\psi_I^{-1} z}_{\text{supply}}.$$

In the representative agent model of Section 1.2.1, only the second term remains (with a coefficient of $\lambda_p = 1$) since $p = \hat{y}$ (see Section 1.2). The inclusion of uninformed investors and stochastic supply introduces the first and last term. As a result, price responsiveness drops below one $\lambda_p < 1$ because the uninformed investors put some weight on their prior. There is no predictability in the supply term and stochastic supply is only relevant in the presence of uninformed investors through its effect on price responsiveness λ_p.

The predictability of the two investor terms can be analyzed by endowing econometricians with an estimation technology corresponding to optimal controls $\boldsymbol{c}_e^*$ that select the vector of signals $\boldsymbol{s}_e$. True out-of-sampling testing is the special case of same estimator $\boldsymbol{c}_e^* = \boldsymbol{c}^*$ and dataset $\boldsymbol{s}_e = \boldsymbol{s}_I$. Problematic in-sample testing corresponds to same dataset but evaluation under the objective measure. If they are available, in-sample testing will include a broader set of signals than the original estimation done by investors as the objective measure is the limiting measure in terms of estimation technology with zero cost of complexity (see A.1.8).

For ease of exposition, I focus on the case where signals employed by investors are a subset of the econometricians' signals $\boldsymbol{s}_I \subseteq \boldsymbol{s}_e$, and the additional signals in $\boldsymbol{s}_e$, if any, are uncorrelated with $\boldsymbol{s}_I$. I denote the vector of additional signals $\tilde{\boldsymbol{s}}_e$. This restriction is not consequential for *whether* econometricians find predictability or not, but helps to disentangle *where* it comes from.

I reserve the notation $\boldsymbol{\mu}_\beta$ and $\boldsymbol{\varepsilon}_\beta$ for the mean and bias of investors' estimator, e.g. $E[\hat{\boldsymbol{\beta}}|\boldsymbol{c}^*] = E[\boldsymbol{\beta}|\boldsymbol{c}^*] = \boldsymbol{\mu}_\beta$, and notice that in the dataset that econometricians work with these are constants. I denote econometricians' bias by $\boldsymbol{\varepsilon}_{\beta e}$. Finally, I assume that econometricians estimate a cross-sectional average, eliminating the variability in their estimate of factor loadings and, as such, its covariance with investors estimate $\hat{\boldsymbol{\beta}}$, which is in effect evaluated at its mean so $E[\hat{\boldsymbol{\beta}}|\boldsymbol{c}_e^*] = \boldsymbol{\mu}_\beta$.[11]

Proposition 4 (Predictability in returns)**.** *The contribution of the informed component is*

$$E[y - \hat{y}_I|\boldsymbol{c}_e^*, \boldsymbol{s}_e] = \boldsymbol{\beta}^\top (E[\boldsymbol{q}|\boldsymbol{s}_e] - E[\boldsymbol{\zeta}|\boldsymbol{s}_e]) + \boldsymbol{\varepsilon}_\beta^\top E[\boldsymbol{\zeta}|\boldsymbol{s}_e] - \boldsymbol{\varepsilon}_{\beta e}^\top E[\boldsymbol{q}|\boldsymbol{s}_e] \tag{1.6}$$

$$= (\boldsymbol{\beta} - \boldsymbol{\varepsilon}_{\beta e})^\top \boldsymbol{\Lambda}_{\tilde{e}}(\tilde{\boldsymbol{s}}_e - \boldsymbol{\mu}_{\tilde{e}}) + (\boldsymbol{\varepsilon}_\beta - \boldsymbol{\varepsilon}_{\beta e})^\top \boldsymbol{\zeta}, \text{ where } \boldsymbol{\Lambda}_{\tilde{e}} = \boldsymbol{\Sigma}_{q\tilde{s}_e}\boldsymbol{\Sigma}_{\tilde{s}_e}^{-1} \tag{1.7}$$

and the contribution of the uninformed component is

$$\begin{aligned} E[y - E[\hat{y}_I]|\boldsymbol{c}_e^*, \boldsymbol{s}_e] &= (\boldsymbol{\beta} - \boldsymbol{\varepsilon}_{\beta e})^\top (E[\boldsymbol{q}|\boldsymbol{s}_e] - \boldsymbol{\mu}_q) + (\boldsymbol{\varepsilon}_\beta - \boldsymbol{\varepsilon}_{\beta e})^\top \boldsymbol{\mu}_q \\ &= (\boldsymbol{\beta} - \boldsymbol{\varepsilon}_{\beta e})^\top \left\{\boldsymbol{\Lambda}_I(\boldsymbol{s}_I - \boldsymbol{\mu}_I) + \boldsymbol{\Lambda}_{\tilde{e}}(\tilde{\boldsymbol{s}}_e - \boldsymbol{\mu}_{\tilde{e}})\right\} + (\boldsymbol{\varepsilon}_\beta - \boldsymbol{\varepsilon}_{\beta e})^\top \boldsymbol{\mu}_q, \\ &\quad \textit{where} \quad \boldsymbol{\Lambda}_I = \boldsymbol{\Sigma}_{qs_I}\boldsymbol{\Sigma}_{s_I}^{-1} \end{aligned}$$

Proof. The explicit expectations in equation (1.6) does not depend on $\boldsymbol{c}_e^*$ due to Assumption 2, estima-

[11] Alternatively, the exercise can be viewed as an analysis of the unconditional expected coefficients on signals.

tor noise independence from factors and signals, and the cross-sectional mean assumption implies $E[\hat{\boldsymbol{\beta}}|\boldsymbol{c}_e^*] = \boldsymbol{\beta} - \boldsymbol{\varepsilon}_\beta$. Equation (1.7) uses the independence between $\tilde{\boldsymbol{s}}_e$ and $\boldsymbol{s}_I$ which yields

$$E[\boldsymbol{q}|\boldsymbol{s}_e] = \boldsymbol{\mu}_q + \boldsymbol{\Lambda}_I(\boldsymbol{s}_I - \boldsymbol{\mu}_I) + \boldsymbol{\Lambda}_e(\tilde{\boldsymbol{s}}_e - \boldsymbol{\mu}_q) = \boldsymbol{\zeta} + \boldsymbol{\Lambda}_{\tilde{e}}(\tilde{\boldsymbol{s}}_e - \boldsymbol{\mu}_q).$$

The derivation of the uninformed component follows from re-arranging terms after substituting in $E[y|\boldsymbol{c}^*] = \boldsymbol{\mu}_\beta^\top \boldsymbol{\mu}_q$ and once again applying the assumption of independence of $\tilde{\boldsymbol{s}}_e$ and $\boldsymbol{s}_I$ in the second line. □

In contrast to Martin and Nagel (2021), out-of-sample estimation will still be biased if econometricians have access to better technology, especially if that technology leads them to use signals that were available, but too complex to be beneficial at the time of investment. In Figure 1.4, the estimation quality of the econometricians is increasing and eventually lead to introduction of two additional signals. This affects not only their own predictive coefficients but also those of the actually used signals even though the two are independent. The latter effect is driven by the change in econometricians' own bias, which shifts discretely away from that of investors, visible in the break in the curve of Figure 1.4(c).

The coefficients on the vector of additional signals is the same for the informed and uninformed component, $\partial E[r|\boldsymbol{c}_e^*, \boldsymbol{s}_e]/\partial \tilde{\boldsymbol{s}}_e = (\boldsymbol{\beta} - \boldsymbol{\varepsilon}_{\beta e})^\top \boldsymbol{\Lambda}_{\tilde{e}}$, so they are unaffected by the price responsiveness λ_p introduced by the presence of uninformed investor. It is also unaffected by investors bias because these signals are ignored. Therefore, the coefficients on these unused signals are relatively large in absolute terms across the two scenarios, positive or negative true factor loadings, in Figure 1.4. Assuming variables are properly demeaned, the constant term of the projection is the difference in biases scaled by factor means $E[r|\boldsymbol{c}_e^*, \boldsymbol{s}_e]|_{\boldsymbol{s}_e = E[\boldsymbol{s}_e]} = (\boldsymbol{\varepsilon}_\beta - \boldsymbol{\varepsilon}_{\beta e})^\top \boldsymbol{\mu}_q$. The coefficients on the used signals also depends on the difference in biases

$$\frac{\partial E[r|\boldsymbol{c}_e^*, \boldsymbol{s}_e]}{\partial \boldsymbol{s}_I} = \left\{\lambda_p(\boldsymbol{\varepsilon}_\beta - \boldsymbol{\varepsilon}_{\beta e}) + (1-\lambda_p)(\boldsymbol{\beta} - \boldsymbol{\varepsilon}_{\beta e})\right\}^\top \boldsymbol{\Lambda}_I = \left\{\lambda_p \boldsymbol{\varepsilon}_\beta + (1-\lambda_p)\boldsymbol{\beta} - \boldsymbol{\varepsilon}_{\beta e}\right\}^\top \boldsymbol{\Lambda}_I, \tag{1.8}$$

so when the technological gap between investors and econometricians is not too large, coefficients on used signals and the constant are smaller, because biases are similar. In the heterogeneous agent model, the true factor loadings in the coefficients on used signals are scaled down by a factor of $(1-\lambda_p)$ compared to the unused signals' coefficients. The representative agent model is at the extreme end of this scaling with the true factor loading term equal to zero.

Only the coefficient on used signals are sensitive to the level of price responsiveness, and as such marks the difference between the representative agent model and the heterogeneous agent model. On the one hand, this means that even if it is possible to control for estimation technology and match investors bias, these coefficients are never zero as would be the null hypothesis in most empirical work and they are sensitive to the out-of-sample echo mentioned above. On the other hand, it is possible to distinguish them from unused signals and other components of returns, such as risk premia, through their sensitivity to variations in price responsiveness.

Return predictability and noise trading

I focus on the level of noise trading σ_z^2 as a source of variation in price responsiveness since it does not affect predictability through other channels, and can more reasonably be taken as exogenous than share of uninformed investors which would be the closest alternative.[12] The scenario in Figure 1.3

[12] While the extension to a variable share of informed versus uninformed investors is straightforward, (see Appendix A.2.6) allowing the share of informed versus uninformed to vary begs an optimization in the spirit of the original paper to find the optimal share and the introduction of a traditional information acquisition problem, which is well-studied elsewhere and

is the easy estimation problem $k_{\sigma 0}$ where there is no difference in investors' and econometricians' information set and all signals, therefore, are affected by shocks to price responsiveness λ_p. In the parametrization of the numerical analysis, the effect of having only half the amount of noise trading, i.e. $\sigma_z^2 = 0.5$ instead of $\sigma_z^2 = 1$, is to lower price responsiveness from 0.92 to 0.87. While this is a modest decrease it is enough to visualize the effect that can also be read of equation (1.8), which is to shift coefficients by $(\boldsymbol{\beta} - \boldsymbol{\varepsilon}_\beta)^\top \boldsymbol{\Lambda}_I$. Signals that, through the variance-adjusted covariance matrix $\boldsymbol{\Lambda}_I$, load heavily on factors with a difference between factor loading and bias $\beta_i - \varepsilon_{\beta i}$ of the same sign as the signals coefficient will have their coefficients amplified. In markets where more noise trading drives price responsiveness down (which might be considered the natural direction), signals with biases that lead to attenuation of true factor loadings towards zero (including special case $\varepsilon_{\beta i} = 0$) have *larger* coefficients in predictive projections when there is *more* noise trading. This is potentially problematic because noise trading, for good reasons, often is associated with behavioral biases. Even if such behavioral biases effectively generate random noise, a pattern identified in a broad sample may be amplified in a sub-sample that would appear especially representative. It is, however, also a possibility to distinguish between return components as discussed above.

1.4.4 Price informativeness

The impact of the specifics of investors' inference problem on price informativeness or market efficiency as referred to by Ozsoylev and Walden (2011a) has only become more prominent in the context of big data and machine learning on financial markets, see Dugast and Foucault (2020) and Farboodi et al. (2020). In the classic formulation of Grossman and Stiglitz (1980), price informativeness is the inverse of the variance of the pay-off conditional on equilibrium price. When price and pay-off are normal random variables, this measure coincides with the mean squared error of a projection, which has led to empirical estimation strategies based on regression analysis, see Dessaint, Foucault, and Frésard (2020). For comparability, I also base my measure of price informativeness on a projection. As discussed in Section 1.3.1, this projection is also the best linear approximation. Price informativeness is given by

$$E[(y - E[y|p, \boldsymbol{\beta}])^2]^{-1} = \left\{ Var[y] - \frac{\lambda_p^2}{\lambda_p^2} \frac{Cov[y, \hat{y}_I]^2}{Var[s_U]} \right\}^{-1} = \left\{ \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_q \boldsymbol{\beta} - \frac{(\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\mu}_\beta)^2}{Var[\hat{y}_I] + \psi_I^{-2} \sigma_z^2} \right\}^{-1}. \tag{1.9}$$

Equation (1.9) shows how price responsiveness λ_p does not affect price informativeness. The effect of uninformed trading on price informativeness is just to scale supply by the square inverse scaling factor of the informed investors, ψ_I^{-2}, while larger supply σ_z^2 unambiguously decreases it. This is true in the baseline model of known factor loadings as well. The difference is in the inflation of variance of the informed predictor compared to the variance of the expectation with known factors, and the attenuation of the covariance

$$Var[\hat{y}_I] - Var\left[E[y|\boldsymbol{\beta}, \boldsymbol{s}_I]\right] = \chi - 2\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta, \quad Cov[y, \hat{y}_I]^2 = (Var\left[E[y|\boldsymbol{\beta}, \boldsymbol{s}_I]\right] - \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta)^2.$$

Higher estimation noise $\boldsymbol{\sigma}_\beta$ unambiguously inflates the variance of the predictor through the cost of complexity, while the role of the bias is less clear cut. However, numerical analysis supports the tendency of variance inflation and covariance attenuation. The presence of the interaction term $\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta$ in both the denominator and numerator implies that if a configuration of bias amplifies rather than attenuates covariance it will with at an even higher inflation of variance.

Taking noisy inference as a given, the unattainable base line is less interesting than the fact that price informativeness under this condition is subject to a trade-off between bias and variance just as the predictor of (informed) investors. Figure 1.6 is based on, respectively, the equilibrium inverse

beyond the scope of this paper.

price informativeness (mean squared error), and the counterfactual of a social planner optimizing price informativeness, i.e. solving the non-linear problem of maximizing equation 1.9 with respect to controls $\boldsymbol{c}$ that control the bias and variance of the estimator $\hat{\boldsymbol{\beta}}$ (see Section 1.2.2). Carried out under the objective measure, a stronger new data source signal (higher k_S) raises the price informativeness (lowers the mean squared error), even though the baseline difficulty is high $k_{\sigma 0} = 0.6$, which means that the mean squared error of the informed predictor is increasing (see Figure 1.2). Of the three dimensions considered, only sophisticated estimation technology (high k_c^2) leads to conversion between the private optimization and the planner's optimization. While the two other parameters are restricted by the requirement that informed predictions outperform (Proposition 3), the tendency to convergence happens in the unrestricted direction towards zero, but convergence is not reached before the hard cut-off at zero. A planner with an ability to invest in improving technology $\uparrow k_c^2$, lowering the baseline difficulty $\downarrow k_{\sigma 0}$ or making new data sources available $\uparrow k_S$, would find all beneficial but might prefer the first as it aligns the goals of private individuals with its own. Especially since the incorporation of new data sources might require better technology to be attractive to private individuals, as it is possible for informed investors' predictive power to be deteriorating even while price informativeness is improving.

1.4.5 Excess price volatility

Another measure of market quality is price volatility and in particularly in excess of the volatility of dividends (Shiller, 1980). In the model, this comparison correspond to the contrast between price variance

$$Var[p] = \lambda_p^2 Var[s_U] = \lambda_p^2 \left\{ Var[\hat{y}_I] + \psi_I^{-2} \sigma_z^2 \right\},$$

and the variane of the pay-off $Var[y]$. In the baseline model with known factor loadings, conditional expectation of the pay-off is $Var[y|\boldsymbol{\beta}, \boldsymbol{s}_I] = Var[y] - Var[E[|y|\boldsymbol{\beta}, \boldsymbol{s}_I]]$, and $Var[E[|y|\boldsymbol{\beta}, \boldsymbol{s}_I]]$ replaces $Var[\hat{y}_I]$ in $Var[p]$. As such, excess price variance is not possible in the representative agent model. In the heterogeneous agent model it requires large amounts of noise trading and the off-setting effect of uninformed traders correcting for it to be small[13] (with no uninformed demand, price responsiveness is unity $\lambda_p = 1$). The additional noise from the estimation process makes excess price variance more prevalent in the parameter space and it can occur even in the representative agent model of Section 1.2, i.e. $Var[\hat{y}_I] > Var[y]$ is possible.

1.4.6 Short-term price reversals

There is a long tradition of using short-term price reversals to disentangle informed trading from liquidity demand, see Hendershott and Menkveld (2014), since the price pressure from liquidity demand can be expected to be reversed in the spirit of Campbell et al. (1993). Predicting price change by price gets at short term price reversals in the context of a two period model assuming a constant before-price of $E[p]$, see Breon-Drish (2015). I first illustrate how the price pressure of liquidity demand relates to price reversals before adding realized estimated factor loadings $\hat{\boldsymbol{\beta}}$ as an additional conditioning variable. I once again apply projection as the linear approximation of the expectation

[13]The effect of more noise trading on λ_p has two counter-acting forces. More noise leads the uninformed to put more weight on their prior ($\downarrow \lambda_U$), which lowers λ_p, but they also take smaller positions ($\downarrow \psi_U$), which increases λ_p as informed investors and noise traders make up a bigger share of the market. Conversely, when uninformed investors form their best estimate by copying informed investors as best as they can (see Section 1.3.1) they do not adjust the same way for estimation noise as noise trading. Because higher estimation noise goes into $Var[\hat{y}_I]$, they will actually *increase* the weight they put on their signal derived from price ($\uparrow \lambda_U$).

and in, addition to price, I condition on negative stochastic supply z (liquidity demand)

$$\tilde{E}[r|p,z] = \boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\mu}_q + \frac{Cov[y,\hat{y}_I] - \lambda_p Var[\hat{y}_I]}{\lambda_p Var[\hat{y}_I]}(p - E[p]) - \frac{Cov[y,\hat{y}_I]}{\lambda_p Var[\hat{y}_I]}\psi_I^{-1} \underbrace{(-z)}_{\substack{\text{liquidity}\\ \text{demand}}},$$

(derivations that can be found in Appendix A.3.1). The non-zero constant of the projection $\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\mu}_q$ generates a tendency for price drift that might over time be picked up as momentum or long term reversal. This component is analysed in Section 1.4.3 on return predictability and here I focus on short-term reversals instead.

It is useful to consider that in the baseline model without noisy inference, covariance of the pay-off and the informed predictor is the variance of the predictor $Cov[y, E[y|\boldsymbol{\beta}, \boldsymbol{s}_I]] = Var[E[y|\boldsymbol{\beta}, \boldsymbol{s}_I]]$. Therefore, the coefficients simplify to $(1-\lambda_p)/\lambda_p$ and $-\psi_I^{-1}/\lambda_p$ respectively, and positive supply shock tends to be followed by negative return. Directionally, the impact of a large supply shock is the same as in the baseline model with known factor loadings if the covariance between informed prediction and pay-off is positive. It is still positive but it is attenuated if $Cov[y,\hat{y}_I] < Var[\hat{y}_I]$, which is likely as a sufficient condition is for this inequality is $Var[y] < Var[\hat{y}_I]$.

More important for interpreting empirical results on reversals, conditioning on the realization of the factor loadings in the projections yields identical coefficients on rescaled negative supply $-\psi_I^{-1} z$ and factor-mean weighted realized factor loadings $\boldsymbol{\mu}^\top \hat{\boldsymbol{\beta}}$.

Proposition 5 (Short-term reversals)**.** *Both liquidity demand* $(-z)$ *and realized estimated factor loadings* $\hat{\boldsymbol{\beta}}$ *generate expected short-term reversals with a common factor in marginal effects*

$$\begin{aligned}\tilde{E}\left[r|p,z,\hat{\boldsymbol{\beta}}\right] &= \boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\mu}_q + \frac{Cov[y,\hat{y}_I] - \lambda_p(Var[\hat{y}_I] - \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q)}{\lambda_p(Var[\hat{y}_I] - \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q}(p - E[p]) \\ &- \frac{Cov[y,\hat{y}_I]}{\lambda_p(Var[\hat{y}_I] - \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q)}(-\psi_I^{-1} z) - \frac{Cov[y,\hat{y}_I]}{\lambda_p(Var[\hat{y}_I] - \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q)}\boldsymbol{\mu}_q^\top(\hat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta).\end{aligned}$$

Proof. See Appendix A.3.1. □

While it is unreasonable to expect econometricians to directly observe either stochastic supply z or factor loadings estimates $\hat{\boldsymbol{\beta}}$, the significance of Proposition 5 is that any variable that correlates with price reversals could be correlated with liquidity demand, but could equally well be correlated with noisy estimation. Whether it is meaningful to group these two types of noise into one will depend on the context. In contrast to the results in Section 1.4.3, which were largely driven by the bias of the noise inference the effect for short-term reversals is a function of the variance of the estimation process.

Short term reversals can also be linked to price variance (see Section 1.4.5) since higher estimation noise tends to attenuate coefficients on both the price, supply shocks, and realized factor loadings per Preposition 5. Higher trading noise only affects coefficients through price responsiveness λ_p which will tend to decrease as uninformed investors put more weight on their prior. Analysing price variance and price reversals is, therefore, useful for understanding whether a likely common factor is: estimation noise (attenuated coefficients), or trading noise (amplified coefficients).

In the limiting case of the estimation technology parameter k_c^2 growing large presented in Figure 1.7, both of these relations exist, and there is additionally a negative relation between the coefficient on price and price variance when driven by estimation noise and a positive one when driven by trading noise. For the parametrization of this numerical analysis, the relations between trading noise, price variance, and reversal coefficients are consistent across different levels of technological sophistication for the case of attenuation bias captured by $+\bar{\beta}$. The case of amplifying bias, $-\bar{\beta}$, is more complicated because the relation between the base-parameter of estimation difficulty $k_{\sigma 0}$ and price variance changes over the range of the estimation technology parameter k_c^2, from being positive for lower values to eventually being negative and some instability in between. The interesting aspect of the limiting case is how the relationship between price variance and the reversal coefficients again prevails despite the changes in the relationship with the base parameter with amplification bias, i.e. price variance is decreasing in $k_{\sigma 0}$.

It is not the only one possible, but from an empirical perspective the most straightforward identifying assumption is that estimation bias is dominated by attenuation bias, in which case the opposing relations between price variance and the price reversal coefficient emerges. It is additionally attractive because of its significance in turning noise trading into predictability in returns as covered in Section 1.4.3.

Trading volume

With several explanations for returns, a natural additional dimension of market data to consider is trading volume following Campbell et al. (1993). Realized trading volume v is given by

$$
\begin{aligned}
v &= \frac{1}{2}\{|\delta_I| + |\delta_U| + |z|\} = \frac{1}{2}\{|\delta_I| + |-\delta_U| + |z|\} \\
&= \frac{\psi_I}{2}\left\{\left|(1-\lambda_p)(\hat{y}_I - E[\hat{y}_I]) - \lambda_p \psi_I^{-1} z\right| + (1-\lambda_p)\left|(\hat{y}_I - E[\hat{y}_I]) + \psi_I^{-1} z\right| + |\psi_I^{-1} z|\right\},
\end{aligned}
$$

see Appendix A.3.2. From this expression it is possible to point out a difference between liquidity demand and noisy estimation. Other things equal, the impact of an absolute increase in stochastic supply/liquidity demand $|z| = |-z|$ is always positive for non-zero demand and supply since

$$
\frac{\partial v}{\partial |z|} = \frac{1}{2}\psi_I^{-1}\left\{-\lambda_p \operatorname{sign}(\delta_I)\operatorname{sign}(z) - (1-\lambda_p)\operatorname{sign}(\delta_U)\operatorname{sign}(z) + 1\right\} \geq 0.
$$

Conversely, an informed prediction with larger absolute deviation from its expected value is ambiguous

$$
\frac{\partial v}{\partial |\hat{y}_I - E[\hat{y}_I]|} = (1-\lambda_p)\operatorname{sign}(\hat{y}_I - E[\hat{y}_I])\left\{\operatorname{sign}(\delta_I) - \operatorname{sign}(\delta_U)\right\},
$$

however, for large enough deviations the effect will be positive as the trading between informed and uninformed investors dominates the trading flow. The classic intuition that conditioning on large trading volume with low expected returns helps to identify liquidity demand, following Campbell et al. (1993), breaks down in the presence of noisy inference.

1.4.7 Fund performance

By identifying informed investors as sophisticated funds who invest in inference technology and the data it requires and uninformed investors as their simpler counterparts, I compare performance as measured by their expected profits. Profit of investors under a measure $\boldsymbol{c}_j^*$ is derived in Appendix A.3.3,

$$
\begin{aligned}
E[\pi_I] &= \psi_I\left\{(1-\lambda_p)\left(Cov[y, \hat{y}_I|\boldsymbol{c}_j^*] - \lambda_p Var[\hat{y}_I]\right) + \lambda_p^2 \psi_I^{-2}\sigma_z^2\right\}, \text{ and} \\
E[\pi_U] &= \psi_U\left\{(\lambda_p - \lambda_U)\left(\lambda_p Var[\hat{y}_I] - Cov[y, \hat{y}_I|\boldsymbol{c}_j^*]\right) + (\lambda_p - \lambda_U)\lambda_p \psi_I^{-2}\sigma_z^2\right\} \\
&= \psi_I(1-\lambda_p)\left\{\left(\lambda_p Var[\hat{y}_I] - Cov[y, \hat{y}_I|\boldsymbol{c}_j^*]\right) + \lambda_p \psi_I^{-2}\sigma_z^2\right\},
\end{aligned}
$$

see Appendix A.3.3. Contrasting ex-ante expected profits under investors' measure $\boldsymbol{c}^*$ with the large numbers average realized profit (expectation under objective measure or equivalent limiting measure $\boldsymbol{c}^*_\infty$) comes down to comparing $Cov[y, \hat{y}_I|\boldsymbol{c}^*]$ and $Cov[y, \hat{y}_I]$. The difference between the two defines the ex-post surprise investors are expected to experience under the model. In Appendix A.3.3, I show that there is no ex-post surprises with respect to the total profits of the investor base and the surprise is a transfer between informed and uninformed investors equal to

$$E[\pi_I] - E[\pi_I|\boldsymbol{c}^*] = -(E[\pi_U] - E[\pi_U|\boldsymbol{c}^*]) = \psi_I(1-\lambda_p)(Cov[y,\hat{y}_I] - Cov[y,\hat{y}_I|\boldsymbol{c}^*]).$$

The condition for a surprise ex-post out-performance of informed investors, and by symmetry the under-performance of uninformed investors, is whether the covariance between the pay-off and the informed predictor is over- or underestimated under the contemporary measure compared to the objective measure. In Appendix A.3.3, I link expected profit under the contemporaneous measure, i.e. expected out-performance by informed investors, to the condition for informed investors making better predictions presented in Proposition 3. Under the parametrization of the numerical analysis, Figure 1.8 establishes a clear relationship between this condition and the sign of $\bar{\beta}$, and Figure 1.9 shows that it carries over to profits. With attenuation bias ($+\bar{\beta}$), the out-of-sample surprise is in favor of informed investors, (sophisticated investors, for some period perhaps quantitative funds) and with amplification bias, the effect is in the direction of hype followed by disappointment. Linking these conditions to Section 1.4.3 and Section 1.4.6 the common environment is characterized by the shared condition of attenuation bias, and fund performance is a way to identify periods where this condition is likely true (under the model). The period leading up to a period of out-performance by quantitative funds (or another identifier of informationally sophisticated funds) is a good candidate for the joint analysis of return predictability and short-term price reversals.

1.5 Predictability of the equity risk premium

In this section, I focus on a specific case of predictability in the asset pricing literature that the noisy estimation with changing technology provides an explanation for: the predictability of the equity risk premium, specifically with respect to the predictive variables surveyed by Welch and Goyal (2008) and the literature inspired hereby (Campbell and Thompson, 2008; Rapach et al., 2010; Neely et al., 2014; Buncic and Tischhauser, 2017; Hammerschmid and Lohre, 2018). Variation in estimation technology offers an explanation for two empirical patterns found within and across these studies: vanishing predictability over time (in the data) but stronger predictability overall between studies (see Section 1.5.1). As a single time series with variation through predictive variables rather than a cross-section of assets the application is better suited of the narrower representative agent model of Section 1.2 than the extended heterogeneous agent model of Section 1.3. Therefore, in Section 1.6, I provide some considerations for further empirical work collecting and extending observations made in Section 1.4 and Section 1.4.1.

1.5.1 Predicting the equity risk premium

An empirical pattern found in Welch and Goyal (2008) and confirmed in later studies with other empirical strategies and auxiliary data (e.g. Buncic and Tischhauser (2017)) is that of, over time, an initial out-performance of the historical mean followed by deteriorating performance[14] driven by the variables identified in the first study. The turn-around falls in the earlier 1990s and as such it follows the rise of the private computer in the 1980s and coincides with the early years of the internet. A

[14] Dependent on specific set-up this might mean under-performance or performance only on par with that of the historical mean. Introducing a floor of zero for the prediction of the equity risk premium as proposed by Campbell and Thompson (2008), generally helps to avoid under-performance.

second pattern that appears *between* studies is one of empirical approaches in subsequent papers out-performing approaches in earlier papers. These effects have a natural interpretation in terms of the estimation technology of respectively investors and econometricians.

In the model, the parameter k_c^2 can be interpreted as the quality of estimation technology. By providing subscripts I for investors and e for econometricians, the first empirical pattern can be understood as fixing k_{ce}^2 and increasing k_{cI}^2. Within a given study the empirical method used by econometricians is the same through time while the data it is applied to changes. Over time, the estimation technology of investors could reasonably be improving as new techniques become available.

The second empirical fact, can be seen as fixing a path for k_{cI}^2 and increasing k_{ce}^2 where later studies represents a better estimation technology than earlier. Abstracting estimation technology into the single dimension of k_c^2 suggests that there is a hierarchy of methods. A priori, it is not obvious how to rank approaches and without assumptions on the data-generating process no such ranking can exists given the no free lunch theorem for learning algorithms (Wolpert, 1996). However, ex posteriori, some empirical strategies should outperform others if there is something to learn from the data. One point to address, is that out-performance between studies coincides with the introduction of new data. As discussed in Section 1.4.1, in the model, data that is used for estimation of the factor loadings is reflected in k_c^2 (or $k_{\sigma 0}$, the estimation difficulty). Conceptually, however, allowing this dimension to vary at the same time as fundamental data, the structure imposed by the model is loosened substantially. Therefore, in my replication of these patterns in Figure 1.12, I do not generate the second pattern by replicating the studies directly. Instead, I fix the data used and shows how regularised linear approaches can outperform plain vanilla ordinary least squares.[15] In the calibration of Section 1.5.1, I focus on the first empirical pattern and investigate which shift in investors estimation technology k_{cI}^2 is required to generate it.

Calibration

In the calibration, I focus on moments that rely only on a subset of parameters. I consider ten of the predictive variables studied by Welch and Goyal (2008) and use the updated time series data from Amit Goyal's website.[16] In addition to the assumption that the representative agent model is a meaningful first-order representation of the phenomenon, I operate under the identifying assumption that investors' signal set is the same over the period. With the discussion of the potency of a mismatch between employed signals and available signals in Section 1.4.3, it is a very relevant alternative hypothesis, and one to keep in mind when considering the results of the calibration. Approaching it directly, however, requires richer data as it introduces more degrees of freedom. The first set of moments is the difference in variance adjusted expected coefficients on the econometricians signals between the two sub-periods

$$\left\{E\left[\frac{\partial E[r_2|\boldsymbol{c}_e^*,\boldsymbol{s}_e]}{\boldsymbol{\partial} s_e}\right]-E\left[\frac{\partial E[r_1|\boldsymbol{c}_e^*,\boldsymbol{s}_e]}{\partial \boldsymbol{s}_e}\right]\right\}Var[\boldsymbol{s}_e]=\left(\boldsymbol{\varepsilon}_{\beta 2}-\boldsymbol{\varepsilon}_{\beta 1}\right)^\top \boldsymbol{D}_{\sigma_q}\boldsymbol{R}_{qs_I}\boldsymbol{R}_{s_I}^{-1}\boldsymbol{R}_{s_I s_e}$$

empirically estimated by the averages of the coefficients from rolling regressions with 30 years windows.[17] Notice how keeping the econometricians estimation technique fixed along with the signal set

[15] Per the discussion of the impossibility of an a priori ranking of methods it is not generally clear that such an out-performance should exists, but the application is inspired by the uses of regularization in later studies (Rapach et al., 2010; Buncic and Tischhauser, 2017)

[16] The variables are listed in Table 1.5.

[17] In Figure 1.10 and Figure 1.11 I investigate the in-sample fit of expanding and contracting as well as rolling windows. The pattern for shorter windows is consistent with some amount of over-fitting in sample as the *shorter* samples achieve noticeably higher scores but have very poor untabulated out-of-sample performance. The interpretation of in-sample analysis as converging to the objective measure, following Martin and Nagel (2021), is not reasonable for the shorter windows and I focus on 20 and 30 years windows instead.

of the investor means that terms involving the bias of the econometricians drop out. The second set, relies on a different assumption to eliminate terms involving the econometricians bias, which is that the benefit of in sample estimation eventually makes the estimation error of econometricians on a given data-set small in comparison to investors. It is the variance adjusted expected coefficients on the econometricians signals over the full sample

$$E\left[\frac{\partial E[r|\boldsymbol{c}_e^*, \boldsymbol{s}_e]}{\boldsymbol{\partial} s_e}\right] Var[\boldsymbol{s}_e] = \left[w\boldsymbol{\varepsilon}_{\beta 2} + (1-w)\boldsymbol{\varepsilon}_{\beta 1}\right]^\top \boldsymbol{D}_{\sigma_q}\boldsymbol{R}_{qs_I}\boldsymbol{R}_{s_I}^{-1}\boldsymbol{R}_{s_I s_e}$$

where the weights are approximated by the number of observations in the two sub-periods. In Figure 1.13, I confirm that the coefficients targeted can generate the empirical pattern of predictive out-performance followed by deteriorating performance when applied to the data. Finally, I include the difference in the unconditional expectation,

$$E[r_2] - E[r_1] = \left(\boldsymbol{\varepsilon}_{\beta 2} - \boldsymbol{\varepsilon}_{\beta 1}\right)^\top \boldsymbol{\mu}_q,$$

as a target moment.

I fix a number of parameters up front and limit the estimation to the correlation structure and the second period investor estimation technology quality parameter k_{cI2}. I maintain the structure from the numerical analysis of two factors with common mean and variance, $\bar{\mu}_q$ and $\bar{\sigma}_q$, but summarize investors' group of signals in one signal which requires two correlation parameters, $\rho_{qs_I 1}$ and $\rho_{qs_I 2}$. Varying $\bar{\mu}_q$, $\bar{\sigma}_q$, the baseline $k_{\sigma 0}$, and first period estimation difficulty k_{cI1} has little impact on the directional effects as long as certain relations are maintained to ensure convergence, $\bar{\mu}_q < \bar{\sigma}_q$ and $k_{\sigma 0} \approx |k_{cI1}|$. I run the estimation iteratively in a two step procedure alternating between minimizing the mean squared error between theoretical and empirical moments over the correlation parameters and then k_{cI2}. I find that then convergence occur, it happens after a few iterations.

I pick $k_{\sigma 0}$ and $\bar{\mu}_q$ to match the baseline values studied in the numerical analysis and calibrate $\bar{\sigma}_q$ such that the correlations between each factor and investors' signal group come out with comparable magnitudes in the estimation, $\rho_{qs_I 1} \approx 0.3$ and $\rho_{qs_I 2} \approx -0.21$. Apart from these two correlation parameters, I estimate a correlation parameter for each of the ten predictive variables,which, from the perspective of the model, are the econometricians signals. Most parameter values can be found in Table 1.6, but the correlation parameters for the predictive signals I instead visualize in Figure 1.14, which clearly shows that the magnitude of correlation with investor signals is the largest for the valuation ratios of dividend- and earnings-to-price. Since the estimation is done with respect to variance-covariance corrected measures, should not be taken to mean that these are the only relevant predictive variables. Rather, valuation is estimated to be the strongest channel through which the variables relate to investors signals which is in line with the view that this group of predictive variables represent signals about fundamentals as opposed to, for instance, sentiment.

The key number of interest is the shift in estimation technology. A direct way to look at it is through the percentage increase in the magnitude of parameter k_{cI} which comes out to $k_{cI2}/k_{cI1} - 1 \approx 233\%$. For a more contextual view that also reflects the estimated information structure, the change in investors optimal bias can be calculated based on the calibration. With two factors the vector has two elements which are virtually the same both experiencing a growth of $\varepsilon_{\beta i,2}/\varepsilon_{\beta i,1} - 1 \approx 82\%$. Taking the information structure into account the magnitude of the shift is attenuated substantially. However, what is perhaps more surprising about this estimate is the direction of the shift. The estimated optimal bias has *grown*. Since there is a trade-off between bias and variance increasing bias and decreasing cost of complexity both follow from the increasing quality of estimation technology. For a visual demonstration of this mechanism see Figure 1.1(a). Recognizing that bias can be optimal helps to explain the empirical pattern.

The shift in estimation technology is substantial, which might to a certain extent reflect the discretization into just two periods. The turnaround in Figure 1.12 is, however, rather sharp and the alternative hypothesis of technological development also changing the composition of investors signal group is reasonable. Either way, the calibration is consistent with technological development historically playing a large role in predictability of returns.

1.6 Further directions for empirical work

One of the key challenges in working empirically with the model, is that taking the concern that both investors and econometricians face an estimation problem in forming predictions about pay-offs or returns seriously makes it problematic to estimate key ingredients of the model such as the cost of complexity directly from the data. Extending to the cross-section, separating the estimation of $\hat{\boldsymbol{\beta}}$ and factor expectations $\boldsymbol{\zeta}$ à la Kelly, Pruitt, and Su (2019) or in a non-linear fashion as in Gu et al. (2021) would allow for more targeted contrasts of sources predictability. E.g. once a factor structure is estimated cost of complexity under various signal configuration and parameter choices could be matched to moments of predictability in the cross-section. Alternatively, separating returns into components that more closely match what is pay-off and price in the model, by proxying expected pay-offs by earnings expectation or focusing on returns around earnings-announcement would again allow a more targeted estimation.

An alternative approach would be to contrast the performance of different methods of analyzing historical returns to imitate the model concept of different levels of technological quality. Since the bias generated by investors is part of the historical data it does not vary with the estimation method applied by econometricians.

Irrespectively of the cross-sectional approach, the heterogeneous agent extension highlights promising ways to sub-sample the data. Most directly related to return predictability, the decomposition of signals available to in Section 1.4.3 shows how only signals employed by investors are affected by price responsiveness. This suggest an empirical strategy of sub-sampling on proxies for noise trading to sort components of predictability on their sensitivity to this cross-sectional variation.[18] Another prediction related to noise trading is that when higher noise trading leads to lower price responsiveness the coefficient on signals estimated with attenuation bias are larger in sub-samples with more noise (see Section 1.4.3). While it might be natural to assume that attenuation bias is more common than amplification that can generally be an identifying assumption that is hard to provide direct support for. In the context of the model, relations between price variance and short-term price reversals with respect to levels of noise trading is particularly stable under attenuation bias (see Section 1.4.6), and in the time series, fund performance can be related to the dominant type of bias, as periods of success (better than expected performance) for sophisticated investors are more likely under attenuation bias (see Section 1.4.7).

An additional avenue for empirical tests are comparative analysis of segregated markets. This could be international markets that exhibit different degrees of technological sophistication, or assets with different but related fundamentals such as stocks and bonds of companies with both instruments outstanding.

Finally, focusing on developments in technological limitations, a promising setting is intra-day data where the constraints on re-training models have become less binding with improvements in computing power.

[18]These component may be predictive variables or relevant transformations of the data.

1.7 Conclusion and perspective

The complex prediction problems faced by investors in financial markets have a number of implications for equilibrium outcomes. Complexity generates a cost of expanding predictive models of the risky pay-off with new signals. Investors optimally trade off bias for precision and the benefit of including a signal for the associated cost of complexity. Advances in estimation techniques such as machine learning methods mitigate the issues of complexity for investors. In the study of financial markets, these advanced methods, however, require careful application to undo rather than amplify the bias optimally introduced by investors. Empirically, the effect of changing estimation technology on predictability is not only relevant for new methods going forward but and can explain historical patterns of predictability as well. For new methods that likely lowers the cost of complexity, a high level of caution is warranted in assessing their likely future performance through historical back-testing. Analysing historic performance over time rather than summarising performance in an aggregate statistic is a mitigating measure, as well as analysis of cross-sectional sub-samples in various dimensions especially around proxies for noise trading.

Tables and Figures of Chapter 1

Table 1.1: Baseline parameter values.

	Notation	Value
Estimation technology level	k_c^2	1
Difficulty of estimation problem	$k_{\sigma 0}$	0.3
New data sources strength	k_S	0.5
Common component in factor loadings	$\bar{\beta}$	0.6
Individual components in factor loadings	$\tilde{\boldsymbol{\beta}}$	$(1\ 0.3)^\top$
Common uncertainty aversion	α_I, α_U	1
Intensity of stochastic supply (noise trading)	σ_z^2	1
Common factor mean	$\bar{\mu}_q$	0.2
Common factor volatility	$\bar{\sigma}_q$	1
Common factor scaling factor	$(1/\lvert\mathbf{1}\rvert)\mathbf{1}$	$0.7 \times \mathbf{1}$
Correlation between factors	ρ_q	0
Correlation between established signals	ρ_{s0}	0.2
Correlation between new signals	ρ_{s1}	-0.5
Correlation inc. signals and factors (diagonal matrix)	$\mathrm{diag}(\boldsymbol{R}_{qs0})$	$(0.5\ 0.25)^\top$
Correlation new signals and factors (diagonal matrix)	$\mathrm{diag}(\boldsymbol{R}_{qs1})$	$(0.25\ 0.5)^\top$
Correlation inc. signals and new (independent)	$\boldsymbol{R}_{ss}$	$0\boldsymbol{I}$

Table 1.2: Baseline factor structure.

	Notation	Value
Factor mean outer product	$\boldsymbol{\mu}_q\boldsymbol{\mu}_q^\top$	$\begin{pmatrix} 0.02 & 0.02 \\ 0.02 & 0.02 \end{pmatrix}$
Factor covariance	$\boldsymbol{\Sigma}_q$	$\begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$
Cond. exp. of factors		
Est. signals covariance term	$\boldsymbol{\Sigma}_{\zeta 0}$	$\begin{pmatrix} 0.13 & -0.01 \\ -0.01 & 0.03 \end{pmatrix}$
Unscaled new signals cov. term	$\boldsymbol{S}$	$\begin{pmatrix} 0.04 & 0.04 \\ 0.04 & 0.17 \end{pmatrix}$
Second moments	$\boldsymbol{\Omega}_\zeta = \boldsymbol{\mu}_q\boldsymbol{\mu}_q^\top + \boldsymbol{\Sigma}_{\zeta 0} + k_S\boldsymbol{S}$	$\begin{pmatrix} 0.17 & 0.03 \\ 0.03 & 0.14 \end{pmatrix}$

Table 1.3: Baseline moments.

	Notation	Value $+\bar{\beta}$	Value $-\bar{\beta}$
Variance of pay-off	$Var[y]$	0.20	0.20
Cost of complexity	χ	0.02	0.02
Cond. var. of pay-off true model	$Var[y\|\boldsymbol{\beta}, \boldsymbol{s}_I]$	0.14	0.14
Mean squared error informed	$E[(y-\hat{y}_I)^2]$	0.15	0.15
Mean squared error uninformed	$E[(y-\hat{y}_U)^2]$	0.17	0.16
Var. informed predictor	$Var[\hat{y}_I]$	0.04	0.11
Expectation of inf. pred.	$E[\hat{y}_I]$	0.07	-0.15
Expectation of pay-off true	$E[y\|\boldsymbol{\beta}]$	0.11	-0.11
Var. inf. pred. true model	$Var[E[\hat{y}_I\|\boldsymbol{\beta}, \boldsymbol{s}_I]]$	0.06	0.06
Cov. inf. predictor and pay-off			
Contemporary measure	$Cov[y, \hat{y}_I\|\boldsymbol{c}^*]$	0.03	0.10
Objective measure	$Cov[y, \hat{y}_I]$	0.04	0.08

Table 1.4: Baseline market structure.

	Notation	Value $+\bar{\beta}$	Value $-\bar{\beta}$
Informed position scaling factor	ψ_I	6.6	6.6
Uninformed position scaling factor	ψ_U	6.0	6.3
Uninformed responsiveness to price-signal	λ_U	0.64	0.82
Price responsiveness to shocks	λ_p	0.82	0.91

Table 1.5: Predictive variables. In depth variable descriptions can be found in Welch and Goyal (2008).

Name	Notation
Dividend price ratio	dp
Earnings price ratio	ep
Stock variance	$svar$
Book to market value	bm
Corporate Issuing Activity	$ntis$
Treasury bills	tbl
Long term yield	lty
Default yield spread	dfy
Default return spread	dfr
Inflation	$infl$

Table 1.6: Parameters of the calibration (see Section 1.5.1) excluding predictive variables correlation parameters (see Figure 1.14).

Name	Notation	Value
Weight in full sample average expected coefficients	w	0.64
Common factor mean	$\bar{\mu}_q$	0.2
Common factor volatility	$\bar{\sigma}_q$	0.8
Difficulty of estimation problem	$k_{\sigma 0}$	0.3
Investors' estimation technology level period 1	k_{cI1}	-0.3
Investors' estimation technology level period 2	k_{cI2}	-1.0
Correlation between investors' signal group and factor 1	$\rho_{qs_I 1}$	0.3
Correlation between investors' signal group and factor 2	$\rho_{qs_I 2}$	-0.21

Figure 1.1: Bias ε_β and volatility σ_β. Panel 1.1(d) shows cost of complexity χ and the left hand side of the inequality of Proposition 3 as $\dagger = Var[y|\boldsymbol{\beta}, \boldsymbol{s}_I] + (\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\mu}_q)^2$.

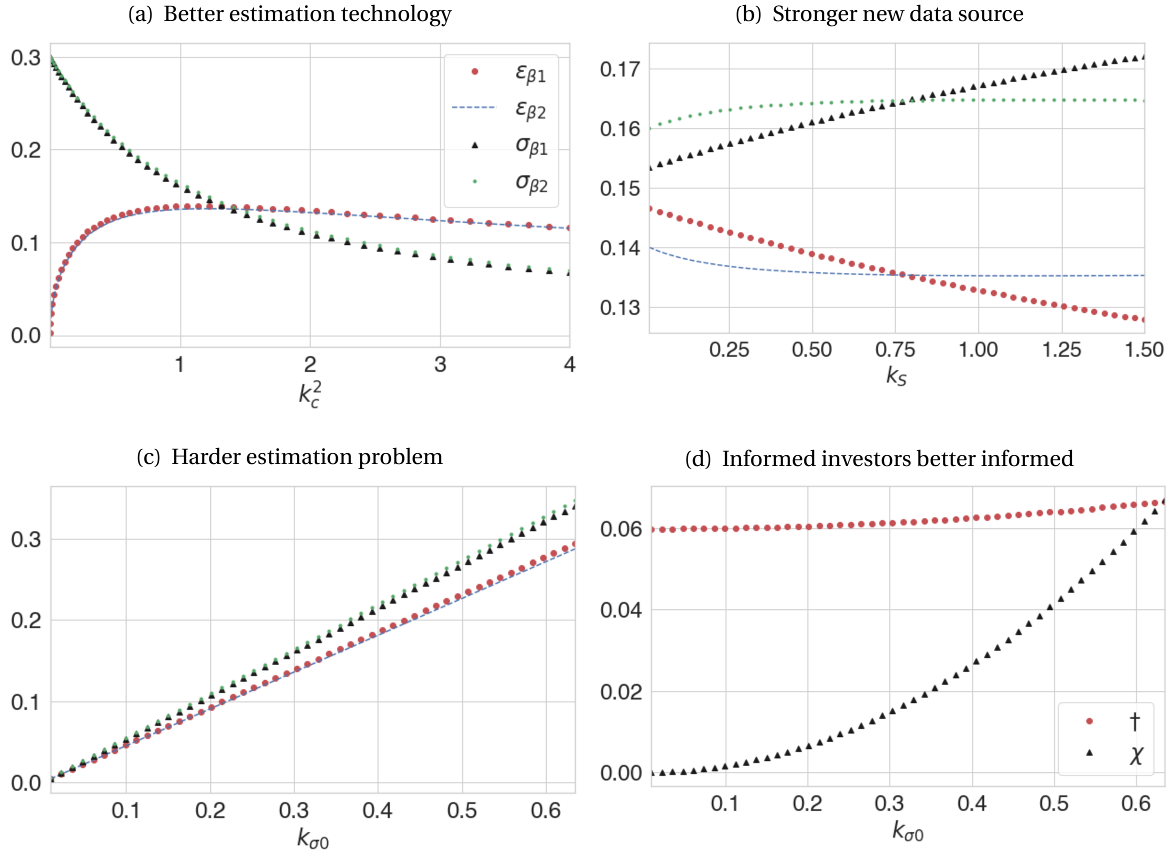


Figure 1.2: Mean squared error of the informed predictor $E[(y - \hat{y}_I)^2]$, cost of complexity χ, and conditional variance of the pay-off under the true model $Var[y|\boldsymbol{s}_I, \boldsymbol{\beta}]$. Stronger new source of data (higher k_S) can lead to higher or lower mean squared error. Given the baseline parametrization the two scenarios are captured by high or low estimation difficulty $k_{\sigma 0}$.

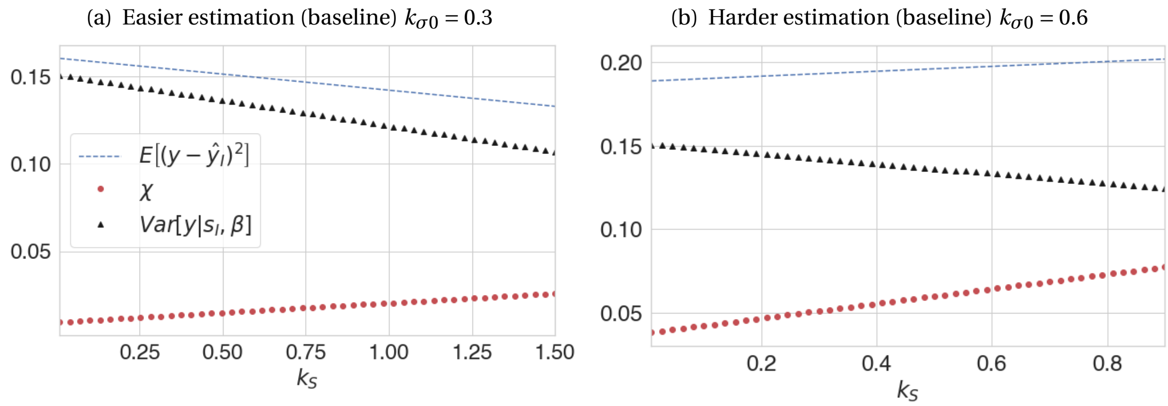


Figure 1.3: Easier estimation problem ($k_{\sigma 0} = 0.3$) means that informed investors uses all information. The microstructure fundamental of trading noise mainly shifts coefficient curves to larger absolute values. Price responsiveness λ_p is respectively ≈ 0.92 and ≈ 0.87.

(a) Less noise trading $\sigma_z^2 = 0.5$ (b) More noise trading $\sigma_z^2 = 1$

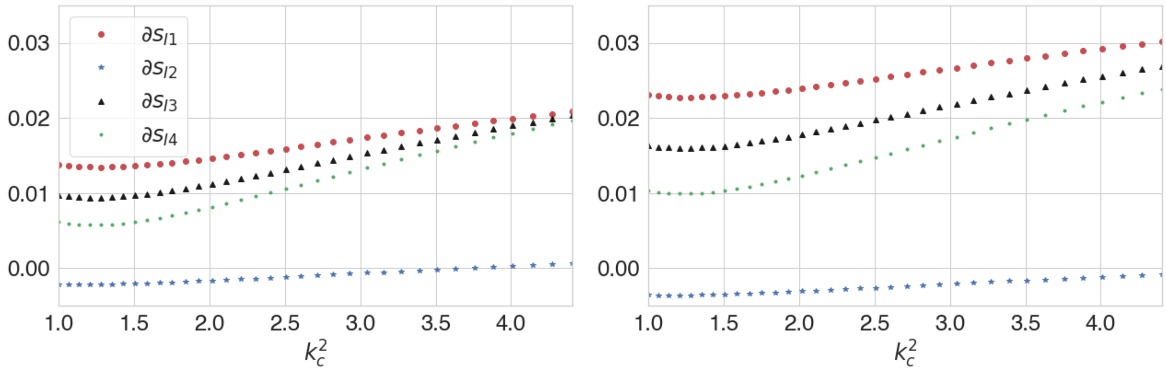
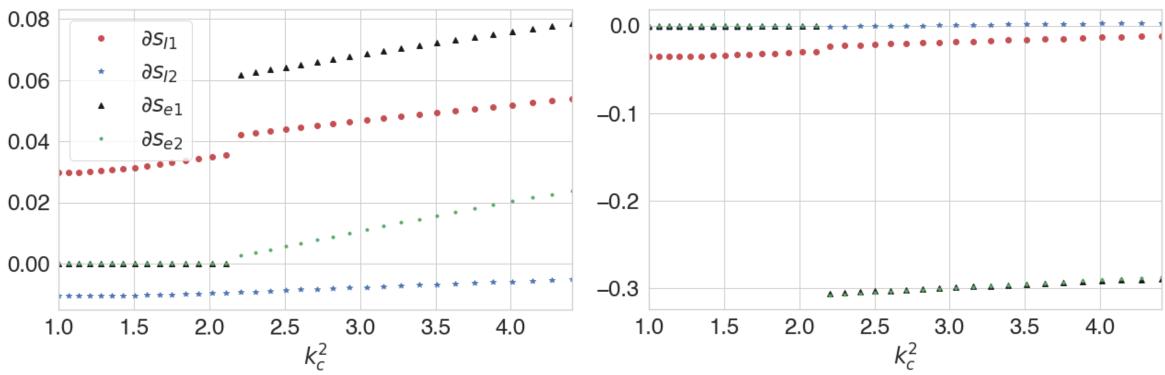


Figure 1.4: Harder estimation problem ($k_{\sigma 0} = 0.6$) means that the information set of informed investors and econometricians eventually differs. Noise trading is at the baseline level $\sigma_z^2 = 1$. Information differences demonstrates a notable break. Signals $\boldsymbol{s}_I$ and $\boldsymbol{s}_e$ are independent, but their coefficients are connected through econometricians' bias.

(a) Positive $\bar{\beta}$ (b) Negative $\bar{\beta}$



(c) Econometricians' bias

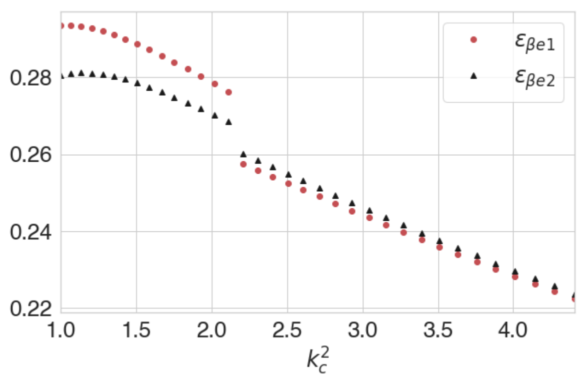


Figure 1.5: The bias in the constant term of the projection of price changes r onto signals is increasing in the gap between investors' and econometricians' bias as the latter a decreases toward zero (see Figure 1.4).

(a) Easier estimation $k_{\sigma 0} = 0.3$ (b) Harder estimation $k_{\sigma 0} = 0.6$

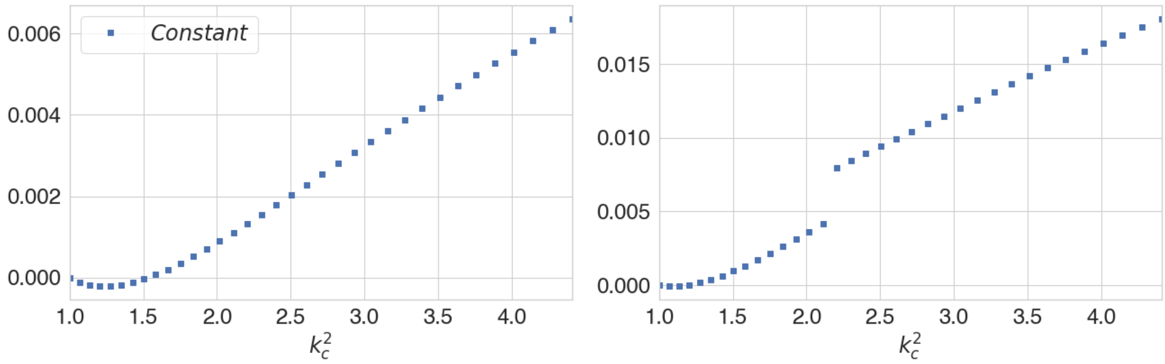


Figure 1.6: Mean squared error of prediction based on price contrasting optimal choice of estimator made by the informed investors (private) and a social planner optimizing for price informativeness. All plots under the hard estimation scenario (baseline) $k_{\sigma 0} = 0.6$. Easy estimation scenario is similar in most cases, except for the contrast between same and opposite sign true factor loadings, $+\bar{\beta}$ and $-\bar{\beta}$ respectively, when varying the new data parameter k_S, which is less pronounced with both graphs looking more like the plot in column 1.6(b).

(a) Positive $\bar{\beta}$ (b) Negative $\bar{\beta}$

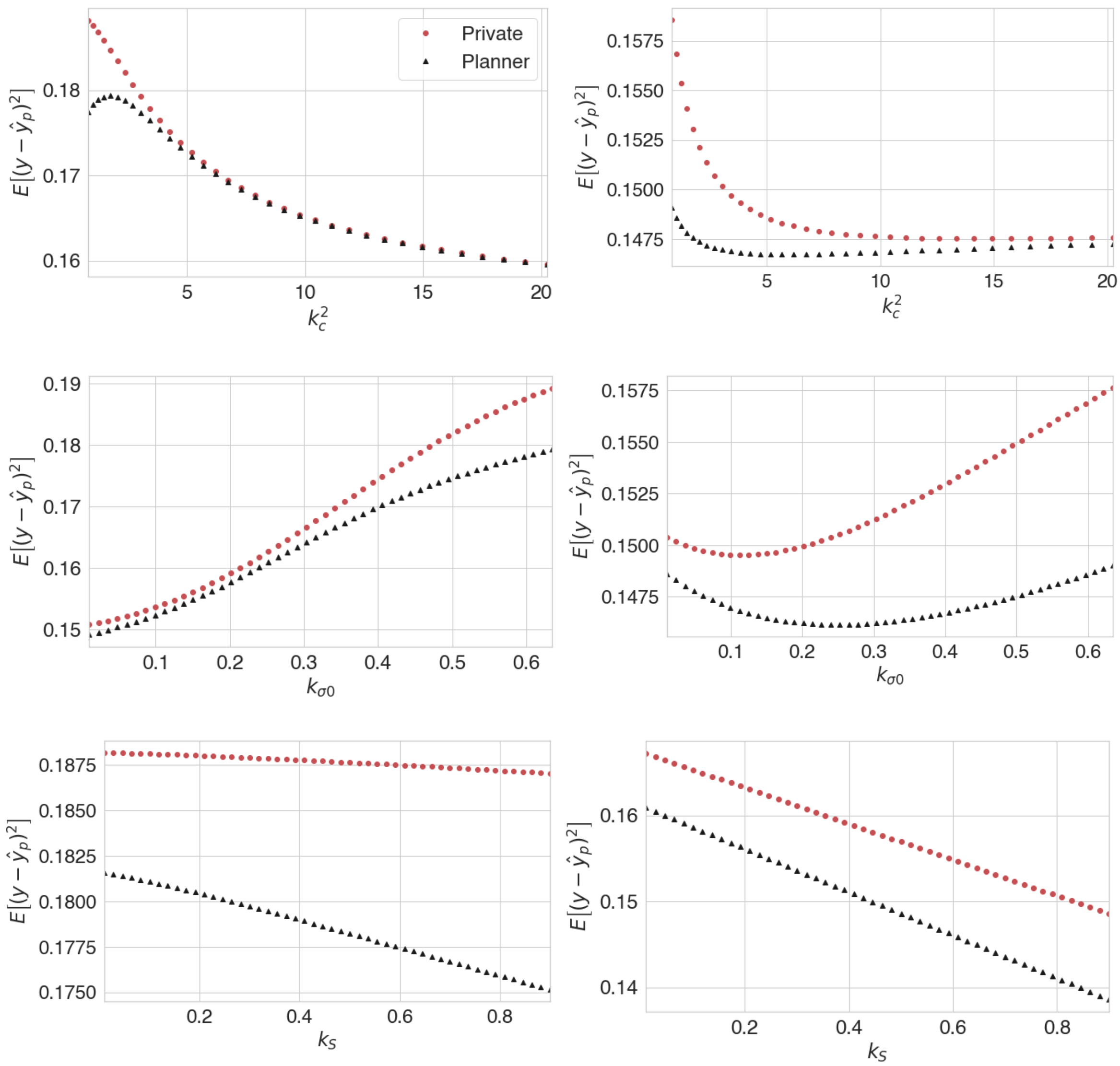
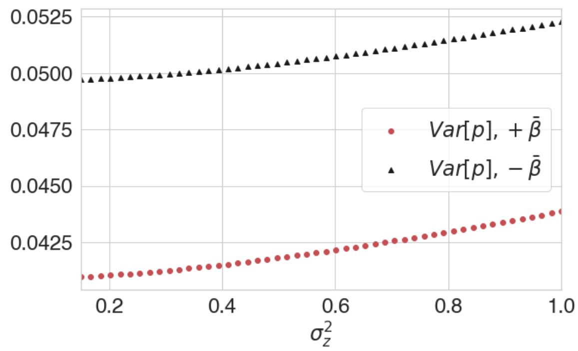
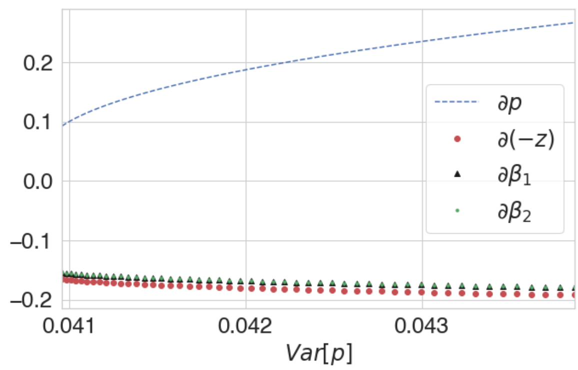


Figure 1.7: Variance of price and reversals coefficients with respect to trading noise σ_z^2, and estimation noise driven by estimation difficulty $k_{\sigma 0}$. Price variance increasing in trading noise also holds at lower levels of estimation technology parameter k_c^2, as well as for same sign bias and true factor loadings $(+\bar{\beta})$. Meanwhile price variance decreasing in estimation difficulty $k_{\sigma 0}$ is reversed at lower levels. Overall the tendency for opposite trends in coefficients is found at different levels of technological sophistication in parameter regions allowed by the constraint of Proposition 3 (informed predictions outperform).

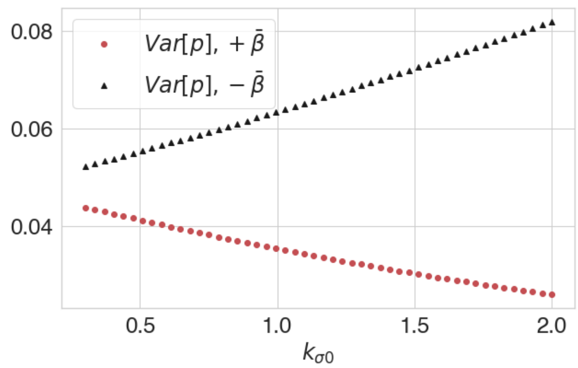
(a) Limiting case $k_c^2 = 100$, trading noise σ_z^2



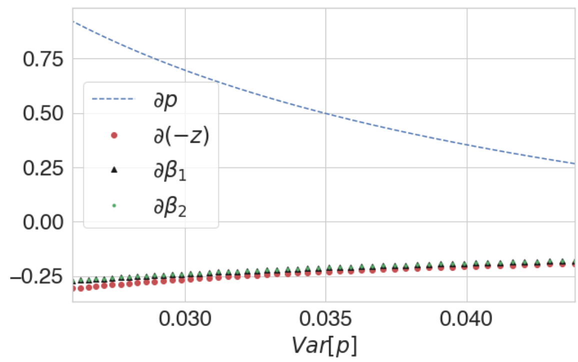
(b) Positive $\bar{\beta}$, trading noise σ_z^2



(c) Limiting case $k_c^2 = 100$, estimation noise $k_{\sigma 0}$



(d) Positive $\bar{\beta}$, estimation noise $k_{\sigma 0}$



(e) Negative $\bar{\beta}$, estimation noise $k_{\sigma 0}$

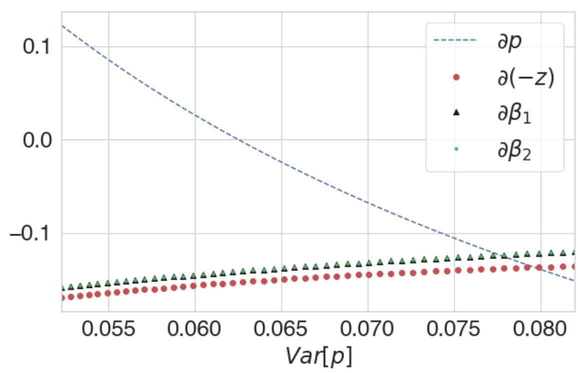


Figure 1.8: Variance and covariance of informed predictor and pay-off under contemporary measure c^* and objective measure. Both the shape of the curves as well as the ordering of the covariances is determined by the sign of the sign of true factor loadings parametrized by $\bar{\beta}$. The variation across estimation difficulty is included for comparability with Figure 1.9.

(a) Easier estimation $k_{\sigma 0} = 0.3$, negative $\bar{\beta}$ (b) Harder estimation $k_{\sigma 0} = 0.6$, positive $\bar{\beta}$

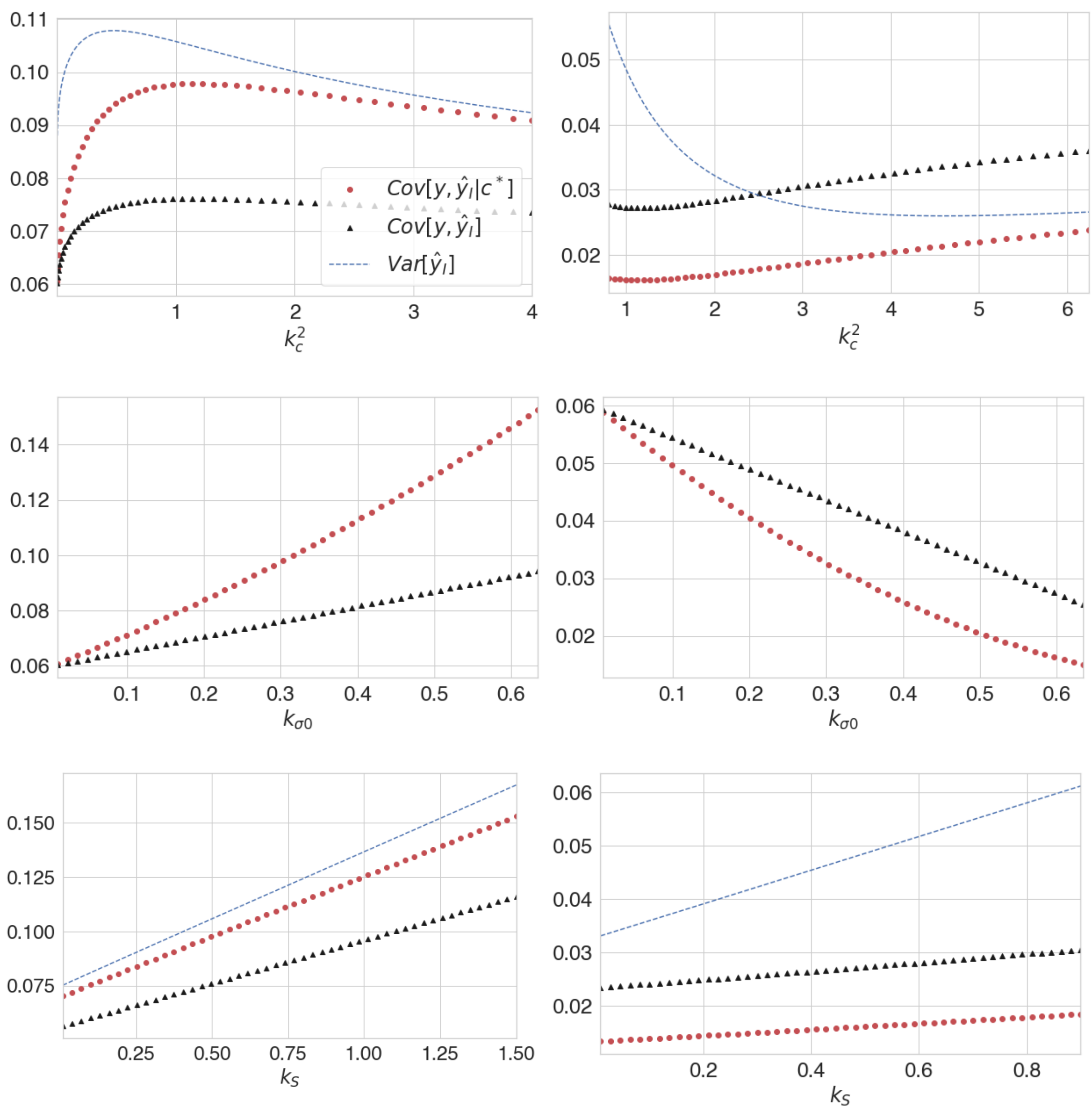


Figure 1.9: Fund performance under the contemporary measure and objective measure. The ordering of the corresponding covariances determine whether average expected profit or average realized profit is higher. The baseline difficulty influences the trend of the two other parameters, most notably new data k_S, which increases profits in the hard estimation scenario.

(a) Easier estimation $k_{\sigma 0} = 0.3$, negative $\bar{\beta}$ (b) Harder estimation $k_{\sigma 0} = 0.6$, positive $\bar{\beta}$

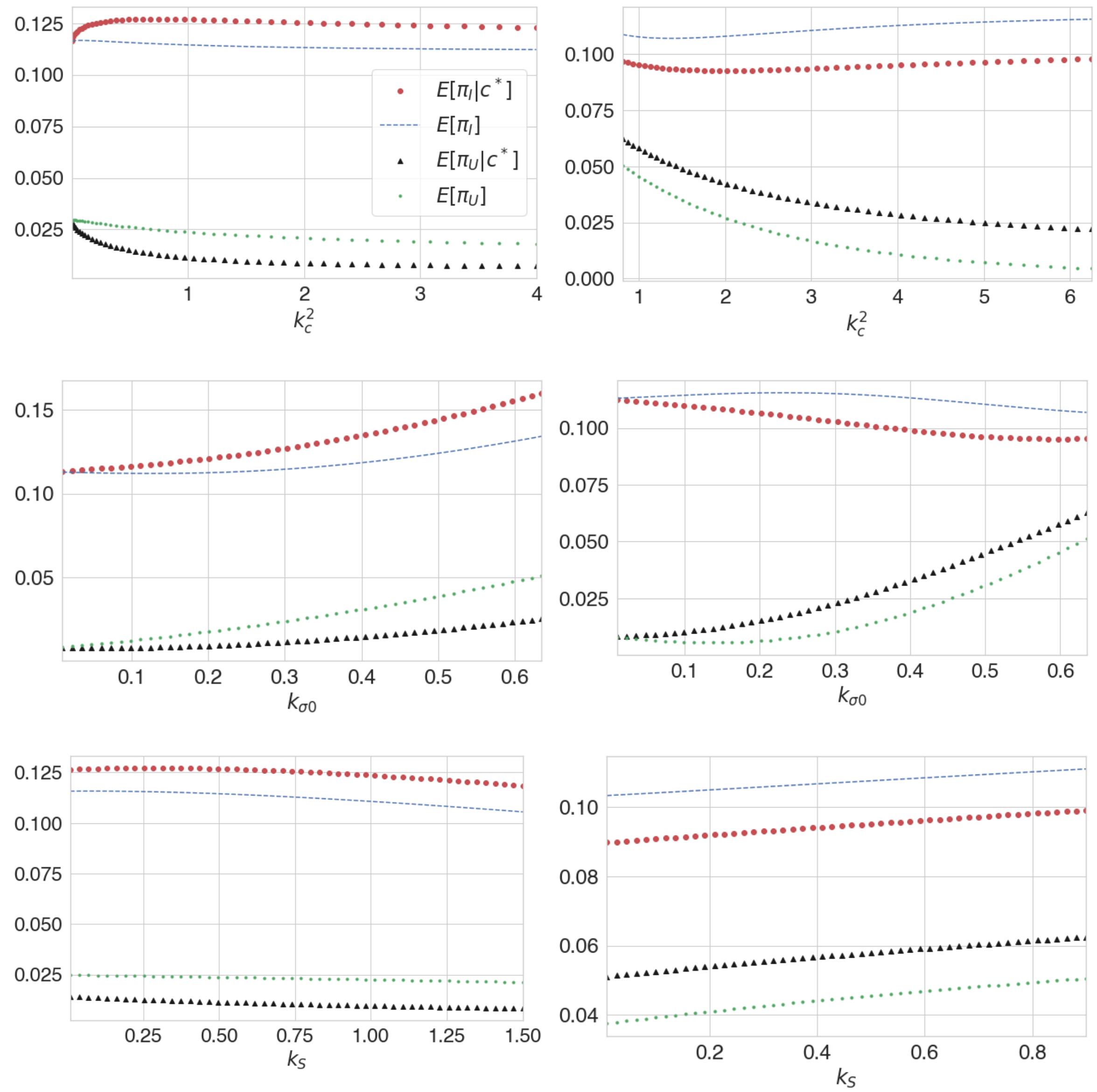
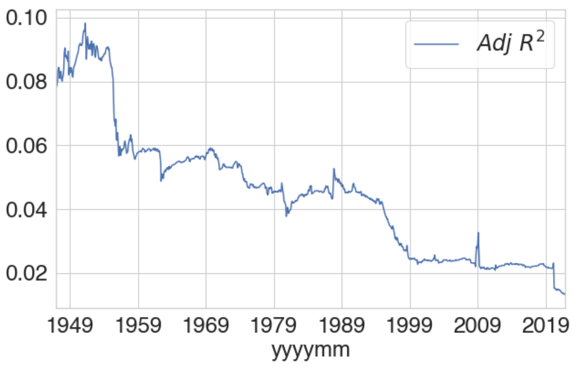


Figure 1.10: Expanding and contracting windows adjusted in-sample R-squared. $Adj\ R^2 = 1 + \frac{n+1}{n-p}(R^2 - 1)$ where n is number of observations and p is number of parameters.

(a) Expanding window



(b) Contracting window

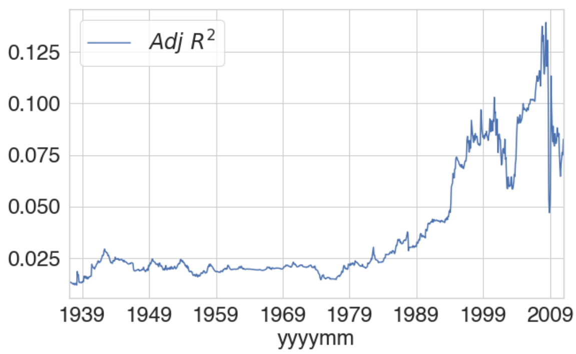


Figure 1.11: Rolling windows adjusted in-sample R-squared. $Adj\ R^2 = 1 + \frac{n+1}{n-p}(R^2 - 1)$ where n is number of observations and p is number of parameters. Each curve represent the adjusted R-squared of a model estimated over the number of years shown in the legend. Plots are smoothened by plotting 10 years rolling averages of the adjusted R-squared.

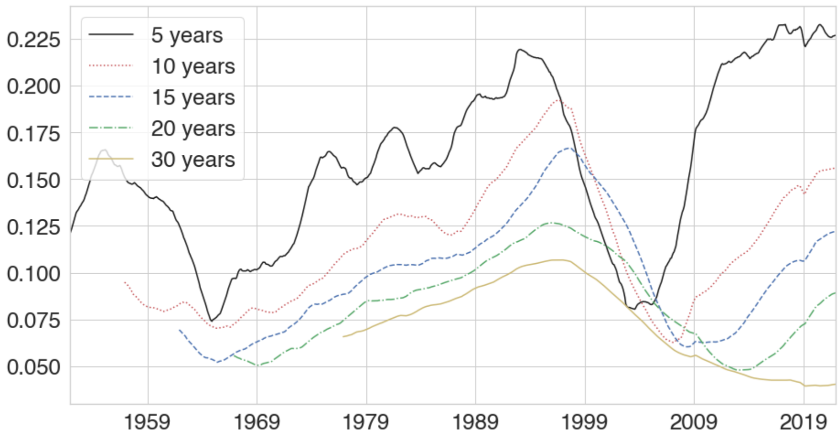
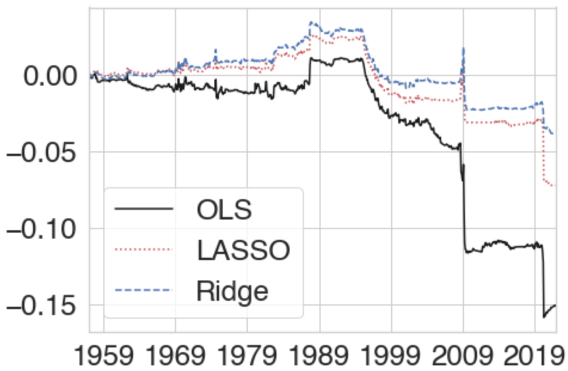
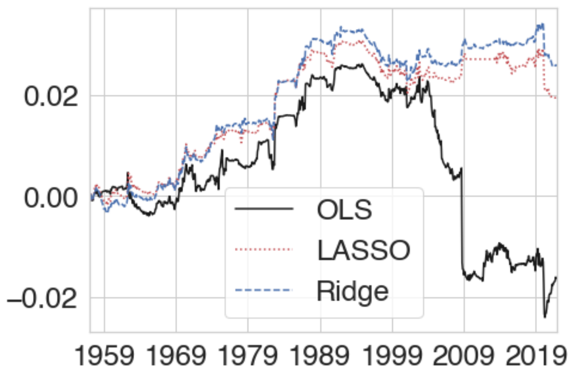


Figure 1.12: Predictive performance compared to historical mean of different methods over rolling windows of 20 and 30 years and with and without the Campbell and Thompson (2008) zero floor that floors predictions at zero denoted CT. Except for 30 years with CT, OLS is a clear under-performer. All plots present deteriorating performance in the second sub-period (starting around 1991) and all expect 20 years OLS without CT present out-performance in the first.

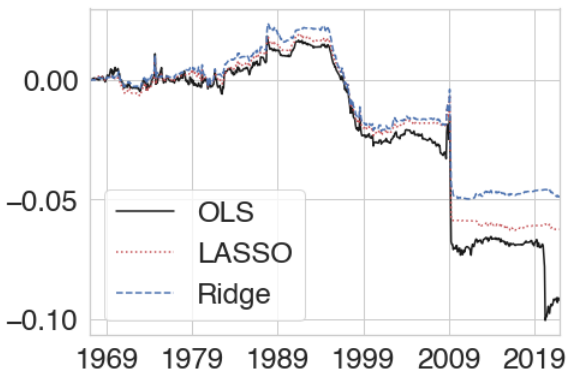
(a) 20 years, no CT



(b) 20 years, CT



(c) 30 years, no CT



(d) 30 years, CT

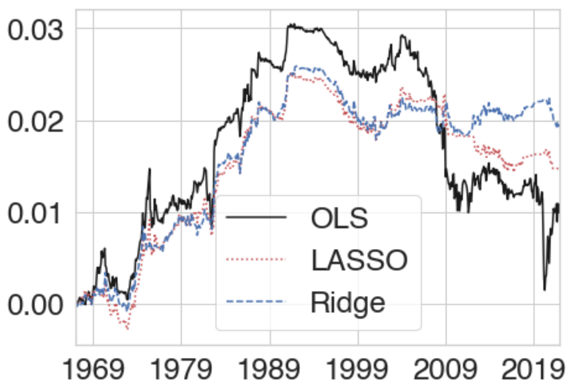
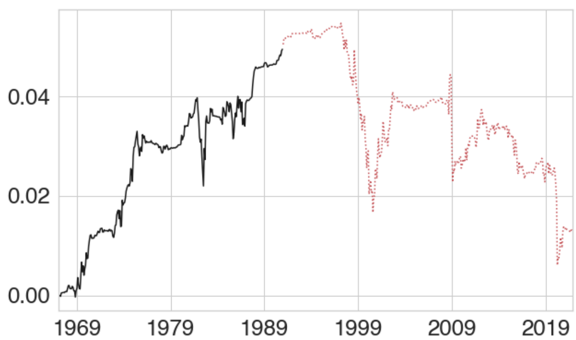


Figure 1.13: Predictive performance compared to historical mean of expected coefficient targets for estimation.

(a) Expected coefficients of sub-periods



(b) Expected coefficients full period

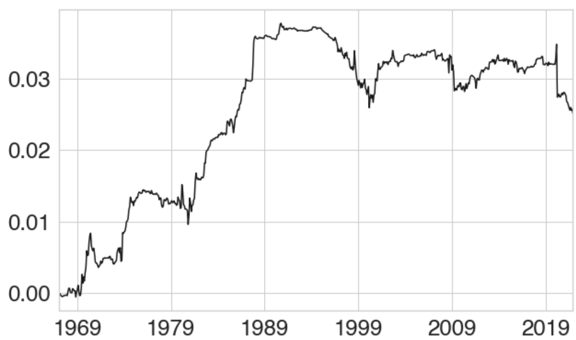
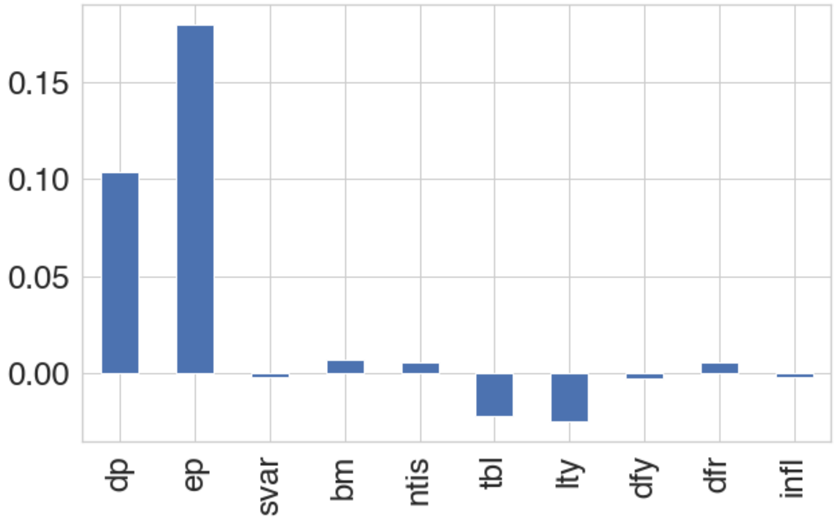


Figure 1.14: Predictive variable correlation parameters estimated in calibration. Description of variables in Table 1.5.



Dividend Restrictions and Asymmetric Information

with Suzanne Vissers (EPFL, Swiss Finance Institute)

We develop a dynamic model of a bank whose management has superior information about the impact of a pending shock to the bank's cash holdings and can signal the bank's type through its dividend policy. Banks that will be adversely affected by the shock have incentives to pool with unaffected banks to increase their market value. To avoid being mimicked, the unaffected banks can credibly *signal via a more aggressive payout strategy. Dividend payout restrictions have the potential to prevent a separating equilibrium from forming. This leads to the bad type adopting a more aggressive payout policy with a higher risk of default but mitigates the distortion of the good type's policy. We identify a number of scenarios where this trade-off presents an opportunity for regulatory intervention and somewhere it does not.*

2.1 Introduction

Payout restrictions have become an increasingly important part of the macro-prudential toolbox of central banks. Examples are the countercyclical capital buffer and the capital conservation buffer introduced in Basel III, which, when triggered, restrain banks from paying out dividends and buying back shares.[1] More recently, both the Federal Reserve and the European Central Bank responded to the Covid-19 outbreak by imposing strict limitations on banks' distributions to shareholders, leading to a decline of 57% of aggregate dividends paid out in 2020 compared to the year before, see Hardy (2021).[2] In the presence of informational frictions generated by the opaque and complex nature of the banking sector, we argue that these measures might have unintended consequences and ultimately be counterproductive in achieving their goal of improving banks' resilience to crises. However, we also show that regulation can, under some circumstances, mitigate the distortions of the laissez-faire equilibrium.

Due to their central role in the financial system, bank defaults can have significant adverse spill-over effects on the real economy, see Acharya, Cooley, Richardson, and Walter (2009). Therefore, regulators

[1] The capital conservation buffer requires banks to hold 2.5% of common Tier equity capital on top of the minimum capital requirement.

[2] In June 2020, the Fed barred banks from share buybacks and capped dividend payments to the amount paid in the second quarter of 2020 and further limited to an amount based on recent earnings, see Federal Reserve System (2020). On March 2021, it was announced that this measure ends for most banks on June 30, 2021, see Federal Reserve System (2021). In May 2020, the European Central Bank asked member banks to refrain from dividend payments completely. This recommendation was addressed to significant institutions (SIs) that are under the direct supervision of the ECB and to national competent authorities (NCAs) that supervise less significant institutions (LSIs). See European Systemic Risk Board (2020b) for an overview of the regulatory announcement made by the NCAs. The recommendation was later revised to a limit on dividend payments and is in place until at least September 2021, see European Central Bank (2020).

aim to limit the likelihood and mitigate the impact of these events. Since the banking industry is among the industries with the highest payout ratios, see Guntay, Jacewitz, and Pogach (2015), one way for the regulator to achieve this is by requiring banks to build sufficient buffers before distributing funds to investors. However, the high level of dividends observed in the banking sector suggests that dividends are essential for its investors. The explanation that we consider is that banks use dividends as a signaling device, e.g., Boldin and Leggett (1995), a strategic behavior that complicates regulatory intervention. Specifically, banks use dividends to signal their future financial health rather than future earnings. This is the hypothesis better supported by the empirical literature, see Tripathy, Wu, and Zheng (2021) for an overview as well as additional evidence.

This paper studies how information asymmetry affects banks' optimal dividend payout policy and what, in relation to that, the consequences of dividend restrictions are. We develop a dynamic model of a bank that controls its cash reserves by paying out dividends. The bank's management has superior information about the impact of a pending shock to the bank's cash reserves. For simplicity, we assume that the good type is unaffected by the liquidity shock, whereas the bad bank loses a fixed amount of its cash reserves. The liquidity shock can be interpreted in several ways, e.g., a large trading loss[3], a regulatory fine[4], or a margin call[5]. The bank's type is private information of bank management and can therefore not be observed by the regulator and potential outside investors. We assume that when the liquidity crisis hits, the market learns about the bank's type. Focusing on a liquidity shock rather than different risk levels of unlike asset compositions allows us to isolate the informational effect from the bank's choice of asset portfolio, a topic that is well-studied in the literature (Ross, 1998; Leland, 1998; Vissers, 2021).

In the tradition of the literature on dividend signaling starting with Miller and Rock (1985), an incentive to signal follows from the assumption that the bank acts in the joint interest of long-term shareholders and short-term shareholders. Whereas the long-term shareholders care about the long-term intrinsic value of the bank, the short-term shareholders are concerned with the market valuation, as they want to be able to sell their stocks at any point in time. A micro-foundation for such short-termism could be that investors themselves are exposed to liquidity shocks or that they face a stochastic investment set leading to a certain probability that re-balancing is necessary. Alternatively, focusing directly on management, this assumption could reflect that management's remuneration scheme is tied to stock price performance.

Under full information and appropriate parameter settings, the good bank pays out dividends at a lower cash level than the bad bank considering the latter wants to hold an additional precautionary savings buffer to withstand the liquidity shock. In the presence of asymmetric information, the bad type has an incentive to mimic the good type to boost its market valuation. When the market cannot distinguish the two bank types, both bank types' market valuation will be a weighted average of its correct valuations. Since this is disadvantageous for the good bank, it has incentives to signal its type by distorting its dividend policy and thereby imposing mimicking costs on the bad bank. By trading off the costs of increased default risk versus the benefit of higher market valuation, the good bank might forgo adopting an aggressive separating strategy and accept being pooled with the bad bank instead. We establish the conditions for the existence of the separating and pooling equilibria.

The model generates several novel insights. First of all, the model predicts that in the separating equilibrium, a good bank pays out dividends more aggressively compared to the symmetric information

[3]For example, the collapse of Archegos Capital Management brought severe trading losses to a group of large banks, see The Economist (2017).

[4]The Economist (2021b) reports that global banks were hit with \$10.4 billion in regulatory fines for money-laundering activities, an increase of more than 80% compared to 2019.

[5]The European Systemic Risk Board (2020a) report considers the implications of large margin calls from cash and derivatives positions on bank and non-bank entities. According to the report, some banks have experienced a significant increase in initial margins and, as a result, increased liquidity constraints in terms of liquid assets and available collateral.

benchmark, whereas a bad bank adopts its first-best strategy. This strategic behavior can result in the good bank having a higher risk of default than the bad bank, despite not being exposed to the liquidity shock. Conversely, a good bank pays dividends at a higher cash level and a bad bank at a lower cash level than their respective first-best targets in the pooling equilibrium. As a result, the bad bank becomes more prone to default and the good bank less than in the symmetric information benchmark.

Second, we study the implications of dividend restrictions imposed by the regulator. We show that a dividend restriction *before* arrival of the liquidity shock has the potential to break the separating equilibrium. When the dividend restriction is set sufficiently high, the bank moves to a pooling equilibrium. This change leads to an increased (decreased) target cash level of the good (bad) bank and the effects on default risk described in the previous paragraph. On an industry level, the regulator faces a trade-off between reduced default risk of the good bank and increased default risk of the bad bank.

Third, considering the additional dimension of industry value, we find that there are up to three regions for the impact of regulation. In the first region, regulation is so lax that it does not affect the equilibrium outcome. In the third region, regulation is so tight that both bank types follow a more conservative dividend policy than their first-best. In this region, regulation in the form of dividend restrictions makes the industry safer, but at the same time, reduces the average value of banks. From a regulatory perspective, the most interesting is the intermediate region, which exists when the unregulated equilibrium is of the separating type. Breaking this separating equilibrium has the potential to *both* lower the average default risk and increase the average bank value. However, depending on the characteristics of the banking industry and the liquidity shock, the outcome might turn out the opposite way, i.e., raising the average default risk *and* destroying value.

We identify two opposing channels through which the *scope* of the liquidity shock (i.e., the fraction of banks exposed to the liquidity shock) affects the impact of regulation and explain the previous result. The *direct* channel captures the fact that, other things equal, the impact of regulation on banks exposed to the shock weighs heavily on aggregate outcomes when they constitute a significant fraction of the industry. The *indirect* channel arises from how banks strategically adapt to the scope of the liquidity shock. In moving from a separating to a pooling equilibrium, the bad bank's dividend policy distortion is more substantial when only a few banks are exposed to the shock, and vice versa. Through scenario analysis, we establish how the *size* of the shock determines which channel dominates. For a large and concentrated shock, regulation tends to be beneficial by simultaneously lowering the bank's average default risk and increasing industry value. For a small and widespread shock, the opposite outcome materializes, making regulation harmful. In both scenarios, the indirect channel dominates. For a shock of medium-sized but concentrated shock, the direct channel dominates, leading to opposite conclusions. In the scenario of a small concentrated shock, banks tend to already be in a pooling equilibrium.

Next to the economic fundamentals of the shock, i.e., the scope and size, the degree to which investors put weight on the bank's (short-term) market valuation influences the effectiveness of regulation. We argue that a high degree of short-term focus improves the outlook for regulation when it can prevent aggressive distortion of the good type's payout policy.

Throughout this paper, analyzed firms are considered to be banks. This is in line with the characteristics of the banking industry as described above, i.e., regulatory scrutiny, high degree of complexity, exposure to tail risk, and high payout ratios. However, as we abstract from certain institutional features on the liability side to keep the analysis tractable, the model can be applied to non-financial firms too.[6]

[6]For example, parts of the aviation sector were subject to dividend restrictions after being bailed out in response to the Covid-19 outbreak; German airliner Lufthansa agreed to not pay out dividends of 2019 in exchange for a 10 billion euro

This paper relates to different strands of the literature. The idea that firms use dividends to signal their quality was suggested by Miller and Modigliani (1961) and later theoretically formalized by Bhattacharya (1979), Miller and Rock (1985), and Kale and Noe (1990).[7] More recently, Guntay et al. (2015) studied the informational role that regulators have in the banking industry with a model in which a regulator with information superior to the market has to approve dividend payments by banks. A dividend restriction that restricts only the weakest banks in the economy might trigger a bank run, whereas the opaqueness resulting from strict dividend restrictions promotes bank stability. In contrast, we find that a pooling equilibrium resulting from dividend restrictions *can* be less stable than a laissez-faire separating equilibrium. Additionally, we pose the problem in a continuous-time liquidity management framework, as opposed to static models of the papers mentioned above, see Moreno–Bromberg and Rochet (2018).

Acharya, Gujral, Kulkarni, and Shin (2011) provide an overview of the dividends paid out by the largest banks before and during the crisis period of 2007-2009. They demonstrate that banks had been paying out significant dividends during the crisis period despite widely anticipated credit losses. The authors attribute this behavior to the short-term nature of the banks' funding and the implicit and explicit government guarantees. As banks are funded with short-term debt, a dividend cut could trigger a market debt run. We take a different approach by abstracting from the bank's liability side and focus on dividends as a signaling device for exposure to adverse shocks.

In the wake of the financial crisis, Acharya et al. (2011) and also Admati, DeMarzo, Hellwig, and Pfleiderer (2013) advocate payout restrictions to promote the stability of the financial industry. Furthermore, Goodhart, Peiris, Tsomocos, and Vardoulakis (2010) and Acharya, Le, and Shin (2017) argue that payout dividend restrictions are desirable when banks' balance sheets are intertwined. Muñoz (2019) adopts a DSGE modeling approach and concludes that bank dividend prudential targets induce welfare gains associated with Basel III-type capital regulation. We contribute to this discussion by adding nuances to the trade-offs involved in imposing such policies due to informational asymmetries.

The paper is organized as follows. Section 2.2 presents the model and the bank's valuation in the symmetric information case. Section 2.3 studies the effects of asymmetric information on the bank's dividend policy. Section 2.4 considers the implications of dividend restrictions on the bank's dividend strategy, valuation, and default likelihood for different parametric scenarios. Section 2.5 concludes.

2.2 The model

2.2.1 Set-up

We develop a continuous-time model of a bank that is owned by shareholders who have limited liability and is run by management. Agents are risk-neutral and discount the future at a rate $\rho > 0$. The bank's assets consists of liquid reserves and a fixed volume of assets in place than generate cash flows X_t with dynamics:

$$dX_t = \overline{\mu} dt + \sigma dZ_t, \quad X_0 = 0.$$

Here, $Z = \{Z_t, t \geq 0\}$ is a Brownian Motion, representing small and continuous movements in the cash flows. The drift and volatility parameters $\overline{\mu}$ and σ are positive and known constants. It is assumed that the bank is partly financed by debt that is already in place and for which the bank pays coupon

bail-out, see Reuters (2021).

[7] Following Myers and Majluf (1984), investment decisions can work as an alternative signaling device, see also Morellec and Schürhoff (2011).

payment $c < \overline{\mu}$ per time period dt. The resulting cumulative earnings C_t evolve according to:

$$dC_t = (\overline{\mu} - c)dt + \sigma dZ_t = \mu dt + \sigma dZ_t, \quad C_0 = 0. \tag{2.1}$$

Here, μ denotes the drift of the bank's cash flows after interest payments to debt holders. Throughout the paper, we abstract from debt financing decisions and focus on the bank's equity valuation.

The bank's type can either be good or bad, denoted by $\ell \in \{G, B\}$. The fraction of good banks in the economy is $\alpha \in (0,1)$, and the complementary fraction $1-\alpha$ is of the bad type. We assume that the bad banks in the economy are subject to a liquidity shock that hits all bad banks at the same time. One could think of a large and widespread trading loss, regulatory fine, or margin call.[8] When hit by a liquidity shock, the bad banks in the economy all suffer a shock $f > 0$ to their liquid reserves. Define the time when the shock takes place as

$$\tau^* = \inf\{t > 0 : N_t = 1\},$$

where $N = \{N_t, t \geq 0\}$ is a Poisson process with intensity λ. For simplicity, we assume that the bad bank is only subject to a single shock. Furthermore, we assume that management (the insiders) knows the bank's type, which the regulator and investors only learn when the shock hits or when the bank credibly signals its type. We find it most natural to think that this applies to investors regardless of whether they have invested in the bank or not, but it is sufficient for our results that investors who want to sell shares in the secondary market cannot credibly signal the type of the bank to potential buyers in the secondary market. Note that the common earnings dynamics as described in Eq. (2.1) and the occurrence of a single shock ensure that outsiders cannot learn the bank's type by observing the liquid reserves before the arrival of the shock.

To add some interpretation to the difference between bank types, one could think of the average earnings parameter μ as the rate banks have to achieve in order to be competitive. One interpretation is that bad banks have to accept some tail risk to get to this level, whereas good banks have projects that do not require the additional tail risk to get to cash flow drift μ. An alternative interpretation is that banks have assets in place and that management learns at time $t = 0$ whether the bank's assets are exposed to the shock or not. As the weights of good and bad banks in the economy are exogenous in our model, the two interpretations are equivalent. However, extending the model to endogenize the choice of type would be more relevant under the former than the latter.[9]

The bank controls its cash reserves by paying out dividends. Let L_t be the cumulative dividends paid out over $[0, t]$. We assume that dividend process $L = \{L_t, t \geq 0\}$ is non-decreasing, $(\mathscr{F}_t)$-adapted and càdlàg. For simplicity, we assume that the reserves are not renumerated. The dynamics of the bank's cash reserves given a strategy L are:

$$M^L_{\ell,t} = m + C_t - L_t - \mathbb{1}_{\{t \leq \tau^*, \ell = B\}} f N_t,$$

where m denotes the initial level of liquid reserves and the last term with zero-one indicator function $\mathbb{1}_{\{\cdot\}}$ reflects the negative liquidity shock that applies to the bad bank. We assume that outsiders can observe the bank's cash reserves, so that the bank's type is revealed when the shock hits, and that the bank has to announce a dividend policy ex-ante.[10]

In order to introduce the risk of liquidation in the model, it is assumed that primary capital markets

[8]We discuss these cases more in detail in Section 2.4.3.

[9]For a sketch of what endogenizing this choice might mean for regulation, see Section 2.4.6.

[10]To focus on dividends, the signaling space is one-dimensional - signaling through other sources, such as regulatory capital is not possible.

are closed.[11] As a result, the bank defaults when its liquid reserves are fully depleted.[12] Let τ_ℓ^L be the liquidation time of bank type ℓ defined for a strategy L:

$$\tau_\ell^L = \inf\{t > 0 : M_{\ell,t}^L \leq 0\}.$$

To generate a signaling incentive, the model relies on managerial short-term incentives. Management acts in the joint interest of short- and long-term shareholders. Short-term shareholders care about the bank's current market value as they might sell their stocks on the secondary market sooner rather than later. Long-term shareholders only care about the intrinsic value of the bank. Let $k \in (0,1)$ denote the fraction of short-term shareholders, and the complementary fraction $1-k$ long-term shareholders. The short-term focus could reflect that investors face liquidity concerns or stochastic investment opportunity sets or that management pay is tied to stock price performance. As such, k can be interpreted more generally as capturing the relative importance of the market value (perceived value) $V_{\tilde{\ell}}^L(m)$ and the intrinsic value $V_\ell^L(m)$ under a strategy L. Therefore, the bank's objective function is given by the weighted sum of the bank's shareholder value as perceived by the market and the intrinsic bank's shareholder value.

$$V_{\ell,\tilde{\ell}}(m) := \max_{L \in \mathscr{A}} \quad k \underbrace{V_{\tilde{\ell}}^L(m)}_{\substack{\text{market value}\\ \text{(perceived value)}}} + (1-k) \underbrace{V_\ell^L(m)}_{\text{intrinsic value}} , \tag{2.2}$$

where $\mathscr{A}$ is the set of all admissible strategies, and

$$V_\ell^L(m) = \mathbb{E}\left[\int_0^{\tau_\ell^L} e^{-\rho t} dL_t | M_{\ell,0}^L = m\right], \quad V_{\tilde{\ell}}^L(m) = \mathbb{E}\left[\int_0^{\tau_{\tilde{\ell}}^L} e^{-\rho t} dL_t | M_{\tilde{\ell},0}^L = m\right],$$

are the present value of future dividends to shareholders for a bank of type ℓ and $\tilde{\ell}$, respectively.

2.2.2 Value function

To determine the bank's shareholder value, we first solve the value function $W(m)$ after the shock has hit. This value function satisfies the following Hamilton-Jacobi-Bellman (HJB) equation for all $m \geq 0$:

$$\max\left\{\frac{1}{2}\sigma^2 W''(m) + \mu W'(m) - \rho W(m), 1 - W'(m)\right\} = 0,$$

together with the boundary condition $W(0) = 0$. Once this function has been established, one can determine the value function defined in Eq. (2.2), which satisfies for all $m \geq 0$:

$$\max\left\{\frac{1}{2}\sigma^2 V_{\ell,\tilde{\ell}}''(m) + \mu V_{\ell,\tilde{\ell}}'(m) - (\rho+\lambda) V_{\ell,\tilde{\ell}}(m), 1 - V_{\ell,\tilde{\ell}}'(m)\right\} = 0,$$

with boundary condition $V_{\ell,\tilde{\ell}}(0) = 0$. The optimal payout strategy is of the so-called barrier type. This strategy is characterized by an optimal target level of liquid reserves $m_{\ell,\tilde{\ell}}$ (or equivalently, an optimal payout strategy), such that all liquid reserves beyond this point are distributed as dividends, and nothing is paid out below this point.

The bank's value function is a weighted sum of its market valuation and its intrinsic valuation. This following proposition presents the intrinsic value function $V_\ell(m; m_\ell)$ of a bank of type $\ell \in \{G, B\}$.[13]

[11] This restriction is equivalent to assuming that the cost of raising new equity is high enough to make it an unattractive alternative to default, see Chapter 2 in Moreno–Bromberg and Rochet (2018).

[12] One could easily generalize this to a set-up with a liquidity requirement, i.e., a strictly positive level of liquid reserves imposed by the regulator similar to Milne and Whalley (2005).

[13] Note that the notation V_ℓ with the single subscript ℓ represents the intrinsic value of a bank of type ℓ. In the presence of

We first determine the value function after the liquidity shock has hit. Note that it is assumed that when the shock occurs, the types are learned. For that reason, the value function after the shock $W(m)$ does not depend on type ℓ, and the corresponding dividend strategy is set optimally, i.e., absent any signaling considerations.

Proposition 6. *The value of a bank of any type $\ell \in \{G, B\}$ after the shock is given by*

$$W(m) = \begin{cases} \sum_{i=1}^{2} \overline{A}_i e^{\overline{r}_i (m - \overline{m}^*)}, & \text{for } m \in [0, \overline{m}^*), \\ m - \overline{m}^* + \sum_{i=1}^{2} \overline{A}_i, & \text{for } m \geq \overline{m}^*, \end{cases}$$

where $\overline{m}^*$ is the optimal cash target after the shock. Let m_ℓ be the target cash level of bank type ℓ. We distinguish two cases for the intrinsic value of the good bank before the arrival of the shock:
(i) For $m_G < \overline{m}^*$:

$$V_G(m) = V_G(m; m_G) = \begin{cases} \sum_{i=1}^{2} A_i^G e^{r_i (m - m_G)} + W(m), & \text{for } m \in [0, m_G), \\ \sum_{i=1}^{2} A_i^G + W(m_G) + m - m_G, & \text{for } m \geq m_G. \end{cases}$$

(ii) For $m_G > \overline{m}^*$:

$$V_G(m) = V_G(m; m_G) = \begin{cases} \sum_{i=1}^{2} A_i^G e^{r_i (m - m_G)} + W(m), & \text{for } m \in [0, \overline{m}^*), \\ \sum_{i=1}^{2} B_i^G e^{r_i (m - m_G)} + \beta_G + \gamma m, & \text{for } m \in [\overline{m}^*, m_G), \\ \sum_{i=1}^{2} B_i^G + \beta_G + (\gamma - 1) m_G + m, & \text{for } m \geq m_G. \end{cases}$$

We distinguish three cases for the value of the bad bank before the shock arrives:
(i) For $0 < m_B - f < \overline{m}^*$:

$$V_B(m) = V_B(m; m_B) = \begin{cases} \sum_{i=1}^{2} A_i^B e^{r_i (m - m_B)}, & \text{for } m < f, \\ \sum_{i=1}^{2} B_i^B e^{r_i (m - m_B)} + W(m - f), & \text{for } m \in [f, m_B], \\ \sum_{i=1}^{2} B_i^B + W(m_B - f) + m - m_B, & \text{for } m > m_B. \end{cases}$$

(ii) For $m_B - f \geq \overline{m}^*$:

$$V_B(m) = V_B(m; m_B) = \begin{cases} \sum_{i=1}^{2} A_i^B e^{r_i (m - m_B)}, & \text{for } m \in [0, f), \\ \sum_{i=1}^{2} B_i^B e^{r_i (m - m_B)} + W(m - f), & \text{for } m \in [f, f + \overline{m}^*], \\ \sum_{i=1}^{2} C_i^B e^{r_i (m - m_B)} + \beta_B + \gamma m, & \text{for } m \in [f + \overline{m}^*, m_B), \\ \sum_{i=1}^{2} C_i^B + \beta_B + \gamma m_B + m - m_B, & \text{for } m \geq m_B. \end{cases}$$

(iii) For $m_B - f \leq 0$:

$$V_B(m) = V_B(m; m_B) = \begin{cases} \sum_{i=1}^{2} A_i^B e^{r_i (m - m_B)}, & \text{for } m < m_B, \\ \sum_{i=1}^{2} A_i^B + m - m_B, & \text{for } m \geq m_B. \end{cases}$$

Proof. The derivations, including the expressions of all the coefficients $\bar{A}_i$, A_i^ℓ, B_i^ℓ, C_i^B, after-shock dividend target $\overline{m}^*$, the constants β_G, β_B, and γ, and the characteristic roots $\bar{r}_i$ and r_i, $i \in \{1, 2\}$, can be found in Appendix B.1. □

Observe that in the region $m < f$, the bad bank defaults upon arrival of the liquidity shock. Further-

asymmetric information, the double subscript $V_{\ell,\tilde{\ell}}$ denotes the weighted sum of a bank of type ℓ that is considered to be of type $\tilde{\ell}$ by the market, as in Eq. (2.2).

more, note that when $m_G > \overline{m}^*$ for the good bank and $m_B > \overline{m}^* + f$ for the bad bank, an additional region shows up in the value function. When $m \in (\overline{m}^*, m_\ell]$ after the shock, shareholders receive a lump-sum payment of $m - \overline{m}^*$. We will encounter this scenario later on in the pooling equilibrium, in which case the market cannot distinguish the good and bad types.

2.2.3 Benchmark case: full information

Before analyzing the effects of asymmetric information on the bank's dividend strategy, we consider the benchmark case in which all agents have full information about the bank's type. In this scenario, both good and bad banks follow their privately optimal dividend policy. The valuations of the two bank types in the full information case are summarized in the following proposition. Define the operator $x^+ := \max\{x, 0\}$.

Proposition 7. *Under full information, the value of a bank of type ℓ before the shock is given by:*

$$V_\ell^*(m) = V_{\ell,\ell}(m; m_\ell^*) = V_\ell(m; m_\ell^*),$$

where optimal target cash level m_ℓ^* is pinned down by high-contact condition:

$$\lim_{m \uparrow m_\ell^*} V_\ell''(m; m_\ell^*) = \lim_{m \downarrow m_\ell^*} V_\ell''(m; m_\ell^*).$$

In specific, the following relation holds at m_ℓ^*:

$$V_\ell^*(m_\ell^*) = \frac{\mu}{\rho+\lambda} + \frac{\lambda}{\rho+\lambda}(m_\ell^* - f\mathbb{1}_{\{\ell=B\}}) = (1-\gamma)W(\overline{m}^*) + \gamma W((m_\ell^* - f\mathbb{1}_{\{\ell=B\}})^+).$$

For the good bank, $m_G^* = \overline{m}^*$ and above relation simplifies to:

$$V_G^*(m_G^*) = W(m_G^*) = \frac{\mu}{\rho}.$$

Proof. See Appendix B.1.4 □

The optimal dividend strategy m_ℓ^* is the cash level at which the marginal benefit of retaining cash equals the marginal benefit of paying out cash. For the good bank, this trade-off boils down to having a precautionary savings motive against the Brownian shock versus the impatience of its shareholders. The above proposition states that the optimal dividend policy of the good bank corresponds to the optimal after-shock strategy, which is characterized in Eq. (B.1). This is an intuitive result, as the good bank is not affected by the liquidity shock and it is fairly priced in the full-information benchmark. As a result, the value of the good bank at its optimal value simplifies to μ/ρ, being the discounted value of a perpetual bond payer a continuous dividend μ per unit of time.

The bad bank faces a more complicated trade-off when deciding on its optimal dividend strategy, as it also has to incorporate the Poisson risk component. Figure 2.2 shows that the relation between the optimal payout boundary of the bad bank m_B^* and shock size f for different values of cash flow volatility σ. What may come across as unexpected is that the link between m_B^* and f can go in two ways. Figure 2.2(a) shows that for a low value of σ, the bad bank pays out dividends at a higher cash level when f increases. For high values of σ, this relationship inverses, and eventually becomes flat, as is visible in Figure 2.2(b).

As the good bank's value at its first-best strategy should be larger than the bad bank's value at its first-best strategy, we have that $W(m_G^*) \geq W(m_B^* - f)$, which by the monotonicity of W implies that

$m_G^* \geq m_B^* - f$. In changing m_B^* in response to a higher f, it is never optimal to increase the buffer more than the increase of the potential loss. Therefore, $m_G^* + f$ serves as an upper bound of m_B^*. As a result, case (ii) of the bad bank's value function in Proposition 6 is irrelevant.

In Figure 2.2(a), the boundaries are ordered as follows: $m_B^* > m_G^* > m_B^* - f > 0$. When $m_B^* < f$, the bad bank has optimally set the default boundary so low that it will be wiped out when the shock hits. In this scenario, their exists a closed-form expression for m_B^*, which is smaller than m_G^* and can be found in Eq. (B.4), so that overall $m_G^* > m_B^* > 0 > m_B^* - f$. This implies that the bad bank hoards less cash to hedge against the Browian default risk compared to the case where there is no Poisson risk. This case corresponds to the flat part of the function in Figure 2.2(b). The decreasing left part of this part corresponds to the ordering $m_G^* > m_B^* > m_B^* - f > 0$.

For the remainder of this paper, we will focus on the case where the precautionary savings motives dominate the effect of impatience and that the bad bank's first-best dividend strategy is more conservative than the good bank's.

Assumption 7. *Parameter values are set such that $m_B^* > m_G^*$.*

We argue that this is a reasonable assumption for the banking industry, which is among the industries with the largest payout ratios. The numerical analysis shows that the above assumption holds for reasonable values of cash flow volatility σ and is only violated for relatively high values of σ. This is in line with our view that banks have relatively stable cash flows in normal times, but are subject to tail risk events.[14]

For the sake of completeness, we provide the set-up of a parallel analysis when Assumption 7 is violated in Appendix B.4. In this scenario, the bad bank can mimic the good bank by increasing its target cash level. In response, the good bank can decide to signal its quality by further increasing its dividend payout level.

2.3 Signaling through dividend policy

In the perfect information case, different bank types choose different target cash levels. This might not be the case under asymmetric information. As the bank's objective function includes a market-valuation component, the bad bank has incentives to mimic the good bank's dividend policy at the cost of adopting a sub-optimal dividend policy. In return, the good bank does not wish to be mimicked by the bad bank, as this reduces its market valuation. For that reason, the good bank has an incentive to impose mimicking costs on the bad bank by distorting its dividend policy. The bad bank might accept a more aggressive dividend policy in return for a higher market valuation, but anticipating the pending shock, it has an additional precautionary savings motive, so that there is a limit to how far the bad bank wants to go in following the good bank's strategy.

In this section, we first confirm that the dividend policy can be used as a signaling mechanism by showing under which conditions the Spence-Mirrlees condition holds. Then, we will analyze the separating and pooling equilibrium.

[14] An example of the opposite case could be some of the most successful companies in the volatile technology sector, building seemingly excessive cash buffers, see The Economist (2017).

2.3.1 Single-crossing condition

We will now show that the dividend policy can be used as a signaling device. When deciding whether to mimic or separate, each bank type makes a trade-off between the cost of distorting the dividend policy with the change in the market valuation. The single-crossing condition in Proposition 8 shows that the elasticity between the change of market valuation and the dividend policy depends positively on the type ℓ. This implies that in the region defined, it is less costly for the good bank to distort its dividend policy and get a higher market valuation than it is for the bad bank. As a result, outside investors can view the dividend policy, or equivalently, the target cash level, as a credible signal. Functions Δ and Γ are defined in Eq. (B.11) and (B.12).

Proposition 8. *If the following condition is satisfied:*

$$\left.\frac{\partial V_G(m;m_D)}{\partial m_D}\right|_{m_D=\overline{m}^*} < \left.\frac{\partial V_B(m;m_D)}{\partial m_D}\right|_{m_D=\overline{m}^*}$$
$$\iff (1-W'(\overline{m}^*-f))\Gamma(\overline{m}^*)+\Delta(\overline{m}^*)W''(\overline{m}^*-f)+W'(0)\Delta'(f)<0,$$

there exists a unique point $m_{SC}\in(f,\overline{m}^*)$ such that for every $m_D\geq m_{SC}$, the single-crossing property holds:

$$\frac{V_{G,G}(m;m_D)-V_{G,B}(m;m_D)}{\partial V_{G,\tilde{\ell}}/\partial m_D} > \frac{V_{B,G}(m;m_D)-V_{B,B}(m;m_D)}{\partial V_{B,\tilde{\ell}}/\partial m_D}.$$

This condition is equivalent to:

$$\frac{\partial V_G(m;m_D)}{\partial m_D} < \frac{\partial V_B(m;m_D)}{\partial m_D}, \quad \forall m\geq 0, m_D\geq m_{SC}.$$

This implies that the high-type bank finds it less costly to distort the dividend policy than the low-type bank regardless of its current cash-level.

Proof. See Appendix B.3. □

This proposition implies that lowering the dividend policy m_D is considered a valid signal for all values $m_D\in(m_{SC},\overline{m}^*)$, which is a sufficient condition for a separating equilibrium to exist. Note that for $m_D=f$, the single-crossing condition is always violated, meaning that a good bank cannot signal its type by pushing down its dividend boundary to a level f. The reason is that for $m_D\leq f$ the bad bank is wiped out when the shock arrives, which alters its optimal behavior (see Chapter 5.1 of Moreno–Bromberg and Rochet (2018)). Limited liability and the impatience introduced by the future wipe-out lead the bank to pay out earlier. Even at levels of m_D above f this effect might dominate the banks' precautionary savings motive, and m_{SC} is therefore strictly above f. In our numerical analysis, under our Assumption 7 the precautionary savings motive is strong enough for the existence of m_{SC}. Furthermore, m_{SC} is systematically lower than $\overline{m}^S$.

Preposition 8 additionally establishes that the validity of dividend policy as a signal does not depend on the current cash levels. There is no incentive for a bank to pretend to follow a different policy than it would at its payout threshold if it, for instance, gets close to default. Therefore an additional commitment device is not necessary for investors to trust banks' policy announcements.

2.3.2 Separating equilibrium

When the market cannot observe the bank's type, the good bank can decide to signal its type by paying out dividends earlier, to the extent that the bad bank does not want to mimic the good bank. We describe the mechanism underlying the separating equilibrium below. When deciding whether to mimic or not, the bad bank balances the cost of having to issue dividends out earlier versus the benefit of having a higher market valuation. By deviating from its first-best strategy, the good bank can affect this trade-off. A more aggressive payout policy lowers the market value of a good bank, making it less attractive to mimic. Furthermore, the bad bank suffers a mimicking cost by hoarding less cash than in the first-best strategy. We establish the existence of a dividend payout level such that the bad bank does not find it profitable to mimic anymore.

First, one needs to determine whether the incentive compatibility constraint (ICC) of the bad bank holds. Suppose the good bank picks a dividend payout strategy m^S. If the bad bank mimics, its value function will be $V_{B,G}(m;m^S)$. If instead the bank refrains from mimicking the good bank, it follows its first-best strategy so that its value function becomes $V_{B,B}(m;m_B^*)$. Evaluated at m^S, the bad bank prefers mimicking the good type at $m^S < m_B^*$ when:

$$\underbrace{V_{B,B}(m^S;m_B^*)}_{\text{first-best}} \geq \underbrace{V_{B,G}(m^S;m^S)}_{\text{value when mimicking}} \tag{2.3}$$

When this condition does not hold at $m^S = m_G^*$, the good bank will have to deviate from its privately optimal strategy m_G^* in the separating equilibrium. The following proposition shows that there exists a solution $\overline{m}^S$ to Eq. (2.3). In all the numerical applications, this solution is unique. See also Figure 2.1 for a graphical representation of the two functions and their intersection points.

Proposition 9. *A solution $\overline{m}^S \in (0, m_B^*)$ exists to the equation $V_{B,B}(m^S;m_B^*) = V_{B,G}(m^S;m^S)$.*

Proof. Note at $m^S = m_B^*$ the value when mimicking dominates the first-best value, i.e., $V_{B,B}(m_B^*;m_B^*) < V_{B,G}(m_B^*;m_B^*)$. At $m^S = 0$, we have $V_{B,B}(0;m_B^*) = V_{B,G}(0;0) = 0$. Denote $\tilde{V}_{B,G}(m_S) := V_{B,G}(m_S;m_S)$, and $\tilde{V}_\ell := V_\ell(m_S;m_S)$ for $\ell \in \{G,B\}$. A sufficient condition for existence of $\overline{m}^S$ is that $V'_{B,B}(0;m_B^*) > \tilde{V}'_{B,G}(0)$. After some algebraic manipulations, it follows that $\tilde{V}'_G(0) = \tilde{V}'_B(0) = 1$, such that $\tilde{V}'_{B,G}(0) = k\tilde{V}'_G(0) + (1-k)\tilde{V}'_B(0) = 1$. Since $V'_{B,B}(0;m_B^*) > 1$, the existence condition is satisfied. □

To determine whether paying out dividends at or below $\overline{m}^S$ is an equilibrium strategy, we check incentive compatibility of the good bank. The good bank has an incentive to separate when its value in the separating equilibrium is larger than its value when mimicking the bad bank. That is:

$$\underbrace{V_{G,G}(m^S;m^S)}_{\text{value in separating eqbm}} \geq \underbrace{V_{G,B}(m^S;m_B^*)}_{\text{value when mimicking}} . \tag{2.4}$$

The threshold $\underline{m}^S$ for which Eq. (2.4) is binding represents the lowest target cash level such that the good bank prefers separation over mimicking and such that observing the target cash level m^S can safely be interpreted as a signal by outsiders. A separating equilibrium exists only if $\underline{m}^S \leq \overline{m}^S$. Note that by the optimality of m_G^* in the full information case, it follows that $\underline{m}^S \leq m_G^*$.

A sufficient condition for $m^S \in [\underline{m}^S, \overline{m}^S]$ to be a Perfect Bayesian Equilibrium (PBE) is that the good bank does not have an incentive to defect to a different strategy given a set of out-of-equilibrium beliefs. It suffices to show that this holds under the pessimistic belief that the good bank is of the bad type instead, which corresponds to the following condition:

$$V_{G,G}(m^S; m^S) \geq V_{G,B}(m^S; m^*_{G,B}), \tag{2.5}$$

where $m^*_{G,B}$ is the cash target chosen by the good bank when it is considered to be of the bad type by the market.

A separating equilibrium exists when there is a m^S for which the three conditions in Eq. (2.3), (2.4) and (2.5) are jointly satisfied. Observe that when $\underline{m}^S \geq m^*_G$, there will be a separating equilibrium in which both banks choose a strategy that coincides with the first-best strategy. In the reverse case, the good bank has to deviate from its optimal strategy to prevent the bad bank from mimicking. Denote by $\underline{m}_L$ and $\underline{m}_H (> \underline{m}_L)$ the two solutions to Eq. (2.4) and let $\tilde{m}_L$ and $\tilde{m}_H (> \tilde{m}_L)$ be the two solutions to Eq. (2.5). It is straightforward to show that $\underline{m}_L < \tilde{m}_L < m^*_G < \tilde{m}_H < \underline{m}_H$. A separating equilibrium exists only if $\overline{m}^S \geq \tilde{m}_L$. Since paying out dividends earlier than m^*_G is costly for good banks, it will select the minimum of $\overline{m}^S$ and m^*_G. The good bank has no incentive to deviate from this strategy which can be sustained under pessimistic beliefs. After applying the Cho-Kreps Intuitive Criterion, see Cho and Kreps (1987), the least-cost separating contract strategy is uniquely selected. The following proposition formalizes this.

Proposition 10. *There exists a separating equilibrium in which both banks' market valuations correspond to their intrinsic valuations when $\overline{m}^S \geq \tilde{m}_L$. In the least-cost separating equilibrium, the good bank pays out dividends more aggressively than in the first-best case and its value is given by:*

$$V_G^{lcs}(m) = \begin{cases} V_G(m; \overline{m}^S), & \textit{for } \overline{m}^S < m^*_G, \\ V_G(m; m^*_G), & \textit{otherwise.} \end{cases}$$

The bad bank pays out at its first-best strategy and has value $V_B(m; m^*_B)$.

2.3.3 Pooling equilibrium

In the pooling equilibrium, outsiders are not able to determine the bank's type. As a result, for dividend strategy m^P, the market valuation component of both bank types is

$$V_p(m; m^P) = \alpha V_G(m; m^P) + (1 - \alpha) V_B(m; m^P).$$

To determine the existence of the pooling equilibrium, we first determine whether mimicking the good type is an optimal strategy for the bad bank. This is the case when the following incentive compatibility constraint holds:

$$\underbrace{V_{B,p}(m^P; m^P)}_{\text{value when pooling}} \geq \underbrace{V_{B,B}(m^P; m^*_B)}_{\text{first-best}} \tag{2.6}$$

Let $\overline{m}^P_L$ and $\overline{m}^P_H (> \overline{m}^P_L)$ be the two solutions to Eq. (2.6). Since $V_p(m; m^P) > V_B(m; m^P)$, the threshold $\overline{m}^P_L \in (\overline{m}^S, m^*_B)$ and $\overline{m}^P_H > m^*_B$.

Without further refinements, we face multiplicity of equilibria, which is a common feature of signaling games. Maskin and Tirole (1992) consider the mechanism design game in which the informed principal (bank management in our setting) offers a contract ex ante to the uninformed outsiders. They show that only those pooling equilibria survive that Pareto-dominate the least-cost separating equilibrium as was characterized in Proposition 10. Therefore, the remaining restriction is that the value of the good bank in the pooling equilibrium is larger than in the least-cost separating equilibrium:

$$\underbrace{V_{G,p}(m^P;m^P)}_{\text{value when pooling}} \geq \underbrace{\mathbb{1}_{\{m_G^* \leq \overline{m}^S\}} V_{G,G}(m^P;m_G^*) + \mathbb{1}_{\{m_G^* > \overline{m}^S\}} V_{G,G}(m^P;\overline{m}^S)}_{\text{value when separating}}. \tag{2.7}$$

Let $\underline{m}_L^P$ and $\underline{m}_H^P$ be the two solutions to Eq. (2.7). There will be pooling equilibria if and only if there is a range of m^P for which conditions (2.6) and (2.7) hold. Note that in the case that $m_G^* \leq \overline{m}^S$, condition (2.7) is violated because the good bank cannot do better than its first-best strategy. Furthermore, if $\overline{m}_L^P > \underline{m}_H^P$, there will not be a strategy m^P for which conditions (2.6) and (2.7) hold.

Let $m_{\ell,p}^*$ be the best pooling equilibrium target cash level of bank type ℓ, see Appendix B.1.5. As $m_{G,p}^* < m_{B,p}^*$ for the parameters considered, there will not be a single Pareto-optimal pooling equilibrium. Notice that a necessary condition for a pooling equilibrium to exist is that $m_{G,p}^*$ satisfies condition (2.7). Furthermore, note that possible pooling strategies outside the range $[m_{G,p}^*, m_{B,p}^*]$ are Pareto-dominated by either $m_{G,p}^*$ or $m_{B,p}^*$.

In the range of Pareto-dominant pooling equilibria, we focus our attention on the best pooling equilibrium for the good type $m_{G,p}^*$. When outsiders infer there is a pooling equilibrium, they expect the good bank to be the first type to start paying out dividends at $m_{G,p}^*$, being a more aggressive policy than $m_{B,p}^*$. The bad bank does not want to reveal its type and follows. The selection of the pooling equilibrium will not alter the results qualitatively, only quantitatively.

Proposition 11. *A pooling equilibrium exists in which the market valuation of both bank types is $V_p(m) := \alpha V_G(m) + (1-\alpha)V_B(m)$ and both types pay out dividends at threshold m^P if conditions (2.6) and (2.7) are satisfied. Compared to the first-best case, good banks pay dividends later and bad banks pay earlier in the pooling equilibrium.*

2.3.4 Numerical analysis

Exogenous parameters

Table 2.1 displays the parameter baseline values. The cash reserve at time $t = 0$ is assumed to be $m = 1$.[15] The risk-free rate is set to $\rho = 0.035$. The mean after-coupon cash flow is $\mu = 0.1$ and the volatility σ is set to 0.1. These parameter values, which all assumed to be annual rates, are similar in size to those used in Hugonnier and Morellec (2017) and Klimenko and Moreno–Bromberg (2016). In the baseline case, we assume that bad banks compromise a fraction $1-\alpha = 0.2$ in the economy being subject to a liquidity shock of size $f = 0.15$ with an expected waiting time of $\lambda^{-1} = 5$ years. Lastly, we assume that the weight that management puts on the market valuation is $k = 0.5$.

Figure 2.3 displays the selected equilibrium strategies for different parameter values in the absence of any regulatory measures. We consider parameters corresponding to three different model elements: shock arrival rate and size, investor beliefs and preferences, and cash flow drift and volatility.

Liquidity shock

Figure 2.3(a) shows the comparative statics for liquidity shock arrival rate λ. Naturally, all lines coincide at $\lambda = 0$, as at this point, the good and bad types are identical, making signaling behavior irrelevant. As λ increases, the optimal cash target of the bad bank increases, while the optimal target of the good bank is unaltered. At first, higher λ makes mimicking more attractive so that the good bank needs to

[15] This value is generally higher than the resulting dividend thresholds in our model, implying that banks make a lump-sum payment at $t = 0$.

lower its cash target more aggressively to achieve a separating equilibrium. Eventually, the effect of a more likely shock on the bad banks' intrinsic value dominates, and the good bank can separate with a relatively smaller deviation from its first-best strategy. Facing this reaction function, the good bank accepts the pooling equilibrium for low values of λ where the cost of being pooled with the bad bank is relatively small. As λ increases, the cost of the pooling equilibrium relative to the cost of separating by distorting from optimal strategy m_G^* decreases. As a result, the equilibrium switches to a separating equilibrium.

Figure 2.3(b) shows the effects of shock size f. Again, all lines coincide at $f = 0$. With effects similar to increasing the shock arrival rate, a pooling equilibrium forms for low values of f. Eventually, it is replaced by a separating equilibrium that gradually converges to a first-best equilibrium where the good bank does not need to distort its payout policy to deter the bad bank.

Investors

Figure 2.3(c) displays the effects of the fraction of good banks α on the equilibrium outcome. One can observe that only the target cash level corresponding to the pooling equilibrium moves in α. With a small fraction of good banks, the market value of the average bank in the pooling equilibrium is dominated by the value of bad banks. This makes the cost of being pooled very high for the good banks in the economy, resulting in a separating equilibrium. As α increases, the payout policy of the pooling equilibrium converges to the first-best policy of a good bank. The value of the bad banks drags down the market value of the average bank, which enters the value function of the good bank in the pooling equilibrium and its optimization for $m_{G,p}^*$, see Section 2.3.3. The decreasing degree of distortion implies that the pooling equilibrium gradually becomes more attractive for a good bank while the separating equilibrium strategy remains unchanged. Therefore, the banks switch to a pooling equilibrium at some point. In the extreme case where $\alpha = 1$ and there are only good banks, the optimal pooling strategy coincides with the good bank's first-best strategy.

The effects of changing short-term investor fraction k (or interpreted alternatively, the strength of investors' liquidity concerns) can be found in Figure 2.3(d). As was the case for α, first-best policies m_G^* and m_B^* are unaffected by changing k as market value and intrinsic value coincide under full information. However, as the weight on the market value in the value function increases, so does the bad bank's benefit of mimicking the good bank. As a result, the good bank cannot deter the bad bank without distorting its payout policy. Since the good bank sets a higher cash target when it is pooled with the bad bank, pooling equilibrium policy $m_{G,p}^*$ is increasing in k. The good bank is motivated to do so because its market value depends on the intrinsic value of the bad bank in the pooling equilibrium. As can be observed in the plot, the (negative) slope of the deterioration policy is steeper than the (positive) slope of the pooling policy, making it increasingly costly for the good bank to separate and leading up to a pooling equilibrium.

Cash flow process

Figure 2.3(e) shows the effects of changing cash flow drift μ. For low values of μ, the relative effect of liquidity shock f is larger than for large values of μ, as the time to build a buffer sufficient to offset the shock is longer (in expectation). This effect creates stronger mimicking incentives for the bad bank, making it more costly for the good bank to deter it, resulting in a pooling equilibrium is selected. As the first-best target cash level of the good bank m_G^* decreases faster in μ than that of the bad bank m_B^*, the good bank's cost of distorting its policy becomes small enough for it to prefer a separating equilibrium.

Lastly, Figure 2.3(f) looks at the effects of cash flow volatility σ. For low values of σ, the probability of default as a result of the Brownian risk component is relatively low. As σ increases, the relative risk of being wiped out by the Poisson risk relative to the Brownian risk becomes smaller, so that the first-best strategies of the good and bad bank slowly converge. When σ is small, the difference between the optimal strategy of the good and bad bank is relatively big, making it rather costly for the bad bank to mimic the good bank. As a result, the good bank does not need to distort its strategy. As σ increases further, the bad bank's benefit of mimicking the good bank increases, so that the good bank needs to distort its dividend strategy (more) to deter it. For high σ, the cost of being pooled for the good bank is smaller since the two banks are relatively more similar and a pooling equilibrium prevails.

2.4 Dividend restrictions

This section considers the effects of the regulator imposing restrictions on the bank's payout policy, or equivalently, the bank's required cash levels. Throughout the analysis, we assume that the dividend restriction is lifted upon arrival of the liquidity shock.[16] We will now look at the effect of dividend restrictions in the case of (i) a pooling equilibrium, (ii) a separating equilibrium, and (iii) the first-best separating equilibrium. We will see that dividend restrictions have the potential to break the separating equilibrium.

2.4.1 Construction of restricted equilibrium

(i) Pooling equilibrium m^P

First, suppose that the bank plays a pooling equilibrium m^P in the absence of a dividend restriction. When $m^R \leq m^P$, the pooling equilibrium is not affected. For $m^R > m^P$, the restriction starts to constrain the pooling equilibrium. The restricted pooling equilibrium $\tilde{m}^P$ needs to satisfy the following conditions. Similar to condition (2.6), the bad bank should prefer to pool rather than to play its now *restricted* first-best strategy:

$$V_{B,p}(\tilde{m}^P;\tilde{m}^P) \geq V_{B,B}(\tilde{m}^P;\max\{m_B^*, m^R\}).$$

Furthermore, the good bank should prefer to pool rather than to mimic the bad bank:

$$\underbrace{V_{G,p}(\tilde{m}^P;\tilde{m}^P)}_{\text{value when pooling}} \geq \underbrace{V_{G,B}(\tilde{m}^P;\max\{m_B^*, m^R\}).}_{\text{value when mimicking bad bank}}$$

To ensure that the good bank has no incentive to deviate, the value in the pooling equilibrium should be larger than when the bank deviates to an out-of-equilibrium strategy. Under pessimistic out-of-equilibrium beliefs, this translates to the following condition:

$$V_{G,p}(\tilde{m}^P;\tilde{m}^P) \geq V_{G,B}(\tilde{m}^P;\max\{m_{G,B}^*, m^R\}).$$

In the unrestricted case, the pooling equilibrium is $\max\{m_{G,p}^*, \overline{m}_L^P\}$. In the scenario where $m_{G,p} > \overline{m}_L^P$, a restriction $m^R > m_{G,p}$, the good bank cannot play its optimal pooling strategy anymore. As there is no gain in setting a higher dividend threshold than the restriction, it sets $\tilde{m}^P = m^R$, which will also be followed by the bad bank. A similar reasoning applies when $m^R > \overline{m}_L^P > m_{G,p}$.

(ii) Separating equilibrium $m^S = \overline{m}^S$

Suppose now that absent a dividend restriction, the bank plays a (least-cost) separating equilibrium with strategy $\overline{m}^S (< m_G^*)$. When $m^R > \overline{m}^S$, the separating equilibrium breaks down and a pooling equilibrium as described above arises.

[16]We have performed the analysis in which the dividend restriction remains active after the liquidity shock has arrived. As this did not significantly change the results, we leave these results untabulated.

(iii) First-best separating equilibrium $m^S = m_G^*$
Lastly, consider the case that absent a dividend restriction, the bank types are in the first-best separating equilibrium, that is, the good bank plays optimal $m_G^* \in [\underline{m}^S, \overline{m}^S]$ and the bad bank m_B^*. A dividend restriction $m^R < m_G^*$ will have no effect on the equilibrium. When m^R is set such that $m^R \in [m_G^*, \overline{m}^S]$, the restricted least-cost separating equilibrium is at the level m^R. Whereas in the first-best case, a pooling equilibrium did not exist, this can change in the presence of the restriction. A pooling equilibrium emerges when conditions (2.6) holds and

$$\underbrace{V_{G,p}(m^P; m^P)}_{\text{value when pooling}} \geq \underbrace{V_{G,G}(m^P; m^R).}_{\substack{\text{value in restricted least-cost} \\ \text{separating equilibrium}}}$$

In the case that $m^R \geq \overline{m}^S$, the first-best separating equilibrium breaks into a pooling equilibrium as described before.

From this analysis, it becomes clear that dividend restrictions have the potential to break the separating equilibrium. Breaking the separating equilibrium makes the good bank less likely to default as it cannot use an aggressive dividend policy to signal its type. At the same time, provided that $m^R < m_B^*$, dividend restrictions that break the separating equilibrium make the bad bank less safe as it abandons its first-best optimal payout policy and pays out dividends at a lower level of cash-holdings to mimic the good bank. In the remainder of this section, we continue the numerical analysis and look at the effect of dividend restrictions on the bank types' payout strategies, 1-year default probability, and valuations.

2.4.2 Effects of dividend restrictions

Figure 2.4 shows the effects of a dividend restriction m^R that is in place *before* the liquidity shock for two different parameter settings. Parameter setting (i) represents a scenario with a large but concentrated shock, whereas the shock is small but widespread in parameter setting (ii).

The main metrics of interest are the average 1-year default probability and the value of the average bank in the economy.[17] In addition to the average bank value, the regulator cares about bank default, as it generally has large negative externalities on the economy. To not impose additional assumptions, we think of these externalities as being the same for both types, so that the regulator puts equal weight on their respective default risk.

Effects of dividend restriction on payout strategies

The dividend payout strategies of the good and bad bank are displayed in Figures 2.4(a) and 2.4(b). For low values of restriction m^R, the separating equilibrium is not affected, so that the good bank can signal its type by adopting a more aggressive dividend policy, and the bad bank sticks to its first-best strategy. As soon as m^R starts constraining the separating equilibrium, the equilibrium switches to the pooling type, which means that the bad bank pays out dividends at a lower cash level than it would have done in the first-best case. When restriction m^R starts binding the (unconstrained) pooling equilibrium, both banks pay dividends at m^R.

[17] We do not distinguish between the market and intrinsic valuation of the average bank, as these metrics coincide. This follows naturally in the separating equilibrium where the banks signal their type to the market. In the pooling case, the market valuation of the average bank is $\alpha V_{G,p}(m) + (1-\alpha)V_{B,p}(m) = \alpha\left(k[\alpha V_G(m) + (1-\alpha)V_B(m)] + (1-k)V_G(m)\right) + (1-\alpha)\left(k[\alpha V_G(m) + (1-\alpha)V_B(m)] + (1-k)V_B(m)\right) = \alpha V_G(m) + (1-\alpha)V_B(m)$. This value corresponds to the average intrinsic value.

Effects of dividend restriction on 1-year default probability

One of the regulator's main concern is the effect of regulatory measures on the default probabilities in the banking industry. Let the probability at $t = 0$ that a bank defaults within a time horizon T be denoted by:

$$\mathrm{PD}_\ell^T = \mathbb{P}\left(\tau_\ell^\pi < T \mid M_{\ell,0}^\pi(t) = m\right).$$

In the numerical analysis of this paper, we consider the 1-year default probability, i.e., we set $T = 1$. A description of the computation of PD_ℓ^T can be found in Appendix B.2 and follows the methodology of Klimenko and Moreno–Bromberg (2016).

Figures 2.4(c) and 2.4(d) show the 1-year default probabilities of the good bank, the bad bank, and the average bank (i.e., the weighted sum of good and bad banks in the economy) for the two parameter settings. In both parameter settings, the good bank has a larger default probability than the bad bank in the separating equilibrium. Even though the good bank is not subject to Poisson risk, it sets its cash target so low that it becomes more likely to default than the bad bank. When m^R starts to constrain the separating equilibrium, the default probability of the bad bank spikes up, as it now hoards less cash than in the separating equilibrium. In contrast, the good bank becomes safer now that it does no longer play the aggressive separating policy. When m^R eventually starts binding the pooling equilibrium, both banks become safer again. The effect of the dividend restriction on the average default probability can go in two ways. Figure 2.4(c) shows that the average default risk increases when the separating equilibrium changes into a pooling equilibrium, whereas the opposite is true in Figure 2.4(d).

Effects of dividend restriction on bank valuations

The effect of dividend restriction m^R on the intrinsic, market, and total bank value can be found in Figures 2.4(e) and 2.4(f). Note that there is only a single line for the average bank, as the market valuation and the intrinsic valuation of the average bank coincide. This follows from the assumption of rational expectations that says that investors' beliefs about the fraction of the two types of banks are correct. An overvaluation of one type means an undervaluation of the other which cancels out on average. In line with this, observe that the market valuations of the good and bad bank in the pooling equilibrium coincide with the intrinsic valuation of the average bank.

Both figures show that when restriction m^R breaks the separating equilibrium, the intrinsic value of the good bank changes as it abandons its aggressive separating policy. In parameter setting (i), the intrinsic value slightly drops, which implies that the value of the good bank under the aggressive separating strategy m_G^* is slightly higher than in the pooling strategy $m_{G,p}^*$. For parameter setting (ii), the good bank's intrinsic value increases, as the good bank is not playing the aggressive separating policy anymore. By contrast, the good bank's market valuation drops in both parameter settings, since it is now indistinguishable from the bad bank for outsiders. The opposite effect applies to the bad bank. Initially, the bad bank's intrinsic value decreases when the separating equilibrium is broken by m^R since the bad bank now deviates from its first-best strategy. Meanwhile, its market valuation jumps up since it is now pooled with the good type. When the restriction is tightened further, both market and intrinsic valuations of the bad bank increase, up to the point where the restriction starts constraining the bank's respective first-best strategies.

A notable outcome of this type of dividend restriction is that in switching from the separating equilibrium to the pooling equilibrium, the payout policy of the good bank jumps up. With such a policy in place, it may very well seem like the dividend restriction is not binding, as it will be the case for a restriction below $m_{G,p}^*$. As such, the regulator might come to believe that a dividend restriction in place does not affect current bank financial policies.

Furthermore, in the event of a change of economic conditions, a previously non-binding restriction can become binding. As an example, consider an upward shift in the shock arrival rate λ from 0 to 0.1 as depicted in Figure 2.3(a). With no pending shock, the two bank types are identical and any restriction set below the current shared policy of 0.4 is not binding. For concreteness, consider dividend restriction level $m^R = 0.3$. Under the threat of a liquidity shock, the good bank would, in the absence of regulation, choose a more aggressive policy of $\overline{m}^S \approx 0.34$ to achieve separation. However, the presence of the dividend restriction makes this unachievable, and the laissez-faire separating equilibrium does not materialize. As an extension, a minor tightening, such as the activation of a counter-cyclical capital buffer, might have a major impact when imposed in response to changes in the economic environment.

Whether or not a dividend restriction is desirable depends on which of the two is the larger evil: a pooling equilibrium in which the bad bank pays out inefficiently early (and the good bank sub-optimally late), or the separating equilibrium, in which the good bank pays out inefficiently early. Next, we quantify these effects further by analyzing a number of relevant scenarios.

2.4.3 Scenario analysis

To illustrate the pitfalls and potential for regulatory intervention, we focus on two fundamental scenarios that can be mapped to the sources of shocks introduced above: a large but concentrated shock, and a smaller but more widespread shock, see Figure 2.4. We consider two relevant dimensions of the liquidity shock: the *scope* of the shock, i.e., the fraction of affected banks $1-\alpha$, and the *size* of the shock, i.e., the impact of the shock on the bad bank's liquid reserves f. We analyze two channels through which the scope of the shock operates when regulation enforces a pooling equilibrium: a *direct* channel and an *indirect* channel. The direct channel works as follows: more bad banks in the economy make, other things equal, dividend regulation less attractive, as the bad banks switch from their first-best to a distorted policy in the pooling equilibrium, thereby lowering their intrinsic values and increasing default risk, see Figure 2.4. The indirect channel represents the additional effect that arises from the banks' adjustment of their policies in response to the scope of the shock. That is, when there are few (many) bad banks in the economy, the pooling equilibrium strategy $m^*_{G,p}$ that is selected by the good bank tilts more (less) towards the first-best strategy of the good bank m^*_G, creating a bigger (smaller) distortion from the first-best of the bad bank m^*_B. The *size* of the shock f determines whether a separating equilibrium would have been selected in the absence of regulation, and if so, which of the two channels dominates.

Large concentrated shock

In the first scenario, we consider that a small number of banks (large α) are subject to a significant negative shock (large f). This scenario could arise from a third-party trading loss (e.g., Archegos capital) or a wave of fines from misconduct (e.g., North European money-laundering scandals). We find that a regulator should be cautious in imposing payout restrictions in this scenario, as it risks having an adverse effect on both of the industry-wide metrics by simultaneously decreasing the average value of banks *and* raising the risk of bank defaults, see the yellow lines in Figures 2.4(c) and 2.4(e), respectively.

Decomposing the average effects into their constituents, the big increase in default risk of bad banks and the large drop in their intrinsic value dominate the industry outcome, even though they only constitute a small fraction of the economy. This illustrates the importance of considering the indirect effect of the concentrated exposure through the strategic equilibrium behavior. On the one hand, the small fraction of bad banks implies that the payout policy under the pooling equilibrium heavily tilts towards the first-best policy of the good banks m_G^*, see Section 2.3.4. This requires a large deviation for bad bank from their first-best strategy m_B^*. On the other hand, this change also makes obtaining the market value of an average bank more attractive for the bad bank. In equilibrium, the relatively small number of bad banks have a seemingly out-sized impact on the industry outcome. This is especially surprising for the regulation-induced pooling equilibria, considering that the presence of the bad banks does not dramatically alter the behavior of the good banks. In summary, regulation has a relatively limited positive impact on the safety of a large base of banks, while making a small group radically more unsafe.

Furthermore, regulation that induces a pooling equilibrium results in a large value transfer from outsiders who were to buy the shares to the insiders of these banks taking the other side of that trade. Figure 2.4(e) shows that the bad bank's market value increases drastically and the intrinsic value decreases dramatically, see the dashed and solid black lines in Figure 2.4(e), respectively.

Small widespread shock

In the second scenario, we consider the presence of a minor shock (small f) that affects a broad segment of the banking industry (small α). Examples of this scenario are the risk of sudden moves in asset prices leading to margin calls on derivative contracts (see European Systemic Risk Board (2020a)), losses related to the deterioration of collateral resulting in higher haircuts (see Shleifer and Vishny (2011)), and costly restructuring of funding conditions. The concerns related to the banking sector during economic crises such as the Covid-19 crisis (see European Systemic Risk Board (2020b)), are more complex than what can be captured by a one-off shock. However, as a first approximation, a pending sector-wide liquidity shock can provide some intuition about the strategic response to regulation under such circumstances.[18]

In contrast to the large but concentrated shock studied above, the outlook for regulation in the second scenario is more promising. Breaking the separating equilibrium decreases the default risk *and* increases the value of the average bank, see the yellow lines in Figure 2.4(d) and 2.4(f). We classify this scenario in which regulation improves both metrics as a *regulatory Goldilock* scenario. The potential for value creation arises from the distortion introduced by signaling. Comparing the policy of the good bank in the previous case to the current, as captured in Figure 2.4(a) and 2.4(b) respectively, it is indeed the case that the good bank's target cash-level in the separating equilibrium is set more aggressively. Comparing the strategy of the bad bank in the same plots, we see that the smaller shock substantially lowers the bad bank's first-best payout boundary, which pushes the good bank to an even lower cash target in the separating equilibrium. Note that α has no impact on the payout levels in the separating equilibrium where information is symmetric. Once regulation breaks the aggressive separating equilibrium, the widespread nature of the shock tilts the pooling policy $m_{G,p}^*$ towards the first-best policy of the bad bank m_B^*. Since the bad bank's policy is not altered much by switching from the separating to the pooling equilibrium, the increase of the default probability and the loss of intrinsic value are limited for the bad bank.

As was the case in Section 2.4.3, the strategic behavior of the *smallest group* has the *biggest impact* on aggregate outcomes. The analysis in Section 2.4.1 already suggested that regulation shifts the distortion from the good bank to the bad bank by breaking the separating equilibrium. However, with this qualitative result in mind, we might expect that a larger fraction of the industry being exposed

[18] In our model, the initial regulatory response of the ECB to ask banks to refrain from paying dividends corresponds to the limiting case of $m^R \to \infty$, which leads to the restricted pooling equilibrium of Section 2.4.1. The analysis of this section is more relevant for the situation when the ECB moved to a recommendation of limiting dividends rather than one to not payout at all (see European Central Bank (2020)), or the (eventual) response of the Federal Reserve (see Federal Reserve System (2020)).

to the shock would render regulation sub-optimal as the effect on those bad banks is mechanically over-weighted on an industry level. Instead, we find that a sharp response of the minority drives the aggregate outcome in both scenarios discussed, reversing the logic of the more straightforward direct channel.

Alternative scenarios

Having described the most extreme scenarios of a large, concentrated shock and a small, widespread shock, we now discuss a few other scenarios without explicitly tabulating the results as in Figure 2.4. In the scenario where the scope and size of the shock are limited, i.e., $1-\alpha = 0.25$ and $f = 0.075$, banks are already in a pooling equilibrium in the absence of regulation. Payout restrictions lower default risk but also lower the value of the banking industry. With a medium jump size of $f = 0.125$, regulation becomes beneficial, like in the case of a small widespread shock, by reducing default risk and creating value. In this scenario, the direct effect that was described earlier dominates the adverse indirect effect of regulation. Combining these observations with that of Section 2.4.3, one see that when there is a small group of bad banks, payout restrictions lower stability and value when the shock is large ($f = 0.2$), fosters stability and creates value when the shock is medium-sized ($f = 0.125$), and increases stability but reduces value when the shock is small ($f = 0.075$). In the scenario where there are many bad banks and the shock is big, i.e., $1-\alpha = 0.75$ and $f = 0.2$, payout restrictions have a similar effect as in the scenario where the shock is big and there are few bad banks, i.e., $1-\alpha = 0.25$.

2.4.4 Short-termism/Investor liquidity concerns

In addition to the economic fundamentals of the liquidity shock, investor preferences can help predict the effectiveness of regulation. As discussed in the model set-up in Section 2.2.1, the parameter k reflects the importance of the possibility of selling shares on short notice to investors, whether it is because some investors have a short-term focus or because they are subject to liquidity concerns. In troubled times, investors might put more weight on the bank's market valuation, because a larger fraction of them have a short-term focus, or because their own liquidity concerns are stronger. For the latter case, investors might have to cover shortfalls in other parts of their portfolios, or want to liquidate their position in the bank to exploit new investment opportunities.

Based on the value of k, three regions of regulatory impact emerge, see Section 2.3.4 and Figure 2.3(d). For low values of k, banks separate without any distortion of the good bank's policy, and while regulatory intervention *might* lower the average default risk, it will surely be value-destroying by causing deviations from the first-best policies. For high values of k, banks are already in a pooling equilibrium, and regulatory intervention makes all banks safer. However, there tends to be an element of value-destruction when the regulator sets the threshold m^R above the optimal pooling strategy $m^*_{G,p}$. To fix ideas, consider the limiting case of $k = 1$, where the good bank optimizes market value, see Eq. (2.2). Since market value coincides with average bank value in the pooling equilibrium (see Section 2.4.2), any deviation from $m^*_{G,p}$ is suboptimal. Finally, in the intermediate region, banks play a separating equilibrium where the bad bank follows its first-best, but the good bank does not. At the upper end of this region, the good bank pays a relatively aggressive dividend policy, which corresponds to high default risk. Therefore, imposing restrictions is more likely to be beneficial for a value of k at the upper end rather than at the lower end of this region.

This observation applies to different scenarios for shock size and scope. While the boundary values of k for the different regions vary, the effect is the same, and for some cases, a large enough increment of k within the intermediate region has a strong enough effect to overturn a pessimistic outlook for regulation. This leads to the rough rule of thumb that regulation is more likely to be beneficial when investors are very focused on the (short-term) market value. A refinement of this guideline is possible if the regulator can identify the laissez-faire equilibrium; a higher k improves the outlook for regulation if it prevents aggressive distortion of the good bank's payout policy. The possibility of undoing this distortion is what enables regulation to not only lower average default risk, but even create value on an industry level.

2.4.5 Macro-prudential regulation

The analysis in Sections 2.4.3 and 2.4.4 shows that the presence of asymmetric information complicates the trade-off that regulators face in designing macro-prudential regulation. The intuitive trade-off between stability and bank value does not always apply. This caveat is particularly relevant around the threshold where restrictions start constraining the separating equilibrium. Rather than suggesting that regulators abstain from interventions, our findings emphasize the importance of a thorough assessment of the economic situation and provide support for allowing a certain level of discretion in applying restrictions. In particular, our analysis shows that it is not only the direct impact of the scope of the liquidity shock that is relevant, but that the indirect effect of strategic adjustments can be substantial as well. The new trade-off that information frictions induce is between signaling and mimicking distortions. These distortions tend to increase with the focus of investors on (short-term) market value. Suppose the build-up of systemic risk during economic upswings is accompanied by a shift in investor focus. In that case, mitigating distortions from asymmetric information could be an additional benefit of activating macro-prudential measures such as the counter-cyclical capital buffer.

2.4.6 Conservative regulation

In light of the risk of possible doubly adverse outcomes as exemplified by the scenario in Section 2.4.3, an approach that a regulator might want to take is to set the payout constraint m^R equal to or above the first-best payout threshold of a bad bank m_B^*. In this way, both good and bad banks are guaranteed not to be more likely to default than in the separating equilibrium, so that the average risk of default *must* be lower. It comes at the cost of lower overall industry value, but this is qualitatively in line with what would happen in a setting without asymmetric information. However, the heterogeneity introduces a value transfer from good to bad banks, since bad banks become more valuable when pooled with good banks at the cost of a reduction of the good bank's market valuation. This observation suggests that it becomes interesting to be a good bank in anticipation of such an intervention. An interesting extension of the model would be to endogenize the shares of good and bad banks in the economy. One way to do this is to allow a continuum of bank managers to incur a private cost to avoid tail risk exposure. We expect that fewer managers would find it beneficial to pay such a cost when regulation lowers the value of a good bank and increases the value of a bad bank. Such a mechanism could lead to the banking industry becoming riskier than the counterfactual.

2.5 Conclusion

This paper studies the effect of dividend restrictions on a bank's payout strategy and default risk in the presence of asymmetric information. We develop a continuous-time model of a bank whose exposure to a pending liquidity shock is private information to its management. To boost their short-term market valuation, the exposed banks have incentives to mimic the dividend policy of the unexposed banks. In response, the unaffected banks can signal their type by aggressively lowering their target cash level. Depending on the economic environment, this strategic interaction results in either a separating or a pooling equilibrium. Dividend restrictions imposed by the regulator have the potential to break the separating equilibrium, thereby decreasing the default risk of the unexposed banks but increasing the default risk of the exposed banks. The effect on the average bank depends on fundamental economic factors of the shock's scope and size, and on investors' focus on short-term market valuation. In the presence of asymmetric information, regulatory intervention has the potential to improve *both* average default risk and banking industry value. However, it comes with the pitfall of causing deterioration of both. A promising avenue for future research is the development of indicators to help regulators navigate this challenging environment and assess the right course of action.

A possible model extension would be the introduction of spillover effects and the resulting implications for the banks' strategic behavior. One way to do so is to assume that the good types are also exposed to the liquidity shock, albeit to a lesser extent. Alternatively, contagion effects could be created by assuming that a bank default affects other banks it is linked to through common assets and interbank market connections. Another direction would be the introduction of recurring shocks rather than the assumption in place of a single shock. Doing so would give more insights into how to set long-term dividend regulations. However, this requires capturing the learning dynamics of outsiders as they observe the bank's cash reserve. Apart from the added complexity, outsiders will eventually learn the bank's type with almost certainty. Absent shocks to bank types, the information asymmetry is resolved after a given period, just as it is in our model after the arrival of the shock.

Tables and Figures of Chapter 2

Table 2.1: Baseline parameter values.

	Notation	Value
Initial level cash reserves	m	1
Discount rate	ρ	0.035
Cash flow drift	μ	0.1
Cash flow volatility	σ	0.1
Shock arrival intensity	λ	0.2
Shock size	f	0.15
Fraction of good banks in economy	α	0.8
Fraction of short-term investors	k	0.5

Figure 2.1: Graphical representation of solution $\overline{m}^S$ to Eq. (2.3). The blue line displays the first-best value function of the bad bank. The red line shows the value of the bad bank for different values of cash level m^S when paying dividends at m^S and being valued by the market as a good bank. The 45° illustrates the tangent line of $V_{B,G}(m_S;m_S)$ at $m_S = 0$, having a slope of 1. Parameters values are according to Table 2.1.

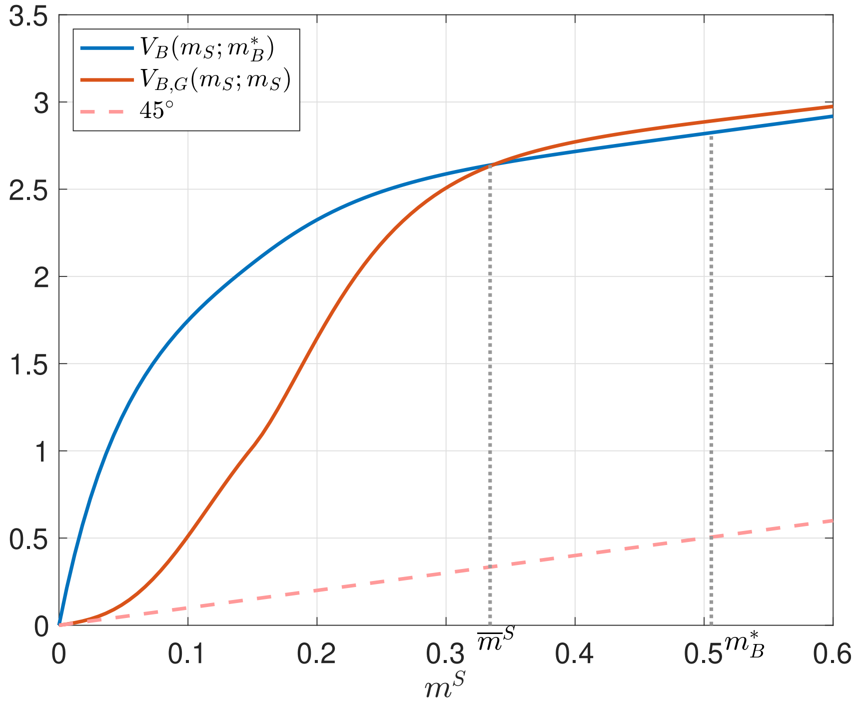


Figure 2.2: Optimal target cash level before the arrival of the shock for different values of liquidity shock f. Parameter values are according to Table 2.1.

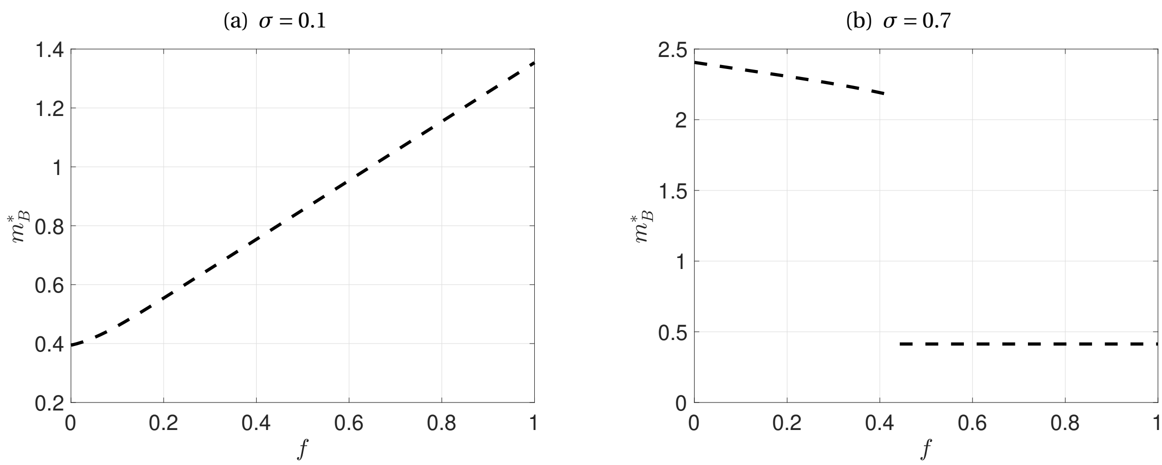


Figure 2.3: Dividend threshold and equilibrium selection. For different parameters, the graphs depict the dividend target (or equivalently, target cash level) in the least-cost equilibrium (green line), in the least-cost separating equilibrium (dashed), the first-best case of the good bank and the bad bank (the solid and dashed-dotted line, respectively), and the optimal target cash level of the good bank when pooled with the bad bank. The other parameters are according to Table 2.1.

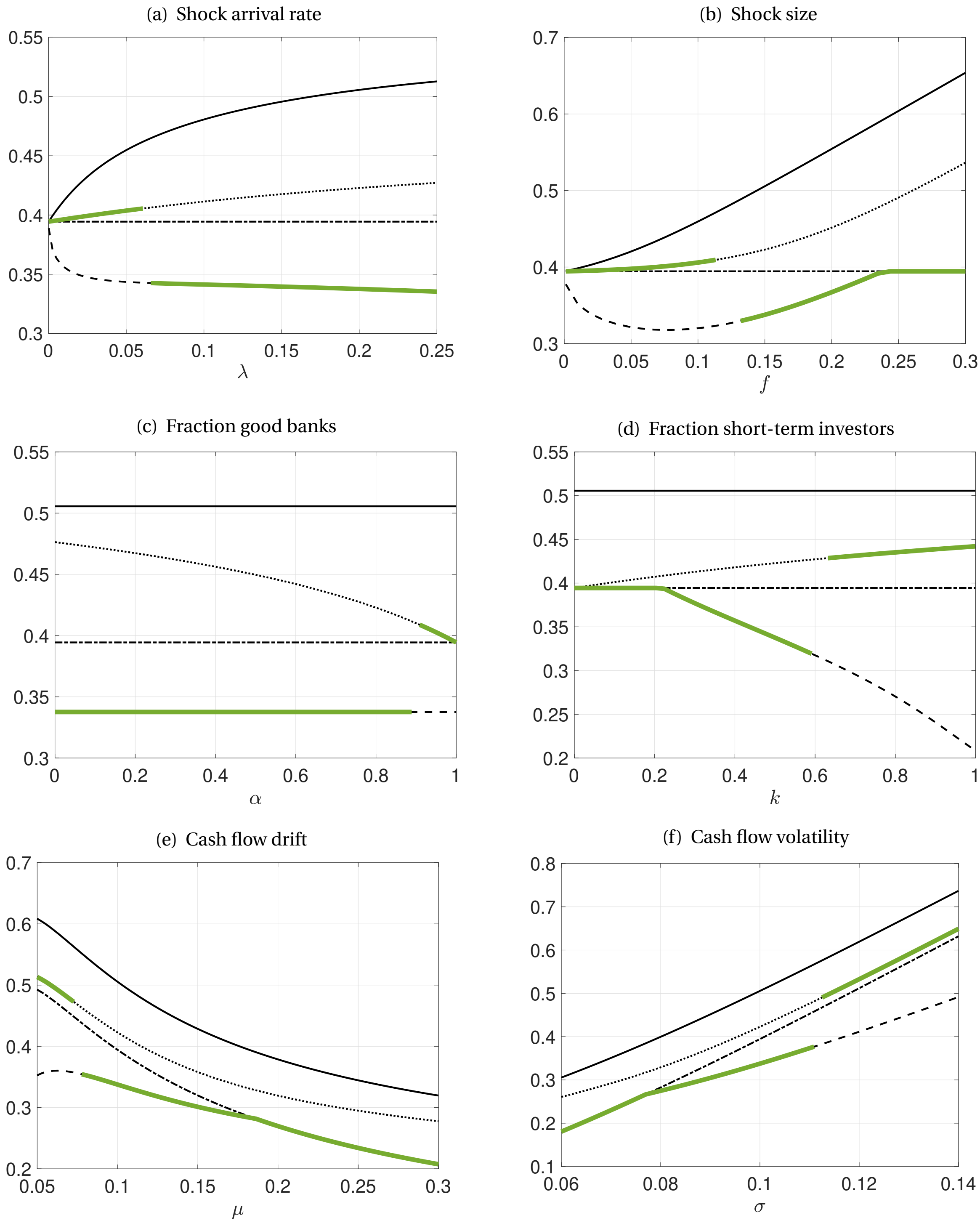
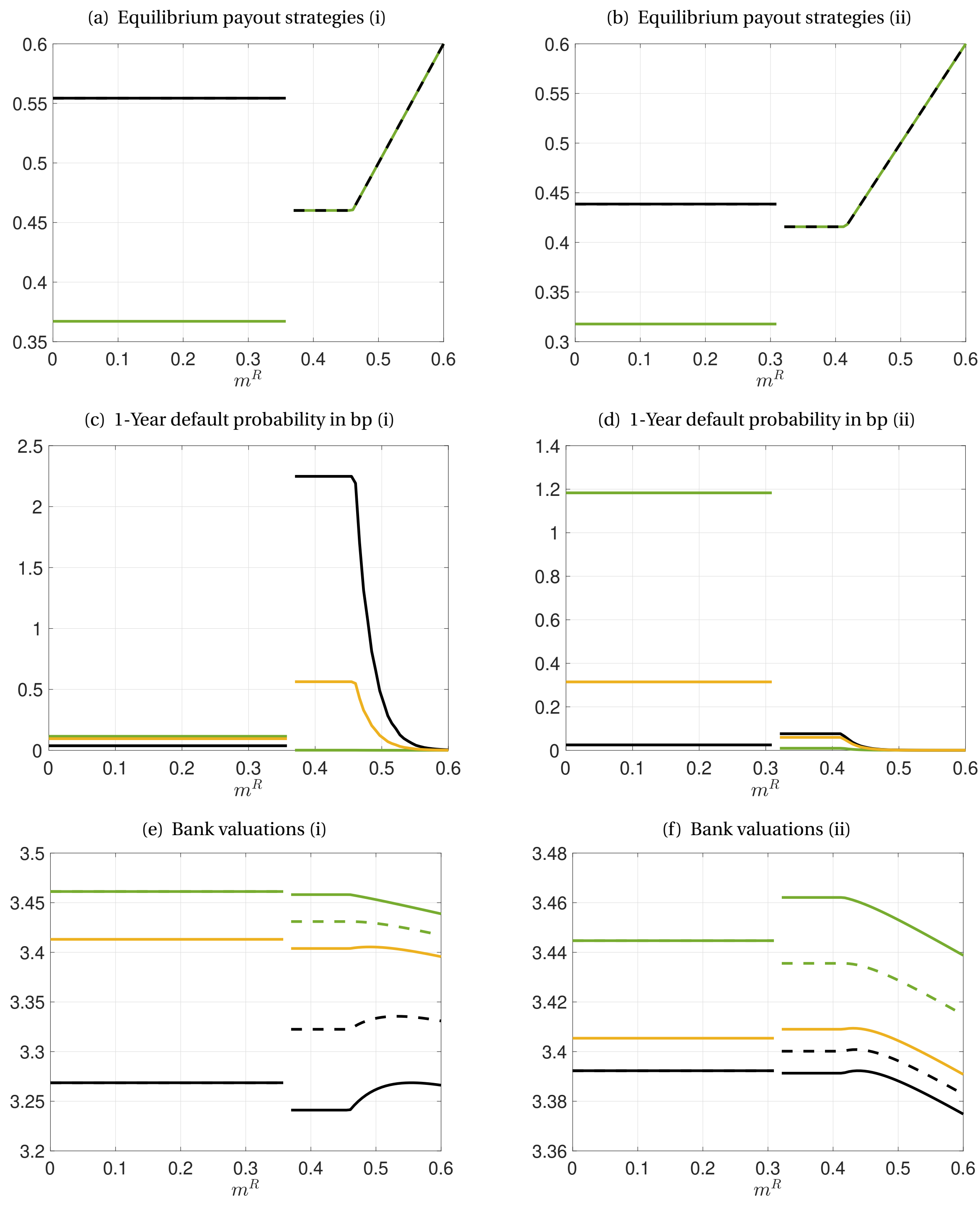


Figure 2.4: Dividend restrictions before shock. The parameters are based on Table 2.1, but parameter set (i) has $\alpha = 0.75$ and $f = 0.2$ (large concentrated shock), and parameter set (ii) has $\alpha = 0.25$ and $f = 0.075$ (small widespread shock). Figures (a) and (b) show the equilibrium pay out strategies of the good bank (green line) and bad bank (black line). Figures (c) and (d) show the 1-year default probability of the good bank (green), bad bank (black) and average bank (yellow). Figures (e) and (f) display for the good bank (green) and bad bank (black) the intrinsic value (solid) and combined intrinsic and market value (dashed), and the average value (yellow).



Social media and equilibrium sentiment

Social media has become a source of data on investor opinions but communication is itself an equilibrium outcome. Platforms are driven by interactions and this paper introduces these interactions to a model of trading under asymmetric information, which yields a communication strategy that does not exist with simple messaging. In posting on social media, an informed investor optimally balances information sharing and disagreement with noise posters. Relations between equilibrium sentiment and market data can generate patterns from the empirical literature of weak positive return predictability, price continuations, and information-driven disagreement that sentiment from an equivalent equilibrium with only noise posters fails to produce.

3.1 Introduction

The rise of social media has generated a treasure trove of unstructured data on human interactions. Improvements in natural language processing have enabled social scientists to mine this new field of data (Gentzkow, Kelly, and Taddy, 2019) and financial economists have used it to establish relations between sentiment, returns and price reversals (Chen, De, Hu, and Hwang, 2014; Bartov, Faurel, and Mohanram, 2018; Gu and Kurov, 2020), and disagreement and trading volume (Cookson and Niessner, 2020). However, as this data is available in real time it is presumably part of the investment process of the investors generating the market data analyzed by researchers and sentiment score vendors[1]. The observed market and communication outcomes jointly represent equilibrium behavior. For financial data this is particularly important because of the benefit of trading on private information. Communication should not contain such information in the first place, and if it did, its exploitation in real time should remove any predictability from the generated trading data. In equilibrium, we would expect a rational informed investor to stay silent. However, in this paper, I establish the existence of a noisy rational expectations equilibrium where such an agent optimally shares information and I derive implications for return predictability, price reversals, disagreement and trading volume in closed form. Notably, social media is not just a channel for this communication – the interactions that make it them *social* media is what makes the communication strategy optimal.

From earlier years' message boards (Yahoo! Finance, Raging Bull) to todays' micro-blogging and blogging platforms (Twitter, StockTwits, Reddit, Seeking Alpha, etc.) it is clear that there is an interest in sharing opinions about investing online. Theoretically, asymmetric information incentivizes communication by traders trading for non-informational reasons (Admanti and Pfleiderer, 1991). By communicating their trading intentions they help uninformed liquidity providers better filter out the noise created by this trading and make prices more informative, which lowers the cost of execution generated by asymmetric information. Whether or not such traders are deliberately employing

[1] Examples of companies active in this space are Dataminr, Ravenpack, but also traditional data providers likelike Bloomberg and Refinitiv (formerly Thomson Reuters Financial & Risk).

it as part of liquidity trading, or they are better understood as noise traders, this mechanism is active. It reduces the informational edge of investors with private information but does not affect the fundamental calculus that any (additional) information sharing is to their detriment. On social media though, communication takes place on an open channel and interactions are not only possible but promoted by algorithms tuned for engagement.

3.1.1 Two equilibria

Extending the asymmetric information workhorse model of Grossman and Stiglitz (1980) with the kind of interactive communication that characterizes social media, I show that the interactivity makes and informed investor (she/her) indifferent between the silence of the baseline model and a communication strategy with a uniquely optimal level of information sharing. The difference in equilibrium communication leads to a fundamental divergence in the properties of observed sentiment. A divergence that produce contrasting empirical predictions despite the equivalence between the two equilibria. These predictions are cleanly defined by the ability of sentiment to explain various market data such as returns, reversals, and turnover, because the equivalence between the two equilibria extends beyond the indifference of the informed investor to the equilibrium price. An analysis purely based on market data would not find a distinguish between these two equilibria, which highlights communication as the basis of the results.

When the informed investor communicates, equilibrium sentiment is correlated with the fundamental value of the traded asset, it's pay-off. The equivalence between the equilibria explains how this is possible. Given her private information, the informed investor's optimal communication strategy consists in designing the presentation of that information. This is characterized by two aspects of her posts: how clear they are, and whether to highlight agreement or lean into disagreement with other posters in her rhetorical style. The other posters in the model are noise traders and their communication makes the equilibrium price more informative to an uninformed investor (he/him) through the mechanism outlined above. For a given price, noise traders being bullish suggests that fundamental value is lower than otherwise. Due to the noise involved in communication this is not a definitive signal though. By choosing an information design that emphasizes disagreement, the informed investor provides the uninformed investor with another condition to observe. He can then reason as follows: for a given price and level of bullishness among noise traders, positive informed posts suggests that noise trader are even more eager to buy than otherwise because of their tendency to disagree with the informed investor in their communication. If this is the only purpose of informed posts, they unambiguously make prices more revealing as the uninformed investor switches his beliefs *away* from the sentiment they reflect. However, because the informed posts also reflect her private information, the uninformed investor simultaneously want to trade *along* with them. In balancing these two countervailing forces the uninformed investor can effectively end up ignoring the informed's communication, which leads to the equivalence with equilibrium where the informed investor does not communicate at all.

It turns out, that the level of clarity that achieves equivalence between the silence and communication, is a simple function decreasing in risk aversion and increasing in the strength of the informed investor's signal, the clarity of noise traders' posts, and how interactive the communication channel is. Meanwhile it is unaffected by the fundamental uncertainty of the pay-off and the level of noise trading. To summarize, due to the interactive nature of social media a communication strategy of disagreement is as attractive as one of complete silence to someone who possesses private information. Occam's razor might be on the side of the silent equilibrium, but the empirical evidence is not.

3.1.2 Related literature

A central question in the empirical literature, is whether sentiment can predict returns.[2] Since there is fundamental information in equilibrium sentiment when informed investor communicate, it is natural to associate this equilibrium with return predictability. However, to the extent that communication by noise traders is not fully revealing, or it is ignored by investors in real time, it can predict returns through its correlation with the price pressure created by their trading or trading by similar traders. In order for this predictability to be detectable empirically, the share of total turnover has to be large enough. In the model, there is return predictability in both equilibria. Actually, the theoretical t-statistic on sentiment is always greater in the price pressure equilibrium, and the sign of the relationship is unambiguously negative.

Early result on stock messaging boards were consistent with the noise trader view Antweiler and Frank (2004); Tumarkin, Management, and Whitelaw (2001) or perhaps that these message board are a side show Kim and Kim (2014). It is challenged in more recent studies extending the scope to blogs and micro-blogs (Chen et al., 2014; Leung and Ton, 2015; McGurk, Nowak, and Hall, 2020; Farrell, Green, Jame, and Markov, 2022), at least for stocks with less analyst coverage. Although there are large differences in social media between now and then, and the appropriate time-frame might be the very short term that has become possible to study with larger online activity (Renault, 2017; Checkley, Higón, and Alles, 2017) this result is also not un-challenged (Behrendt and Schmidt, 2018). A summary of the results in the literature are that they are mixed (Nardo, Petracco-Giudici, and Naltsidis, 2016) but with an overall trend towards finding evidence of a weak positive signal. While this is not definitive empirical evidence against the equilibrium without informed communication, the mixed communication equilibrium is a better fit.

Recent studies focusing specifically on the information content has additionally demonstrated a lack of price reversals (Bartov et al., 2018; Gu and Kurov, 2020; Eierle, Klamer, and Muck, 2022) inconsistent with an explanation relying exclusively on price pressure. In contrast, the mixed communication equilibrium can match the pattern found in these papers. A sufficient condition to rule out price reversals on average is that the uncertainty of the informed investor's private information is greater than the product of risk aversion, magnitude of informed signal noise, and level of noise trading. Since the variability of the pay-off only influences one side of this inequality, enough uncertainty about the fundamental value of the asset can ensure this relationship holds.

In comparison to the results on return predictability, the relationship between online activity and trading volume is robustly positive (Antweiler and Frank, 2004; Das and Chen, 2007; Sprenger, Tumasjan, Sandner, and Welpe, 2014b; Bandara, 2016; Giannini, Irvine, and Shu, 2019). In a recent study by Cookson and Niessner (2020) the authors further decomposes this link into it reliance on disagreement and information using data from the micro-blogging service StockTwits. They find that disagreement between investor that fundamentally agree is a stronger driver of trading volume than disagreement between groups. Based on the self-identification of posters, they find support for the notion that this disagreement is driven by differences in information sets. In the model, such disagreement is mirrored by the disagreement between the informed and uninformed investor. In the silent equilibrium, sentiment does not predict this disagreement because the information about noise trader bullishness contained in sentiment is fully incorporated into the expectation of the uninformed. In the mixed communication equilibrium, sentiment does predict this kind of disagreement. This is due to the correlation of sentiment with the private information of the informed investor, about which investors are differentially informed.

In addition to the empirical literature on using sentiment cited above, this paper is related to a mainly

[2] Sentiment here understood simply as the tone of a text usually quantified by some method of natural language processing Nadkarni, Ohno-Machado, and Chapman (2011) rather than the 'beliefs not justified by the facts at hand' terminology of Baker and Wurgler (2007).

theoretical literature on why people might listen to and spread stock market rumors (Van Bommel, 2003; Schmidt, 2020) and leak private information (Indjejikian, Lu, and Yang, 2014), or more generally why investor might volunteer information due to limits to arbitrage (Ljungqvist and Qian, 2016), short investment horizon (Liu, 2017), or differential quality of private information (Goldstein, Xiong, and Yang, 2020). However, these papers do not directly model institutional details of social media, on the contrary, they rely on communication not being publicly observable[3] or difficult to parse.

This paper is also related to the literature on information sharing and networks in asset pricing (Demarzo, Vayanos, and Zwiebel, 2003; Duffie, Malamud, and Manso, 2009; Colla and Mele, 2010; Ozsoylev and Walden, 2011b; Pedersen, 2021). Due to the complexity of embedding an information sharing structure (a network or a random search model) into tractable models this literature generally has to rely on the assumption that investors are sharing information without strategic considerations. The interaction captured in the model presented here, abstracts away the topology of the network, but in its reduced form maintains the emphasis on correlation, even in noise, as a powerful mechanic for driving outcomes (Colla and Mele, 2010; Ozsoylev and Walden, 2011b). In contrast to the papers that focuses on social networks and how they assign unduly importance to certain opinions over others (Demarzo et al., 2003; Pedersen, 2021), the focus on social media in this paper highlights how opinions become data and how it matters for what we can learn from that data.

Finally, a relevant strand of literature investigates the role of traditional media in affecting outcomes on the stock market and finds that it proxies for noise trader sentiment (Tetlock, 2007a), trails informed investors on news (Hendershott, Livdan, and Schürhoff, 2015), and that this form of communication is strategic as well even though journalists are not themselves investors (Goldman, Martel, and Schneemeier, 2022).

The paper is structured as follow. First the model is presented and equilibria derived in Section 3.2. Second, Section 3.3 contains he exposition of contrasting empirical predictions for sentiment explaining market data in the two equilibria and the baseline scenario of communication ignored at the time of trading. Finally, Section 3.5 concludes and provide perspective.

3.2 Model

The model extends the two-period model of Grossman-Stieglitz by adding posts and a communication decision stage where the informed investor commits to a communication strategy.

3.2.1 Baseline

In the last period ($t = 2$), investors consume their final wealth w_i. Investors have CARA utility with common risk aversion α, i.e. $U(w_i) = -e^{-\alpha w_i}$ for an investor i. In the penultimate period ($t = 1$), two types of investors, informed I and uninformed U, trade a risk-less asset and a risky asset with pay-off $y \sim \mathcal{N}(0,\sigma_y^2)$ in stochastic supply $z \sim \mathcal{N}(0,\sigma_z^2)$ provided by a noise trader Z. The risk-less asset is in elastic supply and its return is normalized to zero. The informed investor receive a noisy signal of the pay-off $s_I = y + \varepsilon$ where $\varepsilon \sim \mathcal{N}(0,\sigma_\varepsilon^2)$. Investor i optimizes their expected utility given their information $\mathscr{F}_i$ that at least contains the equilibrium price p. This optimization yields linear demand functions

$$D_i = \alpha^{-1} Var[y|\mathscr{F}_i]^{-1}(E[y|\mathscr{F}_i] - p) = \arg\max_{d_i} E[U(d_i\{y-p\})|\mathscr{F}_i]. \tag{3.1}$$

[3] Mathematically, the models are based on Kyle (1985) and they, largely, maintain the assumption that the market maker sets the price and clear the market purely based on net order flow.

Given her signal, the informed investor has superior information and her conditional expectation and variance, therefore, only depend on that signal and not price

$$E[y|s_I] = \sigma_y^2 Var[s_I]^{-1} s_I =: \beta_I s_I, \quad Var[y|s_I] = \sigma_y^2 \sigma_\varepsilon^2 Var[s_I]^{-1}.$$

Equilibrium price ensures market clearing $D_I + D_U = z$ and is given by

$$p = \lambda_p \beta_I \left(s_I - \alpha\sigma_\varepsilon^2 z\right) + (1-\lambda_p) E[y|\mathscr{F}_U], \qquad (3.2)$$
$$\text{where} \quad \lambda_p = \left(1 + \frac{Var[y|s_I]}{Var[y|\mathscr{F}_U]}\right)^{-1} \geq \frac{1}{2} \Longleftarrow Var[y|s_I] < Var[y|\mathscr{F}_U].$$

From the price, the uninformed can extract the signal

$$\zeta := \beta_I^{-1} \lambda_p^{-1} \left\{p - (1-\lambda_p) E[y|\mathscr{F}_U]\right\} = s_I - \alpha\sigma_\varepsilon^2 z.$$

In the absence of communication, the uninformed forms expectation

$$E[y|\zeta] = \sigma_y^2 \left\{Var[s_I] + \alpha^2 \sigma_\varepsilon^4 \sigma_z^2\right\}^{-1} \zeta$$

and price is linear in the price signal ζ with a scaling factor that is a convex combination of the informed excpectation coefficient and the uninformed

$$p = \left\{\lambda_p \beta_I + (1-\lambda_p)\beta_{\zeta 0}\right\} \zeta$$

with more weight on the informed coefficient.

3.2.2 Communication

I extend the baseline model to include communication by letting the informed investor and the noise trader communicate through posts observed by the uniformed investor in period $t = 1$ and requiring that the informed investor commits to a communication strategy[4] at time $t = 0$. The posts indicate in which direction the senders will trade, i.e., whether they are bullish or bearish. To formalize this, communication of the senders respectively provide the signals $c_I = s_I + \eta_I$ and $c_Z = -z + \eta_Z$. Since z is stochastic *supply*, its negation $-z$ is noise trader demand, which is the direction they want to trade, e.g. providing positive supply is to sell. The noise

$$\boldsymbol{\eta} = \begin{bmatrix}\eta_I & \eta_z\end{bmatrix}^\top \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_\eta = \begin{bmatrix}\sigma_{\eta_I}^2 & \rho_\eta \sigma_{\eta_I}\sigma_{\eta_Z} \\ \rho_\eta \sigma_{\eta_I}\sigma_{\eta_Z} & \sigma_{\eta_Z}^2\end{bmatrix}\right)$$

prevents messages from being fully revealing, and the interactive nature of social media is reflected in the correlation in the noise ρ_η.

Optimal communication

At the communication stage (time $t = 0$, i.e. before observing s_I), the informed trader chooses a communication strategy σ_{η_I} and sign(ρ_η) to maximize the certainty equivalent of her expected utility

[4] It is possible to extend the model to allow noise traders to choose communication strategy as well. However, it requires the introduction of a cost of or limit to communication to prevent fully revealing communication, see Appendix C.3.1. As this cost or limit would be exogenous, setting the level directly simply abstracts away a step and leaves the focus on the choice of the informed investor.

derived from trading (see Appendix C.1.2)

$$CE_I = \frac{1}{2\alpha}\ln\left(1 + \frac{Var[E[y|s_I] - p]}{Var[y|s_I]}\right).$$

Fixing the magnitude of the communication correlation, $|\rho_\eta|$, represents how the interactivity is a platform feature rather than a strategic decision.[5] In the context of social media communication the pre-commitment correspondence to establishing a presence online with a network of followers. However, without alterations to the set-up, a pre-commitment is not necessary, since the communication of the informed will be ignored in equilibrium and any deviation from her communication strategy, therefore, won't have an effect. Proposition 12 below is that, when the informed investor *has* to communicate, there is a unique interior solution to this problem in terms of noise even though an infinitely noisy strategy is possible. Furthermore, this strategy involves disagreeing with the other posters, the noise traders, on average even though their respective reasons trade are independent. One prediction of the model is, therefore, excess disagreement on social media in the this equilibrium. Proposition 13 is that in terms of price, this equilibrium is equivalent to another one where the informed does not communicate at all.

Communication effects

To understand what the optimal communication strategy achieves, it is useful to consider a number of the outcomes it influences. The difference between the informed investor's conditional expectation and the price is her expected profit per dollar she invests. It can be decomposed into disagreement with the uninformed investor and the compensated risk of the stochastic supply, i.e.,

$$E[y|s_I] - p = (1-\lambda_p)(E[y|s_I] - E[y|\boldsymbol{s}_U]) + \lambda_p \alpha Var[y|s_I]z. \tag{3.3}$$

The informational edge that leads to disagreement with the uninformed investor and more risky stochastic supply drives up the profit of the informed investor. Only the former is directly affected by the message send to the uninformed investor, but through the price-weight λ_p the compensation for the stochastic supply risk is affected as well.

As (3.3) is a convex combination, it is not immediately clear that maximizing the price weight λ_p is optimal. However, because the uninformed extract his signal from price, his conditional expectation covaries with stochastic supply and this means that in the variance of the difference in (3.3) the price-weight only interacts with stochastic supply risk (see Appendix C.1.5). Per (3.2), maximizing the price-weight requires minimizing the informativeness of the uninformed investor's signal vector $\boldsymbol{s}_U$

$$I[\boldsymbol{s}_U] = Var[y|\boldsymbol{s}_U]^{-1}.$$

Given this aspect of maximizing the variance of profit per dollar it is not surprising that the equilibrium of no communication in Proposition 13 exists. Rather it is its equivalence to, and indeed the existence of, the equilibrium in Proposition 12 that is the interesting result.

Finally, another way to view the variance of the expected profit per dollar invested in (3.3), is as the disagreement between the informed investor and the market. Larger disagreement allows her to take a larger position and since she will choose optimal demand D_I according to (3.1), by backward induction, it is optimal to maximize that position in expectation

$$E[|D_I|] = \sqrt{\frac{2}{\pi}}\frac{\sqrt{Var[E[y|s_I] - p]}}{\alpha Var[y|s_I]},$$

[5] In reality, the truth is likely somewhere in between and an analyse of the other extreme, where the informed investor can entirely control communication correlation would be a relevant extension.

where π is the mathematical constant. We are now ready to state the first proposition on optimal communication.

Proposition 12 (Optimal communication strategy)**.** *The unique solution to the informed investors problem of choosing the optimal communication strategy, when she has to communicate, is rhetorical disagreement with other posters, negative correlation between posts* $Corr(c_I, c_Z) = Corr(\eta_I, \eta_Z) = \text{sign}(\rho_\eta) < 0$*, and a finite amount of noise given by*

$$\sigma^*_{\eta_I} = -\frac{\alpha\sigma^2_\varepsilon\sigma_{\eta_z}}{\rho_\eta} = \arg\max_{\sigma_{\eta_I}} CE_I = \arg\max_{\sigma_{\eta_I}} Var[E[y|s_I] - p].$$

The solution maximizes the expected position of the informed by maximizing the disagreement between investors, minimizes the informativeness of the uninformed investor's signal vector, and maximizing the covariance of uninformed expectation and stochastic supply

$$\begin{aligned}\sigma^*_{\eta_I} &= \arg\max_{\sigma_{\eta_I}} E[|D_I|] \\ &= \arg\max_{\sigma_{\eta_I}} Var[E[y|s_I] - E[y|\boldsymbol{s}_U]] = \arg\min_{\sigma_{\eta_I}} I[\boldsymbol{s}_U] = \arg\max_{\sigma_{\eta_I}} Cov[E[y|\boldsymbol{s}_U], z].\end{aligned}$$

Proof. See Appendix C.1.1. □

A few things are worth noticing about this optimal communication strategy. First, it relies on negative correlation in posting noise. The type of interaction on social media that facilitates this optimal strategy is disagreement.

Second, the optimal communication strategy of Proposition 12 does not depend on the fundamental uncertainty of the pay-off σ^2_y or stochastic supply σ^2_z. Instead, larger risk aversion, a less precise informed signal, more noisy communication by noise traders lead to more noisy posting by the informed investor. One thing that pushes the informed investor to reveal more information is higher interactivity $|\rho_\eta| \uparrow$. Finally, a simple corollary is that it is never optimal for the informed investor to not obscure her information without this interaction between posts. From casual observation, strong interaction taking the form of disagreement is not a misleading characterization of of social media as a communication channel.

Corollary 12.1 (Social media as interaction)**.** *As the correlation in posting noise disappears, the informed investor's optimal communication strategy becomes infinitely noisy*

$$\lim_{\rho_\eta \nearrow 0} \sigma^*_{\eta_I} \to \infty.$$

To close of the section, I state a proposition on the equivalence between the equilibrium generated by the communication strategy of Proposition 12 and one where the informed investor does not communicate at all.

Proposition 13 (Equivalence of equilibria)**.** *If the informed investor can choose not to communicate, a second equilibrium exists in which the informed does not communicate. The equilibria are equivalent in the sense that uninformed beliefs are identical which lead to the same equilibrium price*

$$\begin{aligned}E[y|\boldsymbol{s}_U]|_{\sigma^*_{\eta_I}} = E[y|\boldsymbol{s}_U \setminus c_I] \textit{ and } Var[y|\boldsymbol{s}_U]|_{\sigma^*_{\eta_I}} = Var[y|\boldsymbol{s}_U \setminus c_I] \\ \implies p|_{\sigma^*_{\eta_I}} = p|_{\boldsymbol{s}_U \setminus c_I}.\end{aligned}$$

Proof. See Appendix C.1.4. □

It is important to notice that while Proposition 13 lead to the same outcomes in terms of market data, the properties of the sentiment generated by the two equilibria are very different as discussed next in Section 3.3. A cost of communication would favor the silent equilibrium whereas a benefit would favor the noisy disagreement equilibrium. While the two equilibria are equivalent in the model, the choice of silence could be considered more robust since it does not require getting the optimal level of information sharing right. As such, the evidence provided in Section 3.3 can be considered in favor of some benefit of communication. One possible benefit of communication generated only by its effect on followers is explored in Section 3.4 where the model is extended to include an uninformed investor who trades on aggregate sentiment, such as a market maker subscribed to a sentiment score.

3.3 Predictability of market data using sentiment

The sentiment available on social media is the collected communication of investors

$$c_S := c_I + c_Z,$$

which in the equilibrium where the informed investor is silent simplifies to the message from noise traders c_Z. Implicit in this definition is that sentiment is derived from the whole corpus of available posts.

Due to the equivalence of the two equilibria, price is the same under both forms of optimal communication, and it is driven by driven by four shocks: through the private information of the informed investor both the fundamental value and the signal noise enter price $s_I = y + \varepsilon$, the stochastic supply of noise traders z, and the communication noise of noise traders η_Z. For further analysis, the weights on those basic shocks can conveniently be written as

$$\begin{aligned} p &= \beta_I \gamma_{s_I}(y+\varepsilon) - \beta_I \gamma_z \alpha \sigma_\varepsilon^2 z + (1-\lambda_p)\beta_{c_Z}\eta_Z \\ \text{where } \gamma_{s_I} &= \frac{Var[s_I](\sigma_z^2 + \sigma_{\eta_Z}^2) + \lambda_p \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \sigma_{\eta_Z}^2}{Var[s_I](\sigma_z^2 + \sigma_{\eta_Z}^2) + \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \sigma_{\eta_Z}^2} \le 1, \\ \text{and } \gamma_z &= \frac{Var[s_I](\lambda_p \sigma_z^2 + \sigma_{\eta_Z}^2) + \lambda_p \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \sigma_{\eta_Z}^2}{Var[s_I](\sigma_z^2 + \sigma_{\eta_Z}^2) + \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \sigma_{\eta_Z}^2} \le \gamma_{s_I}. \end{aligned}$$

Dollar returns are given by then given by

$$r = y - p = (1 - \beta_I \gamma_{s_I}) y - \beta_I \gamma_{s_I} \varepsilon + \beta_I \gamma_z \alpha \sigma_\varepsilon^2 z - (1-\lambda_p)\beta_{c_Z}\eta_Z, \tag{3.4}$$

where $\beta_I \le 1$ and $\beta_{c_Z} \le 0$. Other things equal, a positive realization of pay-off y, stochastic supply z, or communication noise η_Z yields a positive return, whereas the return for a positive realizations of noise in the informed signal ε is negative.[6]

[6] While it is outside the scope of this paper, one could take the presence of η_Z in returns in (3.4) as the starting point for a discussion of the emerging literature on social media sentiment as a risk premium (Houlihan and Creamer, 2017; Hosseini, Jostova, Philipov, and Savickas, 2020).

3.3.1 Return predictability

Both equilibria generates return predictability. T-stats for projections (regressions) of returns on sentiment in the two equilibria are

$$t\{E[r|c_Z]\} = -\frac{\alpha\sigma_z^2 Var[y|s_I]\lambda_p}{\sqrt{Var[r]Var[c_Z] - Cov[r,c_Z]^2}} \leq 0, \text{ and}$$
$$t\{E[r|c_S]\} = \frac{(1-\gamma_{s_I})\sigma_y^2 - \alpha\sigma_z^2 Var[y|s_I]\lambda_p}{\sqrt{Var[r]Var[c_S] - Cov[r,c_S]^2}},$$

where derivations can be found in Appendix C.2.1. From the t-stats it is clear that the equilibrium with only noise trader communication only can produce a negative relation between sentiment and returns. By comparing the the t-stats, it is possible to show that the predictability is stronger in this equilibrium as well.

Proposition 14 (Predicting returns). *Sentiment without informed communication always produces a lower predictability than sentiment based on mixed communication*

$$|t\{E[r|c_Z]\}| > |t\{E[r|c_S]\}|.$$

Proof. See Appendix C.2.1. □

Taken together, sentiment of the equilibrium without informed communication predicts a negative relation that would be easier to detect than sentiment of the equilibrium with informed communication. This could be important for the strategic consideration of posting on social media in the presence of agents (companies) parsing and distributing in real time to some investors. In the context of the model, this would mean introducing another uninformed investor who relies solely on c_S, which would break the equivalence of the two equilibria, but this extension is left for future work.

3.3.2 Contemporary returns

Extending projections to explain contemporary dollar return,

$$r_0 = p - E[y] = p,$$

the expectation for contemporary bullish sentiment in the equilibrium without informed communication is positive (see Appendix C.2.1)

$$E[r_0|c_Z] = \frac{Cov[p,c_Z]}{Var[c_Z]}c_Z = \frac{\lambda_p\alpha\sigma_z^2 Var[y|s_I]}{Var[c_Z]}c_Z, \text{ where } \lambda_p\alpha\sigma_z^2 Var[y|s_I] \geq 0.$$

For sentiment of communication where the informed investor is active, expectation of contemporary returns are also positive

$$E[r_0|c_s] = \frac{Cov[p,c_I] + Cov[p,c_Z]}{Var[c_s]} = \frac{\beta_I\gamma_{s_I} Var[s_I] + \lambda_p\alpha\sigma_z^2 Var[y|s_I]}{Var[c_s]} \geq 0.$$

This simple result on positivity of contemporary returns holds in a scenario where sentiment represent noise trader bullishness that is ignored at the time of trading with

$$E[r_0|c_Z]|_{s_U=\zeta} = E[p|c_Z]|_{s_U=\zeta} \propto Cov[p|c_Z]|_{s_U=\zeta} \times c_Z$$
$$\text{where } Cov[p|c_Z]|_{s_U=\zeta} = \left\{\lambda_p\beta_I + (1-\lambda_p)\beta_{\zeta 0}\right\} \geq 0,$$

using price from Section 3.2.1 without communication.

Proposition 15 (Explaining contemporary returns)**.** *A positive relationship between sentiment and contemporary returns does not help to distinguish between alternative composition of returns or sentiment.*

Proof. See main text. □

Proposition 15 underlines the importance of how returns are calculated in empirical studies in terms of timing. Conceptually this is more thorny than not mixing up contemporary and future returns as it requires taking a stance on when price pressure is resolved and/or a pay-off can reasonably be assumed to be realized.

3.3.3 Price reversal

For noise trader only sentiment, the expected sign of contemporary return is the oppositie of predicted return, so stronger sentiment (negative or positive) predicts stronger price reversal other things equal, formally

$$\frac{\partial E[r|c_Z]E[r_0|c_Z]}{\partial |c_Z|} = -2\frac{Cov[p,c_Z]^2}{Var[c_Z]}|c_Z| \leq 0.$$

This is in line with the typical logic that a shock to price that is not related to fundamentals is reversed.

For the sentiment generated by the mixed communication equilibrium, the relationship with contemporary returns is also positive, (see Section 3.3.2) so for predicted returns to have the same sign, the requirement is that this relationship is positive as well.

Proposition 16 (Price reversal)**.** *Stronger sentiment always predicts stronger price reversal in the equilibrium with only noise trader communication. In contrast, for mixed communication sentiment, both reversal and continuation is possible.*

Necessary conditions for mixed sentiment to predict price continuation is that communication of noise traders is not perfectly revealing $\sigma_{\eta_Z} > 0$, and high private information uncertainty versus low riskiness of stochastic supply relative to risk tolerance

$$\sigma_\varepsilon^2 > \frac{1}{\alpha} > \sigma_z^2.$$

Proof. See main text and Appendix C.2.1. □

Further intuition about Proposition 16 can be gleaned from considering how the two sides of the inequality arises as limiting cases for the coefficient on sentiment for predicting returns:

$$\lim_{\sigma_z \to 0} \frac{\partial E[r|c_S]}{\partial c_S} > 0 \iff \sigma_\varepsilon^2 > \frac{1}{\alpha}, \text{ and} \tag{3.5}$$
$$\lim_{\sigma_\varepsilon \to \infty} \frac{\partial E[r|c_S]}{\partial c_S} > 0 \iff \frac{1}{\alpha} > \sigma_z^2.$$

The requirement in (3.5) can be related to the high difficulty of predicting stock returns. It is more reasonable to hold in a situation where the informed investor is not an insider trading on privileged information, but a rather a well-informed institutional investor. With this in mind, the second requirement that the level of noise trading is sufficiently low compared to risk tolerance, can be seen understood from the following perspective. The model does not accommodate varying the size (dollars invested) of the informed trader relative to noise traders explicitly, so the level of noise trading σ_z^2 has to account for this aspect and can be interpreted as a relative level.

3.3.4 Disagreement

According to Cookson and Niessner (2020), based on their measure of sentiment derived from Stock-Twits data, disagreement between investor who fundamentally agree but have different information sets is the most important for driving trading volume. In the model presented here this corresponds to the disagreement between the informed and the uninformed investor. This disagreement is indeed a component of both the informed and uninformed investors demand, see (3.3) in Section 3.2.2 for the example of the informed investor.

Proposition 17 (Sentiment predicting disagreement)**.** *Sentiment only predicts disagreement between informed and uninformed investors in the equilibrium where the informed investor communicates.*

The sentiment produced by the equilibrium where the informed investor does not communication does not predict disagreement between informed and uninformed investors in contrast to the equilibrium where she does.

Proof. Noise trader only communication yields

$$\begin{aligned} E[E[y|s_I]-E[y|\mathbf{s}_U]|c_Z] &\propto -Cov[E[y|\mathbf{s}_U],c_Z]c_Z \\ &= (\beta_\zeta \alpha\sigma_\varepsilon^2+\beta_{c_Z})Cov[z,c_Z]-\beta_{m_Z}Cov[\eta_Z,c_Z] \\ &\propto (\alpha\sigma_\varepsilon^2\sigma_{\eta_Z}^2\sigma_z^2-\alpha\sigma_\varepsilon^2\sigma_z^2\sigma_{\eta_Z}^2)c_Z=0 \end{aligned}$$

whereas the expectation for mixed communication is

$$\begin{aligned} E[E[y|s_I]-E[y|\mathbf{s}_U]|c_s] &= \frac{(\beta_I-\beta_\zeta)Cov[s_I,c_I]-Cov[E[y|\mathbf{s}_U],c_Z]}{Var[c_s]}c_s \\ &= \sigma_y^2\frac{\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2}{Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)+\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2}\frac{1}{Var[c_s]}c_s \end{aligned}$$

based on expressions that can be found in Section 3.2.1 and Appendix C.1.4. □

3.4 Extension: trading on aggregate sentiment

This extension of the model introduces an actor relying on aggregate sentiment in their trading decision, i.e., market makers subscribing to sentiment scores. This additional uninformed investor is signified by the subscript U_2 whereas the subscript for the uninformed investor of the model in the main text will be U_1 and the assumption of common risk aversion α maintained. Since the information set of the uninformed investor who observes posts separately is more granular than that of the uninformed investor trading on aggregate sentiment, the former can determine the demand of the latter and extract the same signal from the price as in Section 3.2.1 here relabelled ζ_1. The price signal of the investor trading on aggregate sentiment can be decomposed into this signal ζ_1 and an additional term related to the inference of uninformed investor U_1

$$\zeta_2 := p\left\{\frac{1}{Var[y|s_I]}+\frac{1}{Var[y|\mathbf{s}_{U_1}]}\right\}\sigma_\varepsilon^2-\alpha\sigma_\varepsilon^2 D_{U_2}=\zeta_1+\sigma_\varepsilon^2\frac{E[y|\mathbf{s}_{U_1}]}{Var[y|\mathbf{s}_{U_1}]}.$$

The signal vector of uninformed investor U_2 is therefore $\boldsymbol{s}_{U_2} = \begin{pmatrix}\zeta_2 & c_S\end{pmatrix}^\top$ yielding the conditional mean and variance

$$E[y|\boldsymbol{s}_{U_2}] = \beta_{\zeta_2}\zeta_2 + \beta_{c_S}c_S, \quad Var[y|\boldsymbol{s}_{U_2}] = \sigma_y^2 - \beta_{\zeta_2}Cov[y,\zeta_2] - \beta_{c_S}Cov[y,c_S],$$

with more explicit expressions in Appendix C.4.1.

Denoting the price of the main text by p_1 and the price weight λ_{p_1}, market clearing with the additional investor yields price p_2 (see Appendix C.4.1 for derivations)

$$p_2 = \lambda_{p_2}p_1 + (1-\lambda_{p_2})E[y|\boldsymbol{s}_{U_2}], \quad \text{where } \lambda_{p_2} = \frac{\frac{1}{Var[y|s_I]} + \frac{1}{Var[y|\boldsymbol{s}_{U_1}]}}{\frac{1}{Var[y|s_I]} + \frac{1}{Var[y|\boldsymbol{s}_{U_1}]} + \frac{1}{Var[y|\boldsymbol{s}_{U_2}]}}.$$

As established in the main text, maximizing the certainty equivalent of the informed investor with respect to communication strategy corresponds to maximizing the variance of the dollar profit margin. In the extension the dollar profit margin is given by

$$E[y|s_I] - p_2 = \lambda_{p_2}(E[y|s_I] - p_1) + (1-\lambda_{p_2})(E[y|s_I] - E[y|\boldsymbol{s}_{U_2}]),$$

i.e., a convex combination of the dollar profit margin of the main text and the disagreement between the informed investor and the uninformed investor trading on aggregate sentiment. The variance of this expression is given by

$$\begin{aligned} Var\left[E[y|s_I] - p_2\right] = \lambda_{p_2}^2 Var\left[E[y|s_I] - p_1\right] + (1-\lambda_{p_2})^2 Var\left[E[y|s_I] - E[y|\boldsymbol{s}_{U_2}]\right] \\ + 2\lambda_{p_2}(1-\lambda_{p_2})Cov\left[E[y|s_I] - p_1, E[y|s_I] - E[y|\boldsymbol{s}_{U_2}]\right], \end{aligned}$$

and more explicit expressions for the individual terms can be found in Appendix C.1.5 and Appendix C.4.1.

In contrast, if the informed investor does not communicate, the information sets of the uninformed investors are the same and price is given by

$$p_{2,s_U\backslash c_I} = \lambda_{p_2,s_U\backslash c_I}\beta_I\zeta_1 + (1-\lambda_{p_2,s_U\backslash c_I})E[y|\zeta_1,c_Z], \text{ where } \lambda_{p_2,s_U\backslash c_I} = \left(1 + \frac{2Var[y|s_I]}{Var[y|\zeta_1,c_Z]}\right)^{-1},$$

and the variance of dollar profit margin is formally the same as in the no communication equilibrium in the main text but with updated price weight $\lambda_{p_2,s_U\backslash c_I}$.

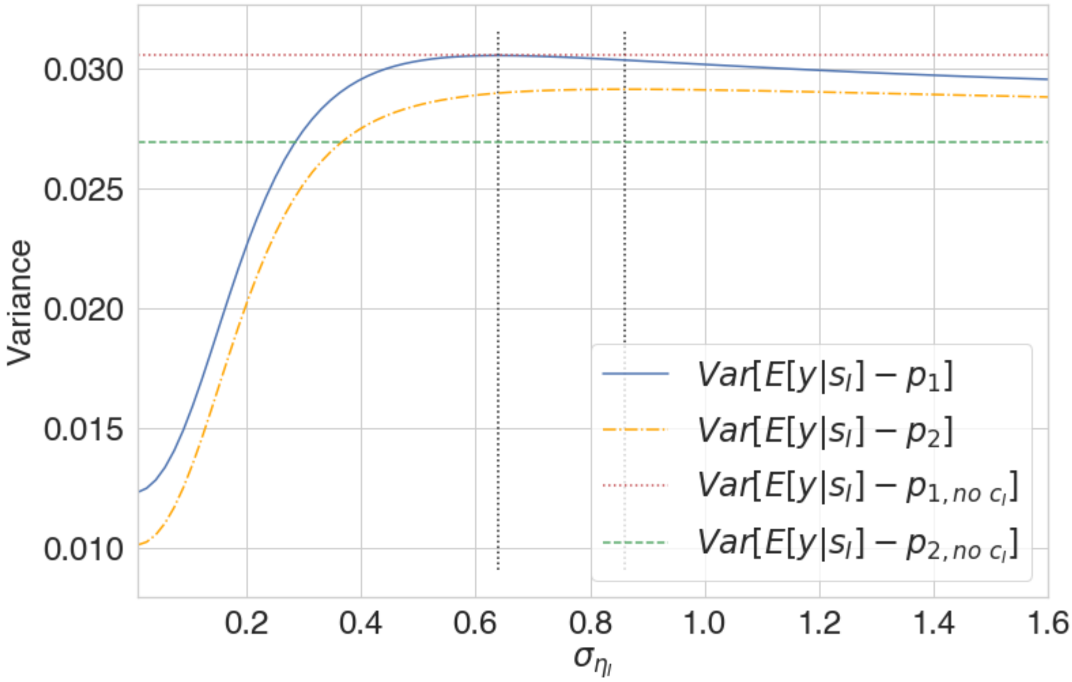
Figure 3.1 shows how the variance of dollar profit margin of the main text is the same for the silent and noisy disagreement equilibrium at the optimal level of communication noise $\sigma^*_{\eta_I}$ of the informed investor. With the extension to include an additional uninformed investor the equivalence is broken and, for the case presented, the optimal level of communication noise yields a higher variance of dollar profit margin. This means a higher certainty equivalent for the informed investor, i.e., noisy communication is the only optimal choice. Finally, the figure demonstrates the tendency of the equilibrium in the extended model to induce a larger amount of communication noise.

3.5 Conclusion and perspective

Due to the interactive nature of social media, an informed investor is indifferent between a communication strategy of information sharing and disagreement and refraining from any communication at all. The empirical evidence in the extant literature on sentiment derived from social media is in favor of the former. The model highlights the importance of interactions for understanding strategic communication and demonstrates that social media sentiment containing fundamental information can exist in equilibrium. Future directions for research include exploring further the extension that introduces an actor relying on aggregate sentiment in their trading decision, i.e., market makers subscribing to sentiment scores, and empirical work on disentangling communication noise from fundamental uncertainty and disagreement.

Figure of Chapter 3

Figure 3.1: Variance of dollar profit margin for parameters $\alpha = 1$, $\sigma_y = 1.2$, $\sigma_\varepsilon = 0.8$, $\sigma_z = 0.5$, $\sigma_{\eta_Z} = 0.5$, $\rho_\eta = -0.5$. Black dotted lines indicate optimal informed communication noise, $\sigma^*_{\eta_I}$, first for p_1 then p_2.



Appendix to Chapter 1

A.1 Inference problem

A.1.1 Variance of dot-product of independent random vectors

For random vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ with mean and variance $\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x \; \forall x \in \{v, w\}$ using the trace operator tr

$$\begin{aligned}
Var\left[\boldsymbol{v}^\top \boldsymbol{w}\right] &= E\left[\left(\boldsymbol{v}^\top \boldsymbol{w}\right)^2\right] - \left(E\left[\boldsymbol{v}^\top \boldsymbol{w}\right]\right)^2 = E\left[\boldsymbol{v}^\top \boldsymbol{w}\boldsymbol{w}^\top \boldsymbol{v}\right] - \left(E\left[\boldsymbol{v}\right]^\top E\left[\boldsymbol{w}\right]\right)^2 \\
&= E\left[\mathrm{tr}\left(\boldsymbol{v}\boldsymbol{v}^\top \boldsymbol{w}\boldsymbol{w}^\top\right)\right] - \left(\boldsymbol{\mu}_v^\top \boldsymbol{\mu}_w\right)^2 = \mathrm{tr}\left(E\left[\boldsymbol{v}\boldsymbol{v}^\top\right] E\left[\boldsymbol{w}\boldsymbol{w}^\top\right]\right) - \left(\boldsymbol{\mu}_v^\top \boldsymbol{\mu}_w\right)^2 \\
&= \mathrm{tr}\left(\left\{\boldsymbol{\mu}_v\boldsymbol{\mu}_v^\top + \boldsymbol{\Sigma}_v\right\}\left\{\boldsymbol{\mu}_w\boldsymbol{\mu}_w^\top + \boldsymbol{\Sigma}_w\right\}\right) - \left(\boldsymbol{\mu}_v^\top \boldsymbol{\mu}_w\right)^2 \\
&= \mathrm{tr}\left(\boldsymbol{\mu}_v\boldsymbol{\mu}_v^\top\boldsymbol{\mu}_w\boldsymbol{\mu}_w^\top\right) + \mathrm{tr}\left(\boldsymbol{\mu}_v\boldsymbol{\mu}_v^\top\boldsymbol{\Sigma}_w\right) + \mathrm{tr}\left(\boldsymbol{\Sigma}_v\boldsymbol{\mu}_w\boldsymbol{\mu}_w^\top\right) + \mathrm{tr}\left(\boldsymbol{\Sigma}_w\boldsymbol{\Sigma}_v\right) - \left(\boldsymbol{\mu}_v^\top \boldsymbol{\mu}_w\right)^2 \\
&= \left(\boldsymbol{\mu}_v^\top \boldsymbol{\mu}_w\right)^2 + \mathrm{tr}\left(\boldsymbol{\mu}_w\boldsymbol{\mu}_w^\top\boldsymbol{\Sigma}_v\right) + \mathrm{tr}\left(\boldsymbol{\Sigma}_w\boldsymbol{\mu}_v\boldsymbol{\mu}_v^\top\right) + \mathrm{tr}\left(\boldsymbol{\Sigma}_w\boldsymbol{\Sigma}_v\right) - \left(\boldsymbol{\mu}_v^\top \boldsymbol{\mu}_w\right)^2 \\
&= \boldsymbol{\mu}_v^\top\boldsymbol{\Sigma}_w\boldsymbol{\mu}_v + \boldsymbol{\mu}_w^\top\boldsymbol{\Sigma}_v\boldsymbol{\mu}_w + \mathrm{tr}\left(\boldsymbol{\Sigma}_w\boldsymbol{\Sigma}_v\right).
\end{aligned}$$

Using the Hadamard product identities $\boldsymbol{v}^\top(\boldsymbol{A}\odot\boldsymbol{B})\boldsymbol{w} = \mathrm{tr}\left(\boldsymbol{D}_v\boldsymbol{A}\boldsymbol{D}_w\boldsymbol{B}^\top\right)$ and $\boldsymbol{D}_v^\top\boldsymbol{A}\boldsymbol{D}_w = \boldsymbol{w}\boldsymbol{v}^\top\odot\boldsymbol{A}$, the covariance matrix identity $\boldsymbol{\Sigma}_w = \boldsymbol{D}_{\sigma_w}\boldsymbol{R}_w\boldsymbol{D}_{\sigma_w}$, and the interchangability of vectors in vector-diagonal matrix products $\boldsymbol{\mu}_v^\top\boldsymbol{D}_{\sigma_w} = \boldsymbol{\sigma}_w^\top\boldsymbol{D}_{\mu_v}$, the variance can be written as

$$\begin{aligned}
Var\left[\boldsymbol{v}^\top \boldsymbol{w}\right] &= \boldsymbol{\sigma}_w^\top\boldsymbol{D}_{\mu_v}\boldsymbol{R}_w\boldsymbol{D}_{\mu_v}\boldsymbol{\sigma}_w + \boldsymbol{\mu}_w^\top\boldsymbol{\Sigma}_v\boldsymbol{\mu}_w + \boldsymbol{\sigma}_w^\top(\boldsymbol{R}_w\odot\boldsymbol{\Sigma}_v)\boldsymbol{\sigma}_w \\
&= \boldsymbol{\sigma}_w^\top(\boldsymbol{R}_w\odot\boldsymbol{\mu}_v\boldsymbol{\mu}_v^\top)\boldsymbol{\sigma}_w + \boldsymbol{\sigma}_w^\top(\boldsymbol{R}_w\odot\boldsymbol{\Sigma}_v)\boldsymbol{\sigma}_w + \boldsymbol{\mu}_w^\top\boldsymbol{\Sigma}_v\boldsymbol{\mu}_w \\
&= \boldsymbol{\sigma}_w^\top(\boldsymbol{R}_w\odot\boldsymbol{\Omega}_v)\boldsymbol{\sigma}_w + \boldsymbol{\mu}_w^\top\boldsymbol{\Sigma}_v\boldsymbol{\mu}_w,
\end{aligned}$$

where $\boldsymbol{\Omega}_v = \boldsymbol{\mu}_v\boldsymbol{\mu}_v^\top + \boldsymbol{\Sigma}_v$.

A.1.2 Bias-variance trade-off solution

Linear-affine case

Substituting vectors of bias and volatility as functions of controls $\boldsymbol{c}$ given by Assumption 5 into the objective in (1.2) yields

$$\begin{aligned}
\Theta &= k_\varepsilon^2\boldsymbol{c}^\top\boldsymbol{\Omega}_\zeta\boldsymbol{c} + k_\sigma^2\boldsymbol{c}^\top\boldsymbol{D}_{\Omega_\zeta}\boldsymbol{c} + 2k_{\sigma0}k_\sigma\boldsymbol{1}^\top\boldsymbol{D}_{\Omega_\zeta}\boldsymbol{c} + k_{\sigma0}^2\boldsymbol{1}^\top\boldsymbol{D}_{\Omega_\zeta}\boldsymbol{1} \\
&= k_\sigma^2\boldsymbol{c}^\top(k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta})\boldsymbol{c} + 2k_{\sigma0}k_\sigma\boldsymbol{1}^\top\boldsymbol{D}_{\Omega_\zeta}\boldsymbol{c} + k_{\sigma0}^2\boldsymbol{1}^\top\boldsymbol{D}_{\Omega_\zeta}\boldsymbol{1}.
\end{aligned}$$

Hessian matrix is positive definite if $\boldsymbol{\Omega}_\zeta$ has full rank and given by

$$\frac{\partial\Theta}{\partial \boldsymbol{c}^2} = k_\varepsilon^2\boldsymbol{\Omega}_\zeta + k_\sigma^2\boldsymbol{D}_{\Omega_\zeta}.$$

Optimal controls from first order condition are

$$\begin{aligned}\boldsymbol{0} = \frac{\partial\Theta}{\partial \boldsymbol{c}} &= 2k_\sigma^2\boldsymbol{c}^\top(k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta}) + 2k_{\sigma 0}k_\sigma\boldsymbol{1}^\top\boldsymbol{D}_{\Omega_\zeta}\\ \iff \boldsymbol{c}^* &= -k_\sigma^{-1}k_{\sigma 0}\left(k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta}\right)^{-1}\boldsymbol{D}_{\Omega_\zeta}\boldsymbol{1},\end{aligned}$$

and substituting back into the objective it simplifies to

$$\begin{aligned}\chi = \Theta|_{\boldsymbol{c}=\boldsymbol{c}^*} &= k_{\sigma 0}^2\boldsymbol{1}^\top\boldsymbol{D}_{\Omega_\zeta}\left(k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta}\right)^{-1}(k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta})\left(k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta}\right)^{-1}\boldsymbol{D}_{\Omega_\zeta}\boldsymbol{1}\\ &\quad - 2k_{\sigma 0}^2\boldsymbol{1}^\top\boldsymbol{D}_{\Omega_\zeta}\left(k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta}\right)^{-1}\boldsymbol{D}_{\Omega_\zeta}\boldsymbol{1} + k_{\sigma 0}^2\boldsymbol{1}^\top\boldsymbol{D}_{\Omega_\zeta}\boldsymbol{1}\\ &= -k_{\sigma 0}^2\boldsymbol{1}^\top\boldsymbol{D}_{\Omega_\zeta}\left(k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta}\right)^{-1}\boldsymbol{D}_{\Omega_\zeta}\boldsymbol{1} + k_{\sigma 0}^2\boldsymbol{1}^\top\boldsymbol{D}_{\Omega_\zeta}\boldsymbol{1}\\ &= -k_{\sigma 0}^2\boldsymbol{1}^\top\left\{\boldsymbol{D}_{\Omega_\zeta} - \left(k_c^2\boldsymbol{\Omega}_\zeta^{-1} + \boldsymbol{D}_{\Omega_\zeta}^{-1}\right)^{-1}\right\}\boldsymbol{1} + k_{\sigma 0}^2\boldsymbol{1}^\top\boldsymbol{D}_{\Omega_\zeta}\boldsymbol{1}\\ &= k_{\sigma 0}^2\boldsymbol{1}^\top\left\{k_c^2\boldsymbol{\Omega}_\zeta^{-1} + \boldsymbol{D}_{\Omega_\zeta}^{-1}\right\}^{-1}\boldsymbol{1}.\end{aligned}$$

Letting $\boldsymbol{X} := k_c^2\boldsymbol{\Omega}_\zeta^{-1} + \boldsymbol{D}_{\Omega_\zeta}^{-1}$, optimal controls and cost of complexity are

$$\boldsymbol{c}^* = -k_\sigma^{-1}k_{\sigma 0}\left\{\boldsymbol{I} - \boldsymbol{D}_{\Omega_\zeta}^{-1}\boldsymbol{X}^{-1}\right\}\boldsymbol{1}, \quad \chi = \Theta|_{\boldsymbol{c}=\boldsymbol{c}^*} = k_{\sigma 0}^2\boldsymbol{1}^\top\boldsymbol{X}^{-1}\boldsymbol{1},$$

and bias and volatility

$$\begin{aligned}\boldsymbol{\varepsilon}_\beta|_{\boldsymbol{c}=\boldsymbol{c}^*} &= k_\varepsilon\left(-k_\sigma^{-1}k_{\sigma 0}\left\{\boldsymbol{I} - \boldsymbol{D}_{\Omega_\zeta}^{-1}\boldsymbol{X}^{-1}\right\}\boldsymbol{1}\right) = -k_c^{-1}k_{\sigma 0}\left\{\boldsymbol{I} - \boldsymbol{D}_{\Omega_\zeta}^{-1}\boldsymbol{X}^{-1}\right\}\boldsymbol{1},\\ \boldsymbol{\sigma}_\beta|_{\boldsymbol{c}=\boldsymbol{c}^*} &= k_\sigma^{-1}\left(-k_\sigma^{-1}k_{\sigma 0}\left\{\boldsymbol{I} - \boldsymbol{D}_{\Omega_\zeta}^{-1}\boldsymbol{X}^{-1}\right\}\boldsymbol{1}\right) + k_{\sigma 0}\boldsymbol{1} = k_{\sigma 0}\boldsymbol{D}_{\Omega_\zeta}^{-1}\boldsymbol{X}^{-1}\boldsymbol{1}.\end{aligned}$$

Feasibility condition simplification and positive optimal bias

Intermediary steps for the condition in Assumption 6. Notice that $k_{\sigma 0} > 0$ and by the Woodbury matrix identity

$$\boldsymbol{X}^{-1} = \boldsymbol{D}_{\Omega_\zeta} - \boldsymbol{D}_{\Omega_\zeta}(k^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta})^{-1}\boldsymbol{D}_{\Omega_\zeta} \implies \boldsymbol{D}_{\Omega_\zeta}^{-1}\boldsymbol{X}^{-1} = \boldsymbol{I} - (k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta})^{-1}\boldsymbol{D}_{\Omega_\zeta}.$$

The non-zero condition on the vector of volatilities of the estimator $\hat{\boldsymbol{\beta}}$ at the optimum $\boldsymbol{c}^*$ can be rewritten as

$$\begin{aligned}\boldsymbol{\sigma}_\beta|_{\boldsymbol{c}=\boldsymbol{c}^*} = k_{\sigma 0}\boldsymbol{D}_\Omega^{-1}\boldsymbol{X}^{-1}\boldsymbol{1} \geq \boldsymbol{0} &\iff \{\boldsymbol{I} - (k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta})^{-1}\boldsymbol{D}_{\Omega_\zeta}\}\boldsymbol{1} \geq \boldsymbol{0}\\ &\iff (k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta})\boldsymbol{1} \geq \boldsymbol{D}_{\Omega_\zeta}\boldsymbol{1} \iff \boldsymbol{\Omega}_\zeta\boldsymbol{1} \geq \boldsymbol{0}.\end{aligned}$$

For the bias, notice $-k_c^{-1}k_{\sigma 0} > 0$ so

$$\begin{aligned}\boldsymbol{\varepsilon}_\beta|_{\boldsymbol{c}=\boldsymbol{c}^*} = \left\{\boldsymbol{I} - \boldsymbol{D}_{\Omega_\zeta}^{-1}\boldsymbol{X}^{-1}\right\}\boldsymbol{1} \geq \boldsymbol{0} &\iff \boldsymbol{1} \geq \boldsymbol{D}_{\Omega_\zeta}^{-1}\boldsymbol{X}^{-1}\boldsymbol{1} \iff \boldsymbol{X}\boldsymbol{D}_{\Omega_\zeta}\boldsymbol{1} \geq \boldsymbol{1}\\ &\iff (k_c^2\boldsymbol{\Omega}_\zeta^{-1} + \boldsymbol{D}_{\Omega_\zeta}^{-1})\boldsymbol{D}_{\Omega_\zeta}\boldsymbol{1} \geq \boldsymbol{1} \iff k_c^2\boldsymbol{\Omega}_\zeta^{-1}\boldsymbol{D}_{\Omega_\zeta}\boldsymbol{1} + \boldsymbol{1} \geq \boldsymbol{1}\\ &\iff \boldsymbol{D}_{\Omega_\zeta}\boldsymbol{1} + k_c^{-2}\boldsymbol{\Omega}_\zeta\boldsymbol{1} \geq k_c^{-2}\boldsymbol{\Omega}_\zeta\boldsymbol{1} \iff \boldsymbol{D}_{\Omega_\zeta}\boldsymbol{1} \geq \boldsymbol{0},\end{aligned}$$

and the bias is always positive.

Affine-affine case

Relax Assumption 5 to allow bias to have intercept $k_{\varepsilon 0}$, let $k_{c0} := k_{\sigma 0}/k_{\varepsilon 0}$, $\tilde{\boldsymbol{\Omega}}_\zeta := \boldsymbol{D}_{\Omega_\zeta}^{-\frac{1}{2}} \boldsymbol{\Omega}_\zeta \boldsymbol{D}_{\Omega_\zeta}^{-\frac{1}{2}}$, $\tilde{\boldsymbol{\Omega}}_{\zeta *1} := \tilde{\boldsymbol{\Omega}}_\zeta + k_c k_{c0} \boldsymbol{I}$, $\tilde{\boldsymbol{\Omega}}_{\zeta *0} := \tilde{\boldsymbol{\Omega}}_\zeta + k_{c0}^2 \boldsymbol{I}$, then optimal control and objective at solution are given by

$$\boldsymbol{c}^* = -\frac{k_{\varepsilon 0}}{k_\varepsilon} \boldsymbol{D}_\Omega^{-\frac{1}{2}} \tilde{\boldsymbol{\Omega}}_{\zeta *}^{-1} \tilde{\boldsymbol{\Omega}}_{\zeta *1} \boldsymbol{D}_\Omega^{\frac{1}{2}} \mathbf{1}, \quad \chi = k_{\varepsilon 0}^2 \mathbf{1}^\top \boldsymbol{D}_\Omega^{\frac{1}{2}} \left\{ \tilde{\boldsymbol{\Omega}}_{\zeta *0} - \tilde{\boldsymbol{\Omega}}_{\zeta *1} \tilde{\boldsymbol{\Omega}}_{\zeta *}^{-1} \tilde{\boldsymbol{\Omega}}_{\zeta *1} \right\} \boldsymbol{D}_\Omega^{\frac{1}{2}} \mathbf{1},$$

based on derivations analogous to Appendix A.1.2. Notice that substituting optimal control into the objective the expression for χ follows from

$$\begin{aligned}
\chi = \mathbf{1}^\top \Big\{ & k_\varepsilon^2 \frac{k_{\varepsilon 0}^2}{k_\varepsilon^2} \boldsymbol{D}_\Omega^{\frac{1}{2}} \tilde{\boldsymbol{\Omega}}_{\zeta *0} \tilde{\boldsymbol{\Omega}}_{\zeta *}^{-1} \tilde{\boldsymbol{\Omega}}_\zeta \tilde{\boldsymbol{\Omega}}_{\zeta *}^{-1} \tilde{\boldsymbol{\Omega}}_{\zeta *0} \boldsymbol{D}_\Omega^{\frac{1}{2}} + k_\sigma^2 \frac{k_{\varepsilon 0}^2}{k_\varepsilon^2} \boldsymbol{D}_\Omega^{\frac{1}{2}} \tilde{\boldsymbol{\Omega}}_{\zeta *0} \tilde{\boldsymbol{\Omega}}_{u*}^{-1} \boldsymbol{I} \tilde{\boldsymbol{\Omega}}_{\zeta *}^{-1} \tilde{\boldsymbol{\Omega}}_{\zeta *0} \boldsymbol{D}_\Omega^{\frac{1}{2}} \\
& - 2 k_\varepsilon k_{\varepsilon 0} \frac{k_{\varepsilon 0}}{k_\varepsilon} \boldsymbol{D}_\Omega^{\frac{1}{2}} \tilde{\boldsymbol{\Omega}}_\zeta \tilde{\boldsymbol{\Omega}}_{\zeta *}^{-1} \tilde{\boldsymbol{\Omega}}_{u*0} \boldsymbol{D}_\Omega^{\frac{1}{2}} - k_\sigma k_{\sigma 0} \frac{k_{\varepsilon 0}}{k_\varepsilon} \boldsymbol{D}_\Omega^{\frac{1}{2}} \boldsymbol{I} \tilde{\boldsymbol{\Omega}}_{\zeta *}^{-1} \tilde{\boldsymbol{\Omega}}_{\zeta *0} \boldsymbol{D}_\Omega^{\frac{1}{2}} \\
& + k_{\varepsilon 0}^2 \boldsymbol{D}_\Omega^{\frac{1}{2}} \tilde{\boldsymbol{\Omega}}_\zeta \boldsymbol{D}_\Omega^{\frac{1}{2}} + \frac{k_{\varepsilon 0}^2}{k_{\varepsilon 0}^2} k_{\sigma 0}^2 \boldsymbol{D}_\Omega \Big\} \mathbf{1},
\end{aligned}$$

by collecting terms in the first and second line.

A.1.3 Recursive formulation of $\boldsymbol{\Gamma}_{n_s}$ single signal

For every entry $\boldsymbol{\Gamma}_{i,j,n_s} = \gamma_{ij,n_s} = \boldsymbol{\rho}_{q_i s,n_s} \boldsymbol{R}_{s,n_s}^{-1} \boldsymbol{\rho}_{q_j s,n_s}^\top$, denoting the vector of correlations between s_{n_s} and other informed signals by $\boldsymbol{\rho}_{s,n_s} = Corr(\boldsymbol{s}_{n_s-1}, s_{n_s})$, and the correlation correction $\rho_{s,n_s|n_s-1} = 1 - \boldsymbol{\rho}_{s,n_s}^\top \boldsymbol{R}_{s,n_s-1}^{-1} \boldsymbol{\rho}_{s,n_s}$, block matrix inversion provides the decomposition

$$\begin{aligned}
\boldsymbol{R}_{s,n_s} &= \begin{pmatrix} \boldsymbol{R}_{s,n_s-1} & \boldsymbol{\rho}_{s,n_s} \\ \boldsymbol{\rho}_{s,n_s}^\top & 1 \end{pmatrix} \\
\implies \boldsymbol{R}_{s,n_s}^{-1} &= \begin{pmatrix} \boldsymbol{R}_{s,n_s-1}^{-1} & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{pmatrix} + \frac{1}{\rho_{s,n_s|n_s-1}} \begin{pmatrix} \boldsymbol{R}_{s,n-1}^{-1} \boldsymbol{\rho}_{s,n_s} \boldsymbol{\rho}_{s,n_s}^\top \boldsymbol{R}_{s,n-1}^{-1} & -\boldsymbol{R}_{s,n_s-1}^{-1} \boldsymbol{\rho}_{s,n_s} \\ -\boldsymbol{\rho}_{s,n_s}^\top \boldsymbol{R}_{s,n_s-1}^{-1} & 1 \end{pmatrix}.
\end{aligned}$$

The vector of correlations between a factor and all signals can be split into $\boldsymbol{\rho}_{q_i s,n_s}^\top = \begin{pmatrix} \boldsymbol{\rho}_{q_i s,n_s-1} & \rho_{q_i s_{n_s}} \end{pmatrix}$, let $\phi_{i,n_s} := \boldsymbol{\rho}_{q_i s,n-1}^\top \boldsymbol{R}_{s,n_s-1}^{-1} \boldsymbol{\rho}_{s,n_s} - \rho_{q_i s_{n_s}}$ and the quadratic form γ_{ij,n_s} can be rewritten as

$$\begin{aligned}
\gamma_{ij,n} = \boldsymbol{\rho}_{q_i s,n_s-1}^\top \boldsymbol{R}_{s,n_s-1}^{-1} \boldsymbol{\rho}_{q_j s,n_s-1} + \frac{1}{\rho_{s,n_s|n_s-1}} \Big\{ & \boldsymbol{\rho}_{q_i s,n-1}^\top \boldsymbol{R}_{s,n-1}^{-1} \boldsymbol{\rho}_{s,n} \boldsymbol{\rho}_{s,n}^\top \boldsymbol{R}_{s,n_s-1}^{-1} \boldsymbol{\rho}_{q_j s,n_s-1} + \rho_{q_i s_n} \rho_{q_j s_n} \\
& - \rho_{q_j s_n} \boldsymbol{\rho}_{q_i s,n-1}^\top \boldsymbol{R}_{s,n-1}^{-1} \boldsymbol{\rho}_{s,n} - \rho_{q_i s_n} \boldsymbol{\rho}_{s,n}^\top \boldsymbol{R}_{s,n-1}^{-1} \boldsymbol{\rho}_{q_j s,n-1} \Big\} = \gamma_{ij,n_s-1} + \frac{\phi_{i,n_s} \phi_{j,n_s}}{\rho_{s,n_s|n_s-1}}.
\end{aligned}$$

Since the correlation correction $\rho_{s,n_s|n_s-1}$ is the same across entries, defining the vector $\boldsymbol{\phi}_n^\top = \begin{pmatrix} \phi_{1,n_s} & \phi_{2,n_s} & \dots & \phi_{n_q,n_s} \end{pmatrix}$ the full matrix can be written recursively as

$$\boldsymbol{\Gamma}_{n_s} = \boldsymbol{\Gamma}_{n_s-1} + \frac{1}{\rho_{s,n_s|n_s-1}} \boldsymbol{\phi}_{n_s} \boldsymbol{\phi}_{n_s}^\top$$

The diagonal elements of $\boldsymbol{\Gamma}_{n_s}$ is weakly increasing in n_s as

$$\gamma_{ii,n_s} = \gamma_{ii,n_s-1} + \frac{\phi_{i,n_s}^2}{\rho_{s,n_s|n_s-1}} \geq \gamma_{ii,n_s-1},$$

and the sum over all entries is as well, since the outer product $\boldsymbol{\phi}_{n_s}\boldsymbol{\phi}_{n_s}^\top$ is positive semi-definite, formally

$$\mathbf{1}^\top(\boldsymbol{\Gamma}_{n_s}-\boldsymbol{\Gamma}_{n_s-1})\mathbf{1}=\frac{1}{\rho_{s,n_s|n_s-1}}\mathbf{1}^\top\boldsymbol{\phi}_{n_s}\boldsymbol{\phi}_{n_s}^\top\mathbf{1}\geq 0.$$

A.1.4 Cost and benefit of complexity single signal

The impact of adding another signal is always to weakly increase χ as can be demonstrated by the positive semi-definiteness of the difference $\boldsymbol{X}_n^{-1}-\boldsymbol{X}_{n-1}^{-1}$. By properties of symmetric positive definite matrices,[1] the difference is semi-positive definite if the difference $\boldsymbol{X}_{n-1}-\boldsymbol{X}_n$ is. Let $\boldsymbol{D}_{\phi,n_s}=\rho_{s,n_s|n_s-1}^{-1}\operatorname{diag}\left(\boldsymbol{\phi}_{n_s}\boldsymbol{\phi}_{n_s}^\top\right)$, then explicit calculation using the Sherman-Morrison formula for the inverse of the sum of a positive definite matrix and the outer product of vectors of the difference yields

$$\begin{aligned}
&\boldsymbol{X}_{n-1}-\boldsymbol{X}_n=\\
&k_c^2\boldsymbol{\Omega}_{\zeta,n_s-1}^{-1}+\boldsymbol{D}_{\Omega_\zeta,n_s-1}^{-1}-k_c^2\boldsymbol{\Omega}_{\zeta,n_s}^{-1}-\boldsymbol{D}_{\Omega_\zeta,n_s}^{-1}=k_c^2\left(\boldsymbol{\Omega}_{\zeta,n_s-1}^{-1}-\boldsymbol{\Omega}_{\zeta,n_s}^{-1}\right)+\boldsymbol{D}_{\Omega_\zeta,n_s-1}^{-1}-\boldsymbol{D}_{\Omega_\zeta,n_s}^{-1}\\
&\quad=k_c^2\left[\boldsymbol{\Omega}_{\zeta,n_s-1}^{-1}-\left(\boldsymbol{\Omega}_{\zeta,n_s-1}^{-1}-\frac{\rho_{s,n_s|n_s-1}^{-1}\boldsymbol{\Omega}_{\zeta,n_s-1}^{-1}\boldsymbol{D}_{\sigma_q}\boldsymbol{\phi}_{n_s}\boldsymbol{\phi}_{n_s}^\top\boldsymbol{D}_{\sigma_q}\boldsymbol{\Omega}_{\zeta,n_s-1}^{-1}}{1+\rho_{s,n|n-1}^{-1}\left(\boldsymbol{\sigma}_q\odot\boldsymbol{\phi}_{n_s}\right)^\top\boldsymbol{\Omega}_{\zeta,n_s-1}^{-1}\left(\boldsymbol{\sigma}_q\odot\boldsymbol{\phi}_{n_s}\right)}\right)\right]\\
&\qquad+\boldsymbol{D}_{\Omega_\zeta,n_s-1}^{-1}-\left[\boldsymbol{D}_{\Omega_\zeta,n_s-1}^{-1}-\left(\boldsymbol{D}_{\Omega_\zeta,n_s}+\boldsymbol{D}_{\Omega_\zeta,n_s}\boldsymbol{D}_{\sigma_q}^{-1}\boldsymbol{D}_{\phi,n_s}^{-1}\boldsymbol{D}_{\sigma_q}^{-1}\boldsymbol{D}_{\Omega,n}\right)^{-1}\right]\\
&\quad=k_c^2\frac{\rho_{s,n_s|n_s-1}^{-1}\boldsymbol{\Omega}_{\zeta,n_s-1}^{-1}\boldsymbol{D}_{\sigma_q}\boldsymbol{\phi}_{n_s}\boldsymbol{\phi}_{n_s}^\top\boldsymbol{D}_{\sigma_q}\boldsymbol{\Omega}_{\zeta,n_s-1}^{-1}}{1+\rho_{s,n_s|n_s-1}^{-1}\left(\boldsymbol{\sigma}_q\odot\boldsymbol{\phi}_{n_s}\right)^\top\boldsymbol{\Omega}_{\zeta,n_s-1}^{-1}\left(\boldsymbol{\sigma}_q\odot\boldsymbol{\phi}_{n_s}\right)}\\
&\qquad+\left(\boldsymbol{D}_{\Omega_\zeta,n_s}+\boldsymbol{D}_{\Omega_\zeta,n_s}\boldsymbol{D}_{\sigma_q}^{-1}\boldsymbol{D}_{\phi,n_s}^{-1}\boldsymbol{D}_{\sigma_q}^{-1}\boldsymbol{D}_{\Omega_\zeta,n_s}\right)^{-1},
\end{aligned}$$

which, as a sum of (semi-)positive definite matrices, is semi-positive definite.

A.1.5 Recursive formulation of $\boldsymbol{\Gamma}_{n_s}$ multiple signals

For every entry $\boldsymbol{\Gamma}_{i,j,n_s}=\gamma_{ij,n_s}=\boldsymbol{\rho}_{q_is,n_s}\boldsymbol{R}_{s,n_s}^{-1}\boldsymbol{\rho}_{q_js,n_s}^\top$, denoting the matrix of correlations between $\boldsymbol{s}_{n_s+}$ and other informed signals by $\boldsymbol{R}_{n_s-n_s+}=Corr(\boldsymbol{s}_{I,n_s},\boldsymbol{s}_{n_s+})$, and the correlation correction $\boldsymbol{R}_{n_s+|n_s-}=\boldsymbol{R}_{s,n_s+}-\boldsymbol{R}_{n_s-n_s+}^\top\boldsymbol{R}_{s,n_s-}^{-1}\boldsymbol{R}_{n_s-n_s+}$, block matrix inversion provides the decomposition

$$\begin{aligned}
\boldsymbol{R}_{s,n_s}&=\begin{pmatrix}\boldsymbol{R}_{s,n_s-} & \boldsymbol{R}_{n_s-n_s+}\\ \boldsymbol{R}_{n_s-n_s+}^\top & \boldsymbol{R}_{s,n_s+}\end{pmatrix}\\
\implies \boldsymbol{R}_{s,n_s}^{-1}&=\begin{pmatrix}\boldsymbol{R}_{s,n_s-}^{-1} & \mathbf{0}\\ \mathbf{0}^\top & \mathbf{0}\mathbf{0}^\top\end{pmatrix}\\
&\quad+\begin{pmatrix}\boldsymbol{R}_{s,n_s-}^{-1}\boldsymbol{R}_{n_s-n_s+}\boldsymbol{R}_{n_s+|n_s-}^{-1}\boldsymbol{R}_{n_s-n_s+}^\top\boldsymbol{R}_{s,n_s-}^{-1} & -\boldsymbol{R}_{s,n_s-}^{-1}\boldsymbol{R}_{n_s-n_s+}\boldsymbol{R}_{n_s+|n_s-}^{-1}\\ -\boldsymbol{R}_{n_s+|n_s-}^{-1}\boldsymbol{R}_{n_s-n_s+}^\top\boldsymbol{R}_{s,n_s-}^{-1} & \boldsymbol{R}_{n_s+|n_s-}^{-1}\end{pmatrix}.
\end{aligned}$$

The vector of correlations between a factor and all signals can be split into $\boldsymbol{\rho}_{q_is,n_s}^\top=\begin{pmatrix}\boldsymbol{\rho}_{q_is,n_s-}^\top & \boldsymbol{\rho}_{q_is,n_s+}^\top\end{pmatrix}$, let $\boldsymbol{\phi}_{i,n_s}^\top:=\boldsymbol{\rho}_{q_is,n_s-}^\top\boldsymbol{R}_{s,n_s-}^{-1}\boldsymbol{R}_{n_s-n_s+}-\boldsymbol{\rho}_{q_is,n_s+}^\top$ and the quadratic form γ_{ij,n_s} can be rewritten as

$$\gamma_{ij,n}=\boldsymbol{\rho}_{q_is,n_s-}^\top\boldsymbol{R}_{s,n_s-1}^{-1}\boldsymbol{\rho}_{q_js,n_s-}$$

[1] For symmetric positive definite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with the same dimensions it is the case that

$$\boldsymbol{v}^\top(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{v}\geq 0\implies \boldsymbol{v}^\top(\boldsymbol{B}^{-1}-\boldsymbol{A}^{-1})\boldsymbol{v}\geq 0\;\forall\boldsymbol{v},$$

i.e. if the difference $\boldsymbol{A}-\boldsymbol{B}$ is positive semi-definite, the difference $\boldsymbol{B}^{-1}-\boldsymbol{A}^{-1}$ is positive semi-definite.

$$
\begin{aligned}
&+\boldsymbol{\rho}_{q_i s,n_s-}^\top \boldsymbol{R}_{s,n_s-}^{-1}\boldsymbol{R}_{n_s-n_s+}\boldsymbol{R}_{n_s+|n_s-}^{-1}\boldsymbol{R}_{n_s-n_s+}^\top \boldsymbol{R}_{s,n_s-}^{-1}\boldsymbol{\rho}_{q_i s,n_s-}\\
&-\boldsymbol{\rho}_{q_i s,n_s-}^\top \boldsymbol{R}_{s,n_s-}^{-1}\boldsymbol{R}_{n_s-n_s+}\boldsymbol{R}_{n_s+|n_s-}^{-1}\boldsymbol{\rho}_{q_i s,n_s+}\\
&-\boldsymbol{\rho}_{q_i s,n_s+}^\top \boldsymbol{R}_{n_s+|n_s-}^{-1}\boldsymbol{R}_{n_s-n_s+}^\top \boldsymbol{R}_{s,n_s-}^{-1}\boldsymbol{\rho}_{q_i s,n_s-}\\
&+\boldsymbol{\rho}_{q_i s,n_s+}^\top \boldsymbol{R}_{n_s+|n_s-}^{-1}\boldsymbol{\rho}_{q_i s,n_s+}\\
=\;&\gamma_{ij,n_s-1}+\boldsymbol{\phi}_{i,n_s}^\top \boldsymbol{R}_{s,n_s+|n_s-}^{-1}\boldsymbol{\phi}_{j,n_s}.
\end{aligned}
$$

Since the correlation correction matrix $\boldsymbol{R}_{n_s+|n_s-}^{-1}$ is the same across entries, defining the matrix $\boldsymbol{\Phi}_{n_s}=\boldsymbol{R}_{qs,n_s-}\boldsymbol{R}_{s,n_s-}^{-1}\boldsymbol{R}_{n_s-n_s+}-\boldsymbol{R}_{qs,n_s+}$, where the rows are $\boldsymbol{\phi}_{i,n_s}^\top$, the full matrix can be written recursively as

$$
\boldsymbol{\Gamma}_{n_s}=\boldsymbol{\Gamma}_{n_s-}+\boldsymbol{\Phi}_{n_s}\boldsymbol{R}_{s,n_s+|n_s-}^{-1}\boldsymbol{\Phi}_{n_s}^\top.
$$

The diagonal elements of $\boldsymbol{\Gamma}_{n_s}$ are weakly increasing in n_s as $\boldsymbol{R}_{s,n_s+|n_s-}^{-1}$ is positive semi-definite so

$$
\gamma_{ii,n_s}=\gamma_{ii,n_s-}+\boldsymbol{\phi}_{i,n_s}\boldsymbol{R}_{s,n_s+|n_s-}^{-1}\boldsymbol{\phi}_{i,n_s}^\top\geq\gamma_{ii,n_s-},
$$

and the sum over all entries is as well, since $\boldsymbol{\Gamma}_{n_s}-\boldsymbol{\Gamma}_{n_s-}=\boldsymbol{\Phi}_{n_s}\boldsymbol{R}_{s,n_s+|n_s-}^{-1}\boldsymbol{\Phi}_{n_s}^\top$ is positive semi-definite.

A.1.6 Cost and benefit of complexity multiple signals

The impact of adding another group of signal is always to weakly increase χ as can be demonstrated by the positive semi-definiteness of the difference $\boldsymbol{X}_{n_s+}^{-1}-\boldsymbol{X}_{n_s-}^{-1}$. By properties of symmetric positive definite matrices,[2] the difference is semi-positive definite if the difference $\boldsymbol{X}_{n_s-}-\boldsymbol{X}_{n_s+}$ is. Explicit calculation using the Sherman-Morrison formula for the inverse of the sum of a positive definite matrix and the outer product of vectors of the difference yields

$$
\begin{aligned}
&\boldsymbol{X}_{n_s-}-\boldsymbol{X}_{n_s+}=\\
&\quad k_c^2\boldsymbol{\Omega}_{\zeta,n_s-}^{-1}+\boldsymbol{D}_{\Omega_\zeta,n_s-}^{-1}-k_c^2\boldsymbol{\Omega}_{\zeta,n_s}^{-1}-\boldsymbol{D}_{\Omega_\zeta,n_s}^{-1}=k_c^2\left(\boldsymbol{\Omega}_{\zeta,n_s-}^{-1}-\boldsymbol{\Omega}_{\zeta,n_s}^{-1}\right)+\boldsymbol{D}_{\Omega_\zeta,n_s-}^{-1}-\boldsymbol{D}_{\Omega_\zeta,n_s}^{-1}\\
&\quad=k_c^2\left[\boldsymbol{\Omega}_{\zeta,n_s-}^{-1}-\left(\boldsymbol{\Omega}_{\zeta,n_s-}^{-1}-\boldsymbol{\Omega}_{\zeta,n_s-}^{-1}\boldsymbol{\Phi}_{n_s}^\top\left(\boldsymbol{R}_{s,n_s+|n_s-}+\boldsymbol{\Phi}_{n_s}^\top\boldsymbol{\Omega}_{\zeta,n_s-}^{-1}\boldsymbol{\Phi}_{n_s}\right)\boldsymbol{\Phi}_{n_s}^\top\boldsymbol{\Omega}_{\zeta,n_s-}^{-1}\right)\right]\\
&\qquad+\boldsymbol{D}_{\Omega_\zeta,n_s-}^{-1}-\left[\boldsymbol{D}_{\Omega_\zeta,n_s-}^{-1}-\left(\boldsymbol{D}_{\Omega_\zeta,n_s}+\boldsymbol{D}_{\Omega_\zeta,n_s}\boldsymbol{D}_{\sigma_q}^{-1}\boldsymbol{D}_{\Phi,n_s}^{-1}\boldsymbol{D}_{\sigma_q}^{-1}\boldsymbol{D}_{\Omega,n}\right)^{-1}\right]\\
&\quad=k_c^2\boldsymbol{\Omega}_{\zeta,n_s-}^{-1}\boldsymbol{\Phi}_{n_s}^\top\left(\boldsymbol{R}_{s,n_s+|n_s-}+\boldsymbol{\Phi}_{n_s}^\top\boldsymbol{\Omega}_{\zeta,n_s-}^{-1}\boldsymbol{\Phi}_{n_s}\right)\boldsymbol{\Phi}_{n_s}^\top\boldsymbol{\Omega}_{\zeta,n_s-}^{-1}\\
&\qquad+\left(\boldsymbol{D}_{\Omega_\zeta,n_s}+\boldsymbol{D}_{\Omega_\zeta,n_s}\boldsymbol{D}_{\sigma_q}^{-1}\boldsymbol{D}_{\Phi,n_s}^{-1}\boldsymbol{D}_{\sigma_q}^{-1}\boldsymbol{D}_{\Omega_\zeta,n_s}\right)^{-1},
\end{aligned}
$$

which, as a sum of (semi-)positive definite matrices, is semi-positive definite.

A.1.7 Impact on including signals of improving technology

The matrix derivative

$$
\begin{aligned}
\frac{\partial\boldsymbol{X}_{n_s+}^{-1}-\boldsymbol{X}_{n_s-}^{-1}}{\partial k_c^2}&=-\boldsymbol{X}_{n_s+}^{-1}\frac{\partial\boldsymbol{X}_{n_s+}}{\partial k_c^2}\boldsymbol{X}_{n_s+}^{-1}+\boldsymbol{X}_{n_s-}^{-1}\frac{\boldsymbol{X}_{n_s-}}{\partial k_c^2}\boldsymbol{X}_{n_s-}^{-1}\\
&=-\boldsymbol{X}_{n_s+}^{-1}\boldsymbol{\Omega}_{\zeta,n_s+}^{-1}\boldsymbol{X}_{n_s+}+\boldsymbol{X}_{n_s-}^{-1}\boldsymbol{\Omega}_{\zeta,n_s-}^{-1}\boldsymbol{X}_{n_s-}^{-1}
\end{aligned}
$$

[2]For symmetric positive definite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with the same dimensions it is the case that

$$
\boldsymbol{v}^\top(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{v}\geq 0\implies\boldsymbol{v}^\top(\boldsymbol{B}^{-1}-\boldsymbol{A}^{-1})\boldsymbol{v}\geq 0\;\forall\boldsymbol{v},
$$

i.e. if the difference $\boldsymbol{A}-\boldsymbol{B}$ is positive semi-definite, the difference $\boldsymbol{B}^{-1}-\boldsymbol{A}^{-1}$ is positive semi-definite.

is positive semi-definite if the matrix difference

$$
\begin{aligned}
&\boldsymbol{X}_{n_s+}\boldsymbol{\Omega}_{\zeta,n_s+}\boldsymbol{X}_{n_s+} - \boldsymbol{X}_{n_s-}\boldsymbol{\Omega}_{\zeta,n_s-}\boldsymbol{X}_{n_s-} \\
&\quad = k_c^4\left\{\boldsymbol{\Omega}_{\zeta,n_s+}^{-1} - \boldsymbol{\Omega}_{\zeta,n_s-}^{-1}\right\} + \left\{\boldsymbol{D}_{\Omega_\zeta,n_s+}^{-1}\boldsymbol{\Omega}_{\zeta,n_s+}\boldsymbol{D}_{\Omega_\zeta,n_s+}^{-1} - \boldsymbol{D}_{\Omega_\zeta,n_s-}^{-1}\boldsymbol{\Omega}_{\zeta,n_s-}\boldsymbol{D}_{\Omega_\zeta,n_s-}^{-1}\right\} \\
&\qquad + k_c^2\left\{\boldsymbol{\Omega}_{\zeta,n_s+}\boldsymbol{D}_{\Omega_\zeta,n_s+}^{-1} + \boldsymbol{D}_{\Omega_\zeta,n_s+}^{-1}\boldsymbol{\Omega}_{\zeta,n_s+} - \boldsymbol{\Omega}_{\zeta,n_s-}\boldsymbol{D}_{\Omega_\zeta,n_s-}^{-1} - \boldsymbol{D}_{\Omega_\zeta,n_s-}^{-1}\boldsymbol{\Omega}_{\zeta,n_s-}\right\}
\end{aligned}
$$

is positive definite. This is always eventually true as $k_c^2 \to \infty$ and only the first term remains since $\boldsymbol{\Omega}_{\zeta,n_s-} - \boldsymbol{\Omega}_{\zeta,n_s+}$ is positive definite.

A.1.8 Objective measure is the limiting measure

This is true as the cost of complexity is reduced to zero as estimation technology parameter goes to infinity $k_c^2 \to \infty$ (k_c is the ratio of slope parameters in the bias variance trade-off minimization, see Section 1.2.3). This can seen by taking the limit of a slightly rearranged cost of complexity χ

$$
\lim_{k_c^2 \to} \chi = \lim_{k_c^2 \to} k_c^{-2} k_{\sigma 0}^2 \boldsymbol{1}^\top \left\{\boldsymbol{\Omega}_\zeta^{-1} + k_c^{-2}\boldsymbol{D}_{\Omega_\zeta}^{-1}\right\}^{-1}\boldsymbol{1} = 0 \times k_{\sigma 0}^2 \boldsymbol{1}^\top \boldsymbol{\Omega}_\zeta \boldsymbol{1} = 0.
$$

A.2 Heterogeneous agents

A.2.1 Profit function from budget constraint

Normalize the price and pay-off of the risk free asset to one (equivalent to a risk free rate of zero). The value of the position in the risk free asset is its size, denote it by B_i. With the initial value of position in the risky asset given by position times price $\delta_i p$, the budget constraint that the total value of investments cannot exceed initial wealth w_{0i} is $w_{0i} \geq B_i + \delta_i p$. For a utility function increasing in wealth the budget constraint binds, and it follows that $B_i = w_{0i} - \delta_i p$. Therefore, after-trade wealth is

$$
w_{1i} = B_i + \delta_i y = w_{0i} - \delta_i p + \delta_i y = w_{0i} + \delta_i (y - p).
$$

It can be shown that in determining position δ_i, for an investor with CARA utility of after-trade wealth, initial wealth (endowment) can be normalized to one without loss of generality, see Breon-Drish (2015). This holds as well for an investor maximizing a mean-variance criterion as initial wealth drops out of the first order condition because it enters wealth additively, which extends to the mean-mean squared error criterion described in the main text. In all three cases optimizing the profit function $\pi_i = \delta_i (y - p)$ is equivalent to optimizing after-trade wealth.

A.2.2 Predicting prediction

To see that $E[y|s_U, \boldsymbol{\beta}] = E[\hat{y}_I|s_U, \boldsymbol{\beta}]$, notice that for known $\boldsymbol{\beta}$

$$
\begin{aligned}
Cov[y, s_U|\boldsymbol{\beta}] &= Cov[y, \hat{y}_I|\boldsymbol{\beta}] = \boldsymbol{\beta}^\top Cov[\boldsymbol{q}, \boldsymbol{u}]\boldsymbol{\beta} = \boldsymbol{\beta}^\top Cov[\boldsymbol{q}, \boldsymbol{s}_I]\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Sigma}_{qs}^\top\boldsymbol{\beta} = \boldsymbol{\beta}^\top\boldsymbol{\Sigma}_{qs}\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Sigma}_{qs}^\top\boldsymbol{\beta} \\
&= \boldsymbol{\beta}^\top\boldsymbol{\Sigma}_{qs}\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Sigma}_s\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Sigma}_{qs}^\top\boldsymbol{\beta} = \boldsymbol{\beta}^\top\boldsymbol{\Sigma}_{qs}\boldsymbol{\Sigma}_s^{-1}Var[\boldsymbol{s}_I]\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Sigma}_{qs}^\top\boldsymbol{\beta} = Var[\hat{y}_I|\boldsymbol{\beta}],
\end{aligned}
$$

and

$$
E[y|\boldsymbol{\beta}] = \boldsymbol{\beta}^\top\boldsymbol{\mu}_q = \boldsymbol{\beta}^\top\{\boldsymbol{\mu}_q + \boldsymbol{\Sigma}_{qs}\boldsymbol{\Sigma}_s^{-1}(E[\boldsymbol{s}_I] - \boldsymbol{\mu}_s)\} = E[\hat{y}_I|\boldsymbol{\beta}] = E[s_U|\boldsymbol{\beta}],
$$

so

$$E[y|s_U,\boldsymbol{\beta}] = E[\hat{y}_I|\boldsymbol{\beta}] + \frac{Var[\hat{y}_I|\boldsymbol{\beta}]}{Var[s_U|\boldsymbol{\beta}]}(s_U - E[s_U|\boldsymbol{\beta}]) = E[\hat{y}_I|s_U,\boldsymbol{\beta}].$$

A.2.3 Demand: robust profit maximization objective

I assume that investors optimize an extended mean-variance objective that consist of the expectation of the scaled profit function $\tilde{\pi}_i(y) := \delta_i \alpha_i (y-p)$ applied to the prediction $\hat{y}_i$ and an uncertainty-adjustment for the fact that investors optimize estimated rather than true profits extended from unconditional variance to unconditional mean squared error. In this two period model, optimizing over profits corresponds to optimizing over second period wealth (see Appendix A.2.1). The specification reflects the fact that investors ultimately care about true profits but are averse to the risk in the pay-off as well as the model uncertainty. To account for both sources of randomness, the uncertainty-adjustment is based on the unconditional mean squared error, and, for consistency with the problem faced by investors with CARA-utility facing a Gaussian gamble, it is half, i.e. $\frac{1}{2}\delta_i^2 E\left[\left(\tilde{\pi}_i(y)-\tilde{\pi}_i(\hat{y})\right)^2\right] = \frac{1}{2}\alpha_i^2 E\left[\left(y-\hat{y}_i\right)^2\right]$. Formally, demand from optimizing the objective function yields

$$\delta_i = \arg\max\ \tilde{\pi}_i(\hat{y}_i) - \frac{1}{2}E\left[\left(\tilde{\pi}_i(y)-\tilde{\pi}_i(\hat{y})\right)^2\right] = \psi_i\left(\hat{y}_i - p\right), \text{ where } \psi_i = \left\{\alpha_i E\left[\left(y-\hat{y}_i\right)^2\right]\right\}^{-1}.$$

where the expression follows from a reorganization of the first-order condition analogous to a classic mean-variance optimization and the second-order condition is satisfied due to the positivity of the mean squared error. If investors know the true model they act as mean-variance optimizers since the covariance between the pay-off and the predictor is the variance of the predictor $Cov[y,\hat{y}_i] = Var[\hat{y}_i]$ and the predictor is unbiased, $E[y]-E[\hat{y}_i]$, so the mean squared of the predictor equals the conditional variance of the pay-off given the predictor, i.e.

$$E\left[\left(y-\hat{y}_i\right)^2\right] = Var[y] - Var[\hat{y}_i] = Var[y|\hat{y}_i].$$

Without the noisy estimation introduced in Section 1.2.1, my specification of demand is, as was the case for uninformed inference in Section 1.3.1, simply a re-formulation of the baseline model provided by Grossman and Stiglitz (1980), the classic mean-variance criterion for utility optimization.

A.2.4 Demand: CARA-utility with ambiguity aversion

For the informed investors I in Section 1.3.1, their linear demand function is the demand of investors with CARA-utility and risk tolerance α_I, who performs a maximization of utility of final wealth w_{1I} over demand δ_I given price p that is robust to miscalculated risk. Conditional on the estimate $\hat{\boldsymbol{\beta}}$, the pay-off y is normally distributed, which extends to final wealth w_{I1}. Utility of final wealth is log-normal and expected utility is

$$E[U_I] = E\left[-e^{-\alpha_I^{-1}w_{1I}}\middle|\hat{\boldsymbol{\beta}},\boldsymbol{s},p\right] = -e^{-\alpha_I^{-1}E\left[w_{1I}\middle|\hat{\boldsymbol{\beta}},\boldsymbol{s},p\right]+\frac{1}{2}\alpha_I^{-2}Var\left[w_{1I}\middle|\hat{\boldsymbol{\beta}},\boldsymbol{s},p\right]}.$$

Maximizing the negative exponential is equivalent to minimizing its exponent, which again can be turned into a maximization by swapping the sign on the objective. Starting wealth can be normalized to zero without loss of generality, so final wealth is $w_{1I} = \delta_I(y-p)$, see Appendix A.2.1. As in the main text, the conditional expectation of the pay-off is the predictor $\hat{y}_I$. Denote the conditional variance of the pay-off by $\sigma^2_{\hat{y}I} \in \left[\underline{\sigma}^2_{\hat{y}I}, \overline{\sigma}^2_{\hat{y}I}\right]$, where the interval is the set of multiple priors of a maxmin expected utility model in the tradition of Gilboa and Schmeidler (1989).

$$\max_{\delta_I}\min_{\sigma^2_{\hat{y}I}} \alpha_I^{-1}\delta_I(\hat{y}_I - p) - \delta_I^2\frac{1}{2}\alpha_I^{-2}\sigma^2_{\hat{y}I}, \quad s.t. \quad \sigma^2_{\hat{y}I} \in \left[\underline{\sigma}^2_{\hat{y}I}, \overline{\sigma}^2_{\hat{y}I}\right].$$

The first order condition of the minimization is the expression $-\delta_i^2\alpha_i^{-2}/2$, which is always negative, meaning that the unconstrained solution would be positive infinity and the constrained solution is the upper bound. Substituting the upper bound into the objective, differentiating, and solving for the optimal position in the risky asset yields a result with a familiar form $\delta_I^* = \alpha_I\overline{\sigma}^{-2}_{\hat{y}I}(\hat{y}_I - p)$. However, rather than being scaled by the inverse conditional variance, the position is scaled by the inverse worst-outcome variance. The product of ambiguity and ambiguity aversion, effective ambiguity, is pinned down by defining the upper bound of the multiple priors set. Setting it equal to the unconditional mean squared error, i.e. $\overline{\sigma}^2_{\hat{y}I} = E[(y-\hat{y}_I)^2]$, yields the demand function in the main text.

The assumption that the uninformed predictor is the best linear approximation of the the informed prediction $\hat{y}_I$ can be tightened to the uninformed investors approximating the distribution of $\hat{y}_I$ by a normal distribution. This stronger assumption also makes $\hat{y}_U$ the projection presented in the main text. Additionally, following the steps outlined for the informed investors, uninformed investors' demand function in Section 1.3.1 corresponds to CARA-utility with risk tolerance α_U optimizing over profit $\delta_U(\hat{y}_I - p)$ robust to miscalculation of risk. Formally, expected utility is

$$E[U_U] = E\left[-e^{-\alpha_U^{-1}\delta_U(\hat{y}_I-p)}\Big|s_U, p\right] = -e^{-\alpha_U^{-1}\delta_U E\left[\hat{y}_I\middle|\boldsymbol{s}_U,p\right] + \frac{1}{2}\alpha_U^{-2}Var\left[\hat{y}_I\middle|\boldsymbol{s}_U,p\right]}.$$

With predictor $\hat{y}_U = E\left[\hat{y}_I\middle|\boldsymbol{s}_U, p\right]$ and conditional variance $Var\left[\hat{y}_I\middle|\boldsymbol{s}_U, p\right] = \sigma^2_{\hat{y}U} \in \left[\underline{\sigma}^2_{\hat{y}U}, \overline{\sigma}^2_{\hat{y}U}\right]$ the robust optimization is

$$\max_{\delta_U}\min_{\sigma^2_{\hat{y}U}} \alpha_U^{-1}\delta_U(\hat{y}_U - p) - \delta_U^2\frac{1}{2}\alpha_U^{-2}\sigma^2_{\hat{y}U}, \quad s.t. \quad \sigma^2_{\hat{y}U} \in \left[\underline{\sigma}^2_{\hat{y}U}, \overline{\sigma}^2_{\hat{y}U}\right].$$

Setting the upper limit equal to the unconditional mean squared error, i.e. $\overline{\sigma}^2_{\hat{y}U} = E[(y-\hat{y}_U)^2]$, yield the demand function in the main text.

A.2.5 Uninformed mean squared error

The mean squared error of the predictor of the uninformed

$$\begin{aligned}
E[(y-\hat{y}_U)^2] &= Var[y] + Var[\hat{y}_U] - 2Cov[y,\hat{y}_U] + (E[y] - E[\hat{y}_U])^2 \\
&= Var[y] + \lambda_U^2\{Var[\hat{y}_I] + \psi_I^{-2}\sigma_z^2\} - 2\lambda_U Cov[y,\hat{y}_I] \\
&\quad + [(\boldsymbol{\beta} - \{(1-\lambda_U)\boldsymbol{\mu}_\beta + \lambda_U\boldsymbol{\mu}_\beta\})^\top\boldsymbol{\mu}_q]^2 \\
&= Var[y] + \lambda_U(Var[\hat{y}_I] - 2Cov[y,\hat{y}_I]) + (\boldsymbol{\varepsilon}_\beta^\top\boldsymbol{\mu}_q)^2 \\
&= Var[y] + \lambda_U(\chi - \boldsymbol{\beta}^\top\boldsymbol{\Sigma}_\zeta\boldsymbol{\beta}) + (\boldsymbol{\varepsilon}_\beta^\top\boldsymbol{\mu}_q)^2 \\
&= (1-\lambda_U)\left\{Var[y] + (\boldsymbol{\varepsilon}_\beta^\top\boldsymbol{\mu}_q)^2\right\} + \lambda_U E[(y-\hat{y}_I)^2]
\end{aligned}$$

so a necessary and sufficient condition for higher lower squared error of informed vs uninformed is

$$\begin{aligned}
E[(y-\hat{y}_U)^2] > E[(y-\hat{y}_I)^2] &\iff \left\{Var[y] + (\boldsymbol{\varepsilon}_\beta^\top\boldsymbol{\mu}_q)^2\right\} > E[(y-\hat{y}_I)^2] \\
&\iff \chi < \boldsymbol{\beta}^\top\boldsymbol{\Sigma}_\zeta\boldsymbol{\beta} + (\boldsymbol{\varepsilon}_\beta^\top\boldsymbol{\mu}_q)^2 = Var[E[y|\boldsymbol{\beta},\boldsymbol{s}_I]] + (\boldsymbol{\varepsilon}_\beta^\top\boldsymbol{\mu}_q)^2.
\end{aligned}$$

A.2.6 Variable share of informed investors

Market clearing $\ell_I \delta_I(p) + (1-\ell_I)\delta_U(p) = z$ so

$$s_U = p - \psi_I^{-1}\ell_I^{-1}(1-\ell_I)\delta_U(p) = \hat{y}_I - \psi_I^{-1}\ell_I^{-1}z$$

and

$$\begin{aligned} p &= \frac{\ell_I\psi_I\hat{y}_I + (1-\ell_I)\psi_U\hat{y}_U - z}{\ell_I\psi_I + (1-\ell_I)\psi_U} = \frac{\left\{\ell_I\psi_I + (1-\ell_I)\lambda_U\psi_U\right\}s_U + (1-\ell_I)(1-\lambda_U)\psi_U E[y|\boldsymbol{c}^*]}{\ell_I\psi_I + (1-\ell_I)\psi_U} \\ &= (1-\lambda_p)E[y|\boldsymbol{c}^*] + \lambda_p s_U \quad s.t. \quad \lambda_p = \frac{\ell_I\psi_I + (1-\ell_I)\lambda_U\psi_U}{\ell_I\psi_I + (1-\ell_I)\psi_U} \end{aligned}$$

so the changes to equilibrium outcomes are captured by price responsiveness λ_p, uninformed responsiveness λ_U, and uninformed signal s_U.

A.3 Predictions

A.3.1 Short-term price reversals

Price and stochastic supply. Moments: variance

$$\begin{aligned} Var[p,z]^{-1} &= \begin{pmatrix} Var[p] & -\lambda_p\psi_I^{-1}\sigma_z^2 \\ -\lambda_p\psi_I^{-1}\sigma_z^2 & \sigma_z^2 \end{pmatrix}^{-1} \\ &= \frac{1}{\lambda_p^2 Var[\hat{y}_I]\sigma_z^2}\begin{pmatrix} \sigma_z^2 & \lambda_p\psi_I^{-1}\sigma_z^2 \\ \lambda_p\psi_I^{-1}\sigma_z^2 & Var[p] \end{pmatrix} \end{aligned}$$

and covariance

$$Cov\left[y-p,\begin{pmatrix}p & z\end{pmatrix}\right] = \begin{pmatrix}Cov[y,p] - Var[p] & \lambda_p\psi_I^{-1}\sigma_z^2\end{pmatrix}$$

Coefficient on price

$$\begin{aligned} &\frac{(Cov[y,p]-Var[p])\sigma_z^2 + \lambda_p^2\psi_I^{-2}\sigma_z^4}{\lambda_p^2 Var[\hat{y}_I]\sigma_z^2}(p-E[p]) = \frac{Cov[y,p] - \lambda_p^2 Var[\hat{y}_I] - \lambda_p^2\psi_I^{-2}\sigma_z^2 + \lambda_p^2\psi_I^{-2}\sigma_z^2}{\lambda_p^2 Var[\hat{y}_I]}(p-E[p]) \\ &= \left(\frac{\lambda_p Cov[y,\hat{y}_I]}{\lambda_p^2 Var[\hat{y}_I]} - 1\right)(p-E[p]) = \frac{Cov[y,\hat{y}_I] - \lambda_p Var[\hat{y}_I]}{\lambda_p Var[\hat{y}_I]}(p-E[p]). \end{aligned}$$

and on stochastic supply/negative liquidity demand

$$\frac{\lambda_p\psi_I^{-1}\sigma_z^2}{\lambda_p^2 Var[\hat{y}_I]\sigma_z^2}\left\{Cov[y,p] - Var[p] + Var[p]\right\}z = -\frac{Cov[y,p]}{\lambda_p Var[\hat{y}_I]}\psi_I^{-1}(-z)$$

in baseline model $Cov[y,\hat{y}_I] = Var[\hat{y}_I]$, so $Cov[y,\hat{y}_I] - \lambda_p Var[\hat{y}_I] = (1-\lambda_p)Var[\hat{y}_I]$.

Noisy factor loadings

Including noisy factor loadings. Let $Var[p,z] = \boldsymbol{\Sigma}_{pz}$ and

$$\boldsymbol{\Sigma}_{pz|\hat{\beta}} = \boldsymbol{\Sigma}_{pz} - \begin{pmatrix}\lambda_p\boldsymbol{\mu}_q^\top\boldsymbol{D}_{\sigma_\beta}^2 \\ \boldsymbol{0}^\top\end{pmatrix}\boldsymbol{D}_{\sigma_\beta}^{-2}\begin{pmatrix}\lambda_p\boldsymbol{D}_{\sigma_\beta}^2\boldsymbol{\mu}_q & \boldsymbol{0}\end{pmatrix}$$

$$= \boldsymbol{\Sigma}_{pz} - \begin{pmatrix} \lambda_p \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{D}^{-2}_{\sigma_\beta} \\ \mathbf{0}^\top \end{pmatrix} \begin{pmatrix} \lambda_p \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q & \mathbf{0} \end{pmatrix} = \boldsymbol{\Sigma}_{pz} - \begin{pmatrix} \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q & 0 \\ 0 & 0 \end{pmatrix}$$
$$= \begin{pmatrix} Var[p] - \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q & -\lambda_p \psi_I^{-1} \sigma_z^2 \\ -\lambda_p \psi_I^{-1} \sigma_z^2 & \sigma_z^2 \end{pmatrix}$$

so

$$\begin{aligned} |\boldsymbol{\Sigma}_{pz|\hat{\beta}}| &= (Var[p] - \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q) \sigma_z^2 - \lambda_p^2 \psi_I^{-2} \sigma_z^4 \\ &= \lambda_p^2 Var[\hat{y}_I] \sigma_z^2 + \lambda_p^2 \psi_I^{-2} \sigma_z^4 - \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q \sigma_z^2 - \lambda_p^2 \psi_I^{-2} \sigma_z^4 \\ &= \lambda_p^2 (Var[\hat{y}_I] - \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q) \sigma_z^2 \end{aligned}$$

and

$$\boldsymbol{\Sigma}^{-1}_{pz|\hat{\beta}} = \frac{1}{\lambda_p^2 (Var[\hat{y}_I] - \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q) \sigma_z^2} \begin{pmatrix} \sigma_z^2 & \lambda_p \psi_I^{-1} \sigma_z^2 \\ \lambda_p \psi_I^{-1} \sigma_z^2 & Var[p] - \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q. \end{pmatrix}$$

Notice that

$$\begin{aligned} -\boldsymbol{\Sigma}^{-1}_{pz|\hat{\beta}} \begin{pmatrix} \lambda_p \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \\ \mathbf{0}^\top \end{pmatrix} \boldsymbol{D}^{-2}_{\sigma_\beta} &= -|\boldsymbol{\Sigma}_{pz|\hat{\beta}}|^{-1} \begin{pmatrix} \sigma_z^2 & \lambda_p \psi_I^{-1} \sigma_z^2 \\ \lambda_p \psi_I^{-1} \sigma_z^2 & Var[p] - \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q \end{pmatrix} \begin{pmatrix} \lambda_p \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \\ \mathbf{0}^\top \end{pmatrix} \boldsymbol{D}^{-2}_{\sigma_\beta} \\ &= -|\boldsymbol{\Sigma}_{pz|\hat{\beta}}|^{-1} \begin{pmatrix} \sigma_z^2 \\ \lambda_p \psi_I^{-1} \sigma_z^2 \end{pmatrix} \lambda_p \boldsymbol{\mu}_q^\top, \end{aligned}$$

and

$$\begin{aligned} -\boldsymbol{D}^{-2}_{\sigma_\beta} \begin{pmatrix} \lambda_p \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q & \mathbf{0} \end{pmatrix} \boldsymbol{\Sigma}^{-1}_{pz|\hat{\beta}} &= -|\boldsymbol{\Sigma}_{pz|\hat{\beta}}|^{-1} \boldsymbol{D}^{-2}_{\sigma_\beta} \begin{pmatrix} \lambda_p \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q & \mathbf{0} \end{pmatrix} \begin{pmatrix} \sigma_z^2 & \lambda_p \psi_I^{-1} \sigma_z^2 \\ \lambda_p \psi_I^{-1} \sigma_z^2 & Var[p] - \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q \end{pmatrix} \\ &= -|\boldsymbol{\Sigma}_{pz|\hat{\beta}}|^{-1} \lambda_p \boldsymbol{\mu}_q \begin{pmatrix} \sigma_z^2 & \lambda_p \psi_I^{-1} \sigma_z^2 \end{pmatrix}, \end{aligned}$$

as well as

$$|\boldsymbol{\Sigma}_{pz|\hat{\beta}}|^{-1} \begin{pmatrix} \lambda_p \boldsymbol{\mu}_q & \mathbf{0} \end{pmatrix} \begin{pmatrix} \sigma_z^2 \\ \lambda_p \psi_I^{-1} \sigma_z^2 \end{pmatrix} \lambda_p \boldsymbol{\mu}_q^\top = |\boldsymbol{\Sigma}_{pz|\hat{\beta}}|^{-1} \lambda_p^2 \boldsymbol{\mu}_q \boldsymbol{\mu}_q^\top \sigma_z^2.$$

Moments: variance

$$\begin{aligned} Var\left[p, z, \hat{\boldsymbol{\beta}}\right]^{-1} &= \begin{pmatrix} \boldsymbol{\Sigma}_{pz} & \begin{pmatrix} \lambda_p \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \\ \mathbf{0}^\top \end{pmatrix} \\ \begin{pmatrix} \lambda_p \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q & \mathbf{0} \end{pmatrix} & \boldsymbol{D}^2_{\sigma_\beta} \end{pmatrix}^{-1} \\ &= \begin{pmatrix} 0 \boldsymbol{I}_{2\times 2} & 0 \boldsymbol{I}_{2\times n} \\ 0 \boldsymbol{I}_{n\times 2} & \boldsymbol{D}^{-2}_{\sigma_\beta} \end{pmatrix} + |\boldsymbol{\Sigma}_{pz|\hat{\beta}}|^{-1} \begin{pmatrix} \boldsymbol{\Sigma}^{\text{adj}}_{pz|\hat{\beta}} & -\lambda_p \begin{pmatrix} \sigma_z^2 \\ \lambda_p \psi_I^{-1} \sigma_z^2 \end{pmatrix} \boldsymbol{\mu}_q^\top \\ -\lambda_p \boldsymbol{\mu}_q \begin{pmatrix} \sigma_z^2 & \lambda_p \psi_I^{-1} \sigma_z^2 \end{pmatrix} & \lambda_p^2 \boldsymbol{\mu}_q \boldsymbol{\mu}_q^\top \sigma_z^2 \end{pmatrix} \end{aligned}$$

covariance

$$\begin{aligned} Cov\left[r, \begin{pmatrix} p & z & \hat{\boldsymbol{\beta}} \end{pmatrix}\right] &= \begin{pmatrix} Cov\left[r, \begin{pmatrix} p & z \end{pmatrix}\right] & -\lambda_p \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \end{pmatrix} \\ &= \begin{pmatrix} \lambda_p (Cov[y, \hat{y}_I] - \lambda_p Var[s_U]) & \lambda_p \psi_I^{-1} \sigma_z^2 & -\lambda_p \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \end{pmatrix} - \end{aligned}$$

Coefficients

$$Cov\left[r, \begin{pmatrix} p & z & \hat{\boldsymbol{\beta}} \end{pmatrix}\right] Var\left[p, z, \hat{\boldsymbol{\beta}}\right]^{-1}$$

$$
\begin{aligned}
&= \begin{pmatrix} \mathbf{0}^\top & -\boldsymbol{\mu}_q^\top \end{pmatrix} \\
&\quad + |\boldsymbol{\Sigma}_{pz|\hat{\beta}}|^{-1} \Bigg(Cov\left[r, \begin{pmatrix} p & z \end{pmatrix}\right] \boldsymbol{\Sigma}^{\text{adj}}_{pz|\hat{\beta}} + \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q \begin{pmatrix} \sigma_z^2 & \lambda_p \psi_I^{-1} \sigma_z^2 \end{pmatrix} \\
&\qquad\qquad - \lambda_p Cov\left[r, \begin{pmatrix} p & z \end{pmatrix}\right] \begin{pmatrix} \sigma_z^2 \\ \lambda_p \psi_I^{-1} \sigma_z^2 \end{pmatrix} \boldsymbol{\mu}_q^\top - \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q \boldsymbol{\mu}_q^\top \sigma_z^2 \Bigg)
\end{aligned}
$$

where

$$
\begin{aligned}
&|\boldsymbol{\Sigma}_{pz|\hat{\beta}}|^{-1} \left(Cov\left[r, \begin{pmatrix} p & z \end{pmatrix}\right] \boldsymbol{\Sigma}^{\text{adj}}_{pz|\hat{\beta}} + \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q \begin{pmatrix} \sigma_z^2 & \lambda_p \psi_I^{-1} \sigma_z^2 \end{pmatrix} \right) \begin{pmatrix} p - E[p] \\ z \end{pmatrix} \\
&= |\boldsymbol{\Sigma}_{pz|\hat{\beta}}|^{-1} \Bigg(\begin{pmatrix} \lambda_p (Cov[y, \hat{y}_I] - \lambda_p Var[s_U]) & \lambda_p \psi_I^{-1} \sigma_z^2 \end{pmatrix} \begin{pmatrix} \sigma_z^2 & \lambda_p \psi_I^{-1} \sigma_z^2 \\ \lambda_p \psi_I^{-1} \sigma_z^2 & Var[p] - \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q \end{pmatrix} \\
&+ \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q \begin{pmatrix} \sigma_z^2 & \lambda_p \psi_I^{-1} \sigma_z^2 \end{pmatrix} \Bigg) \begin{pmatrix} p - E[p] \\ z \end{pmatrix} \\
&= \frac{Cov[y, \hat{y}_I] - \lambda_p (Var[\hat{y}_I] + \psi_I^{-1} \sigma_z^2) + \lambda_p \psi_I^{-1} \sigma_z^2 + \lambda_p \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q}{\lambda_p^2 (Var[\hat{y}_I] - \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q) \sigma_z^2} \lambda_p \sigma_z^2 (p - E[p]) \\
&\quad + \lambda_p \psi_I^{-1} \sigma_z^2 \frac{\lambda_p Cov[y, \hat{y}_I] - Var[p] + Var[p] - \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q + \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q}{\lambda_p^2 (Var[\hat{y}_I] - \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q) \sigma_z^2} z \\
&= \frac{Cov[y, \hat{y}_I] - \lambda_p (Var[\hat{y}_I] - \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q)}{\lambda_p (Var[\hat{y}_I] - \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q)} (p - E[p]) - \frac{Cov[y, \hat{y}_I]}{\lambda_p (Var[\hat{y}_I] - \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q)} (-\psi_I^{-1} z)
\end{aligned}
$$

and

$$
\begin{aligned}
&- \left\{ 1 + \sigma_z^2 \frac{\lambda_p Cov[y, \hat{y}_I] - \lambda_p^2 Var[s_U] + \lambda_p^2 \psi_I^{-2} \sigma_z^2 + \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q}{\lambda_p^2 (Var[\hat{y}_I] - \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q) \sigma_z^2} \right\} \boldsymbol{\mu}_q^\top (\hat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta) \\
&= - \frac{Cov[y, \hat{y}_I]}{\lambda_p (Var[\hat{y}_I] - \boldsymbol{\mu}_q^\top \boldsymbol{D}^2_{\sigma_\beta} \boldsymbol{\mu}_q)} \boldsymbol{\mu}_q^\top (\hat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta).
\end{aligned}
$$

A.3.2 Trading volume

Realized trading volume v is given by

$$
\begin{aligned}
v &= \frac{1}{2} \{ |\delta_I| + |\delta_U| + |z| \} \\
&= \frac{1}{2} \left\{ \psi_I \left| (1 - \lambda_p)(\hat{y}_I - E[\hat{y}_I]) - \lambda_p \psi_I^{-1} z \right| + \psi_U (\lambda_p - \lambda_U) \left| (E[\hat{y}_I] - \hat{y}_I - \psi_I^{-1} z) \right| + |z| \right\} \\
&= \frac{1}{2} \left\{ \psi_I \left| (1 - \lambda_p)(\hat{y}_I - E[\hat{y}_I]) - \lambda_p \psi_I^{-1} z \right| + \psi_I (1 - \lambda_p) \left| -(\hat{y}_I - E[\hat{y}_I]) - \psi_I^{-1} z \right| + |z| \right\} \\
&= \frac{\psi_I}{2} \left\{ \left| (1 - \lambda_p)(\hat{y}_I - E[\hat{y}_I]) - \lambda_p \psi_I^{-1} z \right| + (1 - \lambda_p) \left| (\hat{y}_I - E[\hat{y}_I]) + \psi_I^{-1} z \right| + \psi_I^{-1} |z| \right\},
\end{aligned}
$$

which in the third line uses the equality

$$
\psi_U (\lambda_p - \lambda_U) = \psi_U \frac{\psi_I + \lambda_U \psi_U - \lambda_U (\psi_I + \psi_U)}{\psi_I + \psi_U} = \psi_I \frac{(1 - \lambda_U) \psi_U}{\psi_I + \psi_U} = \psi_I (1 - \lambda_p).
$$

A.3.3 Expected profit and ex-ante expected utility

Expected profit

Using that

$$
\begin{aligned}
&\hat{y}_I - p = (1-\lambda_p)\hat{y}_I - (1-\lambda_p)\boldsymbol{\mu}_\beta^\top \boldsymbol{\mu}_q - \lambda_p \psi_I^{-1} z = (1-\lambda_p)(\hat{y}_I - E[\hat{y}_I]) - \lambda_p \psi_I^{-1} z \\
&\implies E[\hat{y}_I - p] = 0 \\
&\implies E[(\hat{y}_I - p)^2] = Var[\hat{y}_I - p] = (1-\lambda_p)^2 Var[\hat{y}_I] + \lambda_p^2 \psi_I^{-2} \sigma_z^2
\end{aligned}
$$

key expectations are

$$
\begin{aligned}
E[(\hat{y}_I - p)y] &= (1-\lambda_p)Cov[y, \hat{y}_I] = (1-\lambda_p)\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\mu}_\beta \\
E[(\hat{y}_I - p)p] &= Cov[\hat{y}_I, p] - Var[p] = \lambda_p Var[\hat{y}_I] - \lambda_p^2 \left\{Var[\hat{y}_I] + \psi_I^{-2}\sigma_z^2\right\} \\
&= \lambda_p(1-\lambda_p)Var[\hat{y}_I] - \lambda_p^2 \psi_I^{-2}\sigma_z^2
\end{aligned}
$$

and informed profit under the objective measure is

$$
\begin{aligned}
E[\pi_I] &= \psi_I E\left[(\hat{y}_I - p)(y-p)\right] = \psi_I \left\{E[(\hat{y}_I - p)y] - E[(\hat{y}_I - p)p]\right\} \\
&= \psi_I \left\{(1-\lambda_p)\left(Cov[y,\hat{y}_I] - \lambda_p Var[\hat{y}_I]\right) + \lambda_p^2 \psi_I^{-2}\sigma_z^2\right\}.
\end{aligned}
$$

Under the contemporary measure $E[\boldsymbol{\beta}|\boldsymbol{c}^*] = \boldsymbol{\mu}_\beta$, covariance is the quadratic form $Cov[y, \hat{y}_I|\boldsymbol{c}^*] = \boldsymbol{\mu}_\beta^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\mu}_\beta$, profits are

$$
E[\pi_I|\boldsymbol{c}^*] = \psi_I \left\{(1-\lambda_p)\left(Cov[y,\hat{y}_I|\boldsymbol{c}^*] - \lambda_p Var[\hat{y}_I]\right) + \lambda_p^2 \psi_I^{-2}\sigma_z^2\right\},
$$

and out of sample surprise is

$$
E[\pi_I] - E[\pi_I|\boldsymbol{c}^*] = \psi_I(1-\lambda_p)(Cov[y,\hat{y}_I] - Cov[y,\hat{y}_I|\boldsymbol{c}^*]).
$$

Similar analysis for the uninformed profit yields

$$
\begin{aligned}
\hat{y}_U - p = (\lambda_p - \lambda_U)(E[\hat{y}_I] - s_U) &\implies E[\hat{y}_U - p] = 0 \\
&\implies E[(\hat{y}_U - p)^2] = (\lambda_p - \lambda_U)^2 Var[s_U]
\end{aligned}
$$

so

$$
E[(\hat{y}_U - p)y] = -(\lambda_p - \lambda_U)Cov[y,\hat{y}_I], \quad E[(\hat{y}_U - p)p] = -(\lambda_p - \lambda_U)\lambda_p Var[s_U]
$$

and expected profit under the objective measure is

$$
\begin{aligned}
E[\pi_U] &= \psi_U E\left[(\hat{y}_U - p)(y-p)\right] = \psi_U \left\{E[(\hat{y}_U - p)y] - E[(\hat{y}_U - p)p]\right\} \\
&= \psi_U \left\{(\lambda_p - \lambda_U)\left(\lambda_p Var[\hat{y}_I] - Cov[y,\hat{y}_I]\right) + (\lambda_p - \lambda_U)\lambda_p \psi_I^{-2}\sigma_z^2\right\}.
\end{aligned}
$$

Notice that

$$
\psi_U(\lambda_p - \lambda_U) = \psi_U \frac{\psi_I + \lambda_U \psi_U - \lambda_U(\psi_I + \psi_U)}{\psi_I + \psi_U} = \psi_I \frac{(1-\lambda_U)\psi_U}{\psi_I + \psi_U} = \psi_I(1-\lambda_p)
$$

so out of sample surprise is

$$
E[\pi_U] - E[\pi_U|\boldsymbol{c}^*] = \psi_I(1-\lambda_p)(Cov[y,\hat{y}_I|\boldsymbol{c}^*] - Cov[y,\hat{y}_I]).
$$

Difference in profit under contemporaneous measure

Expected difference in profit under the contemporaneous measure is

$$\begin{aligned} E[\pi_I|\boldsymbol{c}^*] - E[\pi_U|\boldsymbol{c}^*] &= \psi_I \left\{(1-\lambda_p)\left(Cov[y,\hat{y}_I|\boldsymbol{c}^*] - \lambda_p Var[\hat{y}_I]\right) + \lambda_p^2 \psi_I^{-2}\sigma_z^2\right\} \\ &\quad - \psi_I(1-\lambda_p)\left\{\left(\lambda_p Var[\hat{y}_I] - Cov[y,\hat{y}_I|\boldsymbol{c}^*]\right) + \lambda_p \psi_I^{-2}\sigma_z^2\right\} \\ &= \psi_I \left\{2(1-\lambda_p)\left(Cov[y,\hat{y}_I|\boldsymbol{c}^*] - \lambda_p Var[\hat{y}_I]\right) + \left[2\lambda_p - 1\right]\lambda_p \psi_I^{-2}\sigma_z^2\right\} \end{aligned}$$

A necessary condition for the differential to be positive is that the last term in the curly bracket, $\left[2\lambda_p - 1\right]\lambda_p \psi_I^{-2}\sigma_z^2$, which require that informed investors trade more aggresively $\psi_I > \psi_U$, since

$$2\lambda_p > 1 \iff 2\psi_I + 2\lambda_U\psi_U > \psi_I + \psi_U \iff \psi_I + (2\lambda_U - 1)\psi_U > 0 \impliedby \psi_I > \psi_U.$$

For symmetric uncertainty aversion, this simplifies to the informed making better prediction (see Proposition 3).

Ex-post performance

Ex-post performance surprises are symmetric and only exist with non-zero bias

$$\begin{aligned} &E[\pi_I] - E[\pi_I|\boldsymbol{c}^*] = \psi_I(1-\lambda_p)\left\{Cov[y,\hat{y}] - Cov[y,\hat{y}|\boldsymbol{c}^*]\right\} = \psi_I(1-\lambda_p)\left\{\boldsymbol{\varepsilon}_\beta \boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta - \boldsymbol{\beta}\boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta\right\}, \\ &\text{and} \quad E[\pi_U] - E[\pi_U|\boldsymbol{c}^*] = -\psi_I(1-\lambda_p)\left\{\boldsymbol{\varepsilon}_\beta \boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta - \boldsymbol{\beta}\boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta\right\}. \end{aligned}$$

By algebraic manipulation

$$Cov[y,\hat{y}] - Cov[y,\hat{y}|\boldsymbol{c}^*] = \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\mu}_\beta - \boldsymbol{\mu}_\beta^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\mu}_\beta = \boldsymbol{\varepsilon}_\beta \boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta - \boldsymbol{\beta}\boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta$$

and

$$\psi_U(\lambda_p - \lambda_U) = \psi_U \frac{\psi_I + \lambda_U\psi_U - \lambda_U(\psi_I + \psi_U)}{\psi_I + \psi_U} = \psi_I \frac{(1-\lambda_U)\psi_U}{\psi_I + \psi_U} = \psi_I(1-\lambda_p).$$

Of the two components of the cost of complexity, ex-post performance surprises are entirely driven by the bias, and while the sign of the first term in the curly bracket could be both negative or positive, the quadratic form is always positive due to the positive definiteness of $\boldsymbol{\Sigma}_\zeta$, suggests that the term might be positive more often than not.

Ex-post performance surprises are a transfer between investors, and due to their symmetry it leaves certain results from the baseline model unaltered regardless of its sign. Ex-post performance surprises are a transfer between investors and nets out in total

$$E[\pi_I] - E[\pi_I|\boldsymbol{c}^*] - (E[\pi_U] - E[\pi_U|\boldsymbol{c}^*]) = 0.$$

It follows that for matters concerning the total profits of investors the distinction between objective measure and contemporary measure is irrelevant. The corresponding result in the baseline model arises trivially because $Var[E[y|\boldsymbol{\beta},\boldsymbol{s}_I]] = Cov[y, E[y|\boldsymbol{\beta},\boldsymbol{s}_I]]$ so there are no ex-post surprises. Due to the common component in expected profit, total profit of investors simplifies to

$$E[\pi_I] + E[\pi_U] = E[\pi_I|\boldsymbol{c}^*] + E[\pi_U|\boldsymbol{c}^*] = \psi_I^{-1}\sigma_z^2 = \alpha_I E[(y-\hat{y}_I)^2]\sigma_z^2.$$

This result mirrors a result in the baseline model where it holds with the modification that the mean squared error collapses to the conditional variance under the true model. Introducing noisy estimation does not alter the intuition of the baseline model that total profits are increasing in the quality of predictions made by informed investors.

A.4 Value of data

A.4.1 Rational function formulation of cost of complexity

Let the adjacency matrix of any matrix $\boldsymbol{A}$ be indicated by superscript $\boldsymbol{A}^{\text{adj}}$ and the determinant by $|\boldsymbol{A}|$, and define

$$\boldsymbol{W} := k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta} = k_c^{-2}\boldsymbol{\Omega}_{\zeta 0} + \boldsymbol{D}_{\Omega_\zeta 0} + k_S(k_c^{-2}\boldsymbol{S} + \boldsymbol{D}_S) := \boldsymbol{W}_0 + k_S\boldsymbol{W}_S$$

such that

$$\begin{aligned}
\mathbf{1}^\top \boldsymbol{X}^{-1}\mathbf{1} &= \mathbf{1}^\top \left\{k_c^2\boldsymbol{\Omega}_\zeta^{-1} + \boldsymbol{D}_{\Omega_\zeta}^{-1}\right\}^{-1}\mathbf{1} = \mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta}\mathbf{1} - \mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta}\left\{k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta}\right\}^{-1}\boldsymbol{D}_{\Omega_\zeta}\mathbf{1} \\
&= \mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta}\mathbf{1} - \mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta}\boldsymbol{W}^{-1}\boldsymbol{D}_{\Omega_\zeta} = \boldsymbol{D}_{\Omega_\zeta}\mathbf{1} - |\boldsymbol{W}|^{-1}\mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta}\boldsymbol{W}^{\text{adj}}\boldsymbol{D}_{\Omega_\zeta}\mathbf{1} \\
&= \mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta}\mathbf{1} \\
&\quad - |\boldsymbol{W}|^{-1}\left\{\mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta 0}\boldsymbol{W}^{\text{adj}}\boldsymbol{D}_{\Omega_\zeta 0}\mathbf{1} + k_S^2\mathbf{1}^\top\boldsymbol{D}_S\boldsymbol{W}^{\text{adj}}\boldsymbol{D}_S\mathbf{1} + 2k_S\mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta 0}\boldsymbol{W}^{\text{adj}}\boldsymbol{D}_S\mathbf{1}\right\}.
\end{aligned}$$

For a sum of matrices where one is scaled by a scalar k such as $\boldsymbol{W}$, it can be shown, see Appendix A.4.2, that the determinant is a polynomial in k of degree equal to the number of rows (or columns) of the matrix, and that the sum over its adjugate matrix is a polynomial in k of one degree less. Scaling the entries of the adjugate matrix before taking the sum, as is done in the terms of the curly bracket above, does not change the degree of the resulting polynomial. However, the multiplication of the second term by the square of k_S, which is the variable of the polynomial for $\boldsymbol{W}$, yields a polynomial of a degree one *higher* than the number of rows, which for $\boldsymbol{W}$ is equal to the number of factor n_q. The curly bracket divided by the determinant is therefore a rational function (ratio of polynomials) of k_S. By extension, the cost of complexity is a rational function of k_S

$$\begin{aligned}
\chi(k_S) = k_{\sigma 0}^2|\boldsymbol{W}|^{-1}\Big\{&|\boldsymbol{W}|\mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta 0}\mathbf{1} + k_S|\boldsymbol{W}|\mathbf{1}^\top\boldsymbol{D}_S\mathbf{1} - \mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta 0}\boldsymbol{W}^{\text{adj}}\boldsymbol{D}_{\Omega_\zeta 0}\mathbf{1} \\
&- k_S^2\mathbf{1}^\top\boldsymbol{D}_S\boldsymbol{W}^{\text{adj}}\boldsymbol{D}_S\mathbf{1} - 2k_S\mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta 0}\boldsymbol{W}^{\text{adj}}\boldsymbol{D}_S\mathbf{1}\Big\} = k_{\sigma 0}^2\frac{\sum_{\ell=0}^{n_q+1} a_\ell k_S^\ell}{\sum_{\ell=0}^{n_q} b_\ell k_S^\ell}.
\end{aligned}$$

The degree of the polynomial in the numerator is exactly one higher than the degree of the polynomial in the denominator and the rational function therefore has a an oblique asymptote, which is linear in k_S. The slope of the asymptote is the coefficient of the highest power of the polynomial in the numerator divided by the coefficient of the highest power in the denominator a_{n_q+1}/b_{n_q}, or the limit of the derivative with respect to k_S, i.e. $\lim_{k_S\to\infty}\partial\chi(k_S)/\partial k_S$. A special case is that of a diagonal $\boldsymbol{S}$ matrix, in which case the derivative simplifies to

$$\lim_{k_S\to\infty}\frac{\partial\chi(k_S)}{\partial k_S} = \frac{k_{\sigma 0}^2}{1+k_c^2}\mathbf{1}^\top\boldsymbol{D}_S\mathbf{1},$$

and in the even more special case where all true factor loadings are the same $\boldsymbol{\beta} = \bar{\beta}\mathbf{1}$, whether more data (stronger signal) is asymptotically valuable or value destroying is entirely determined by the base-parameters $k_{\sigma 0}$, k_c and $\bar{\beta}$, specifically

$$\lim_{k_S\to\infty}\frac{\partial\chi(k_S)}{\partial k_S} + \frac{\partial Var[y|\boldsymbol{\beta},\boldsymbol{s}_I]}{\partial k_S} = \mathbf{1}^\top\boldsymbol{D}_S\mathbf{1}\left\{\frac{k_{\sigma 0}^2}{1+k_c^2} - \bar{\beta}^2\right\}.$$

A.4.2 Matrix inverses as rational functions

If a square matrix $\boldsymbol{L} = \boldsymbol{A} + k\boldsymbol{B}$ is of dimensions two by two matrix, its determinant is a polynomial in k of the form

$$|\boldsymbol{L}_{2\times 2}| = |\boldsymbol{A}| + k^2|\boldsymbol{B}| + k(a_{11}b_{22} + a_{22}b_{11} - a_{12}b_{21} - a_{21}b_{12}).$$

Let $\boldsymbol{M}_{ij}$ be the $(n-1)\times(n-1)$ sub-matrix of the square $n \times n$ matrix $\boldsymbol{L}$ that deletes row i and column j, i.e. its determinant $|\boldsymbol{M}_{ij}|$ is the ij-minor of $\boldsymbol{L}$. By a Laplace expansion over row i, the determinant of $\boldsymbol{L}$ is given by $|\boldsymbol{L}| = \sum_j (-1)^{i+j} l_{ij}|\boldsymbol{M}_{ij}|$, where the entries of $\boldsymbol{L}$ are of the form $l_{ij} = a_{ij} + kb_{ij}$. Therefore, if $|\boldsymbol{M}_{ij}|$ is a polynomial in k, the product $l_{ij}|\boldsymbol{M}_{ij}|$ is a polynomial in k of one degree higher. Since $|\boldsymbol{L}_{2\times 2}|$ is a polynomial of second degree, the determinant of $\boldsymbol{L}_{n\times n}$ is a polynomial of n'th degree. Meanwhile, the co-factor matrix of $\boldsymbol{L}$ denoted $\boldsymbol{C}$ has entries $c_{ij} = (-1)^{i+j} l_{ij}|\boldsymbol{M}_{ij}|$ which are polynomials in k of degree $n-1$. The sum over the adjugate matrix $\boldsymbol{1}^\top \boldsymbol{L}^{\text{adj}} \boldsymbol{1} = \boldsymbol{1}^\top \boldsymbol{C}^\top \boldsymbol{1}$ is, therefore, a polynomial of degree $n-1$.

Appendix to Chapter 2

B.1 Computation of value functions

B.1.1 Value function after shock

Let $W(m) = W(m;\overline{m}^*)$ denote the value function after the shock. Note that the value function is independent of the bank's type, as the market learns the bank's type when the shock hits. The ordinary differential equation corresponding to the cash-flows dynamics is:

$$\rho W(m) = \mu W'(m) + 0.5\sigma^2 W''(m).$$

The characteristic roots $\overline{r}_1$ and $\overline{r}_2$ are given by characteristic equation

$$\rho - \mu r - 0.5\sigma^2 r^2 = 0 \implies \overline{r}_1 = \frac{-\mu - \sqrt{\mu^2 + 2\sigma^2\rho}}{\sigma^2}, \quad \overline{r}_2 = \frac{-\mu + \sqrt{\mu^2 + 2\sigma^2\rho}}{\sigma^2}.$$

As a result, the value function can be written as

$$W(m) = \begin{cases} \overline{A}_1 e^{\overline{r}_1(m-\overline{m}^*)} + \overline{A}_2 e^{\overline{r}_2(m-\overline{m}^*)}, & \text{for } m < \overline{m}^*, \\ m - \overline{m}^* + \overline{A}_1 + \overline{A}_2, & \text{for } m \geq \overline{m}^*. \end{cases}$$

The coefficients $\overline{A}_1$ and $\overline{A}_2$ are pinned down by smooth pasting and high-contact conditions:

$$W'(\overline{m}^*) = 1, \quad W''(\overline{m}^*) = 0,$$

which results in

$$\overline{A}_1 = \frac{\overline{r}_2^2}{\overline{r}_1\overline{r}_2(\overline{r}_2 - \overline{r}_1)}, \quad \overline{A}_2 = \frac{-\overline{r}_1^2}{\overline{r}_1\overline{r}_2(\overline{r}_2 - \overline{r}_1)}.$$

Furthermore, the optimal dividend boundary is pinned down by value matching at 0:

$$\overline{m}^* = \frac{2}{\overline{r}_2 - \overline{r}_1} \log\left(-\frac{\overline{r}_1}{\overline{r}_2}\right). \tag{B.1}$$

Note that at this optimal boundary, $W(\overline{m}^*) = \overline{A}_1 + \overline{A}_2 = \mu/\rho$.

B.1.2 Value function bad bank before shock

We will now compute the value function of a bad bank before the arrival of the liquidity shock for a given dividend policy m_B. We will distinguish three cases that depend on the relative position of the dividend policy before and after the shock.

Case i) $m_B - f \in (0, \overline{m}^*)$

The bank defaults when the bank's cash reserves are smaller than the shock, i.e. $m < f$. In this region, the ordinary differential equation is:

$$\rho V_B(m) = \mu V_B'(m) + 0.5\sigma^2 V_B''(m) - \lambda V_B(m),$$

so that the solution can be written as

$$V_B(m) = \sum_{i=1}^{2} A_i^B e^{r_i(m-m_B)},$$

where r_1 and r_2 are the solution to characteristic equation

$$(\rho+\lambda) - \mu r - 0.5\sigma^2 r^2 = 0 \implies r_1 = \frac{-\mu - \sqrt{\mu^2 + 2\sigma^2(\rho+\lambda)}}{\sigma^2}, \quad r_2 = \frac{-\mu + \sqrt{\mu^2 + 2\sigma^2(\rho+\lambda)}}{\sigma^2}.$$

For the region $m \in [f, m_B]$ the ordinary differential equation is

$$\rho V_B(m) = \mu V_B'(m) + 0.5\sigma^2 V_B''(m) + \lambda(W(m-f) - V_B(m)).$$

The homogeneous solution can be written as

$$V_B^h(m) = \sum_{i=1}^{2} B_i^B e^{r_i(m-m_B)},$$

To find the particular solution, plug in conjecture $\sum_{i=1}^{2} \alpha_i \overline{A}_i e^{\overline{r}_i(m-f)}$ in the ODE:

$$\sum_{i=1}^{2} (\rho+\lambda)\alpha_i \overline{A}_i e^{\overline{r}_i(m-f)} = \sum_{i=1}^{2} \mu\alpha_i \overline{r}_i \overline{A}_i e^{\overline{r}_i(m-f)} + 0.5\sigma^2 \alpha_i r_i^2 \overline{A}_i e^{\overline{r}_i(m-f)} + \lambda \overline{A}_i e^{\overline{r}_i(m-f)},$$

where α_i, $i \in \{1,2\}$ can be simplified as follows:

$$(\rho + \lambda - \mu\overline{r}_i - 0.5\sigma^2 \overline{r}_i^2)\alpha_i = \lambda \implies \alpha_i = 1.$$

Bringing this together:

$$V_B(m) = V_B(m; m_B) = \begin{cases} \sum_{i=1}^{2} A_i^B e^{r_i(m-m_B)}, & \text{for } m < f, \\ \sum_{i=1}^{2} B_i^B e^{r_i(m-m_B)} + W(m-f), & \text{for } m \in [f, m_B), \\ \sum_{i=1}^{2} B_i^B + W(m_B - f) + m - m_B, & \text{for } m \geq m_B. \end{cases} \tag{B.2}$$

The coefficients A_i^B and B_i^B are pinned down by the following boundary equations:

- Value matching at $m = 0$: $\lim_{m \downarrow 0} V_B(m) = 0$;
- Value matching at $m = f$: $\lim_{m \uparrow f} V_B(m) = \lim_{m \downarrow f} V_B(m)$;
- Smooth pasting at $m = f$: $\lim_{m \uparrow f} V_B'(m) = \lim_{m \downarrow f} V_B'(m)$;
- Smooth pasting at $m = m_B$: $\lim_{m \uparrow m_B} V_B'(m) = 1$.

Denote $\mathbf{A^B} = \begin{bmatrix} A_1^B & A_2^B \end{bmatrix}^\top$ and $\mathbf{B^B} = \begin{bmatrix} B_1^B & B_2^B \end{bmatrix}^\top$. These boundary equations can be summarized into the following system of equations:

$$\begin{bmatrix} \mathbf{A^B} \\ \mathbf{B^B} \end{bmatrix} = (M_1^B)^{-1} v_1^B,$$

where

$$M_1^B := \begin{bmatrix} e^{-r_1 m_B} & e^{-r_2 m_B} & 0 & 0 \\ e^{r_1(f-m_B)} & e^{r_2(f-m_B)} & -e^{r_1(f-m_B)} & -e^{r_2(f-m_B)} \\ r_1 e^{r_1(f-m_B)} & r_2 e^{r_2(f-m_B)} & -r_1 e^{r_1(f-m_B)} & -r_2 e^{r_2(f-m_B)} \\ 0 & 0 & r_1 & r_2 \end{bmatrix}, \quad v_1^B := \begin{bmatrix} 0 \\ 0 \\ W'(0) \\ 1 - W'(m_B - f) \end{bmatrix}.$$

Case ii) $m_B - f \geq \overline{m}^*$

In the region $m < f$, the ordinary differential equation and solution is the same as in case i). In the region $m \in [f, f + \overline{m}^*)$, the ordinary differential equation is again given by

$$(\rho + \lambda) V_B(m) = \mu V_B'(m) + 0.5\sigma^2 V_B''(m) + \lambda W(m - f).$$

In the region $m \in [\overline{m}^* + f, m_B)$, the ordinary differential equation is given by

$$(\rho + \lambda) V_B(m) = \mu V_B'(m) + 0.5\sigma^2 V_B''(m) + \lambda \left(m - f - \overline{m}^* + \frac{\mu}{\rho} \right).$$

Conjecture the following solution to the particular solution: $V_B^p(m) = \beta_B + \gamma m$, and plug this into the ordinary differential equation:

$$(\rho + \lambda)(\beta_B + \gamma m) = \mu\gamma + \lambda \left(m - f - \overline{m}^* + \frac{\mu}{\rho} \right).$$

Solving for β_B and γ gives

$$\begin{aligned} (\rho + \lambda)\gamma = \lambda &\implies \gamma = \frac{\lambda}{\rho + \lambda}, \\ (\rho + \lambda)\beta_B = \mu\gamma + \lambda \left(-f - \overline{m}^* + \frac{\mu}{\rho} \right) &\implies \beta_B = \gamma \left(\frac{\mu}{\rho + \lambda} + \frac{\mu}{\rho} - f - \overline{m}^* \right). \end{aligned}$$

Bringing all of this together the following value function:

$$V_B(m) = \begin{cases} \sum_{i=1}^2 A_i^B e^{r_i(m - m_B)}, & \text{for } m \in [0, f), \\ \sum_{i=1}^2 B_i^B e^{r_i(m - m_B)} + W(m - f), & \text{for } m \in [f, f + \overline{m}^*], \\ \sum_{i=1}^2 C_i^B e^{r_i(m - m_B)} + \beta_B + \gamma m, & \text{for } m \in [f + \overline{m}^*, m_B), \\ \sum_{i=1}^2 C_i^B + \beta_B + (\gamma - 1) m_B + m, & \text{for } m \geq m_B. \end{cases}$$

The coefficients A_i^B, B_i^B, C_i^B, for $i \in \{1, 2\}$ are solved by the following boundary conditions:

- Value matching at $m = 0$: $\lim_{m \downarrow 0} V_B(m) = 0$;
- Value matching at $m = f$: $\lim_{m \downarrow f} V_B(m) = \lim_{m \uparrow f} V_B(m)$;
- Smooth pasting at $m = f$: $\lim_{m \downarrow f} V_B'(m) = \lim_{m \uparrow f} V_B'(m)$;
- Value matching at $m = f + \overline{m}^*$: $\lim_{m \downarrow f + \overline{m}^*} V_B(m) = \lim_{m \uparrow f + \overline{m}^*} V_B(m)$;
- Smooth pasting at $m = f + \overline{m}^*$: $\lim_{m \downarrow f + \overline{m}^*} V_B'(m) = \lim_{m \uparrow f + \overline{m}^*} V_B'(m)$;

- Smooth pasting at $m = m_B$: $\lim_{m\uparrow m_B} V_B'(m) = 1$.

Denote $\mathbf{A^B} = \begin{bmatrix} A_1^B & A_2^B \end{bmatrix}^\top$, $\mathbf{B^B} = \begin{bmatrix} B_1^B & B_2^B \end{bmatrix}^\top$ and $\mathbf{C^B} = \begin{bmatrix} C_1^B & C_2^B \end{bmatrix}^\top$. The system of equations can then be summarized as follows:

$$\begin{bmatrix} \mathbf{A^B} \\ \mathbf{B^B} \\ \mathbf{C^B} \end{bmatrix} = (M_2^B)^{-1} v_2^B, \quad \text{where } v_2^B := \begin{bmatrix} 0 \\ 0 \\ W'(0) \\ \frac{\mu}{\rho+\lambda}(\gamma-1) \\ \gamma-1 \\ 1-\gamma \end{bmatrix},$$

and $M_2^B :=$

$$\begin{bmatrix} e^{-r_1 m_B} & e^{-r_2 m_B} & 0 & 0 & 0 & 0 \\ e^{r_1(f-m_B)} & e^{r_2(f-m_B)} & -e^{r_1(f-m_B)} & -e^{r_2(f-m_B)} & 0 & 0 \\ r_1 e^{r_1(f-m_B)} & r_2 e^{r_2(f-m_B)} & -r_1 e^{r_1(f-m_B)} & -r_2 e^{r_2(f-m_B)} & 0 & 0 \\ 0 & 0 & e^{r_1(f+\overline{m}^*-m_B)} & e^{r_2(f+\overline{m}^*-m_B)} & -e^{r_1(f+\overline{m}^*-m_B)} & -e^{r_2(f+\overline{m}^*-m_B)} \\ 0 & 0 & r_1 e^{r_1(f+\overline{m}^*-m_B)} & r_2 e^{r_2(f+\overline{m}^*-m_B)} & -r_1 e^{r_1(f+\overline{m}^*-m_B)} & -r_2 e^{r_2(f+\overline{m}^*-m_B)} \\ 0 & 0 & 0 & 0 & r_1 & r_2 \end{bmatrix}.$$

Case iii) $m_B - f \leq 0$
In this case, the bad bank defaults as soon as the shock arrives. Although this is an unrealistic case, we will add the derivations of its value function for the sake of complexity. The ordinary differential equation of said case is:

$$(\rho+\lambda)V_B(m) = \mu V_B'(m) + 0.5\sigma^2 V_B''(m),$$

so that

$$V_B(m) = \begin{cases} \sum_{i=1}^2 A_i^B e^{r_i(m-m_B)}, & \text{for } m < m_B, \\ \sum_{i=1}^2 A_i^B + m - m_B, & \text{for } m \geq m_B. \end{cases}$$

The coefficients A_1^B and A_2^B are determined as follows:

$$\begin{bmatrix} A_1^B \\ A_2^B \end{bmatrix} = (M_3^B)^{-1} v_3^B, \quad M_3^B := \begin{bmatrix} e^{-r_1 m_B} & e^{-r_2 m_B} \\ r_1 & r_2 \end{bmatrix}, \quad v_3^B := \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

B.1.3 Value function good bank before shock

Case i) $m_G \leq \overline{m}^*$

$$V_G(m) = \begin{cases} \sum_{i=1}^2 A_i^G e^{r_i(m-m_G)} + W(m), & \text{for } m \in [0, m_G), \\ \sum_{i=1}^2 A_i^G + W(m_G) + m - m_G, & \text{for } m \geq m_G. \end{cases} \tag{B.3}$$

where A_1^G and A_2^G are pinned down by

- Value matching at $m = 0$: $\lim_{m\downarrow 0} V_G(m) = 0$;
- Smooth pasting at $m = m_G$: $\lim_{m\downarrow m_G} V_G'(m) = 1$.

which gives us

$$\begin{bmatrix} A_1^G \\ A_2^G \end{bmatrix} = (M_1^G)^{-1} v_1^G, \quad M_1^G := \begin{bmatrix} e^{-r_1 m_G} & e^{-r_2 m_G} \\ r_1 & r_2 \end{bmatrix}, \quad v_1^G := \begin{bmatrix} 0 \\ 1 - W'(m_G) \end{bmatrix}.$$

Case ii) $m_G > \overline{m}^*$
Suppose now that the good bank pays out later than what is optimal,

$$V_G(m) = \begin{cases} \sum_{i=1}^2 A_i^G e^{r_i(m-m_G)} + W(m), & \text{for } m \in [0, \overline{m}^*), \\ \sum_{i=1}^2 B_i^G e^{r_i(m-m_G)} + \beta_G + \gamma m, & \text{for } m \in [\overline{m}^*, m_G), \\ \sum_{i=1}^2 B_i^G + \beta_G + (\gamma - 1) m_G + m, & \text{for } m \geq m_G. \end{cases}$$

where $\gamma = \lambda/(\rho + \lambda)$ and

$$(\rho + \lambda)\beta_G = \mu\gamma + \lambda\left(\frac{\mu}{\rho} - \overline{m}^*\right) \implies \beta_G = \gamma\left(\frac{\mu}{\rho + \lambda} + \frac{\mu}{\rho} - \overline{m}^*\right).$$

The coefficients are pinned down by the following conditions

- Value matching at $m = 0$: $\lim_{m\downarrow 0} V_G(m) = 0$;
- Value matching at $m = \overline{m}^*$: $\lim_{m\downarrow \overline{m}^*} V_G(m) = \lim_{m\uparrow \overline{m}^*} V_G(m)$;
- Smooth pasting at $m = \overline{m}^*$: $\lim_{m\downarrow \overline{m}^*} V_G'(m) = \lim_{m\uparrow \overline{m}^*} V_G'(m)$;
- Smooth pasting at $m = m_G$: $\lim_{m\uparrow m_G} = 1$.

This can be summarized by the following system of equations:

$$\begin{bmatrix} \mathbf{A^G} \\ \mathbf{B^G} \end{bmatrix} = (M_2^G)^{-1} v_2^G,$$

where

$$M_2^G := \begin{bmatrix} e^{-r_1 m_G} & e^{-r_2 m_G} & 0 & 0 \\ e^{r_1(\overline{m}^* - m_G)} & e^{r_2(\overline{m}^* - m_G)} & -e^{r_1(\overline{m}^* - m_G)} & -e^{r_2(\overline{m}^* - m_G)} \\ r_1 e^{r_1(\overline{m}^* - m_G)} & r_2 e^{r_2(\overline{m}^* - m_G)} & -r_1 e^{r_1(\overline{m}^* - m_G)} & -r_2 e^{r_2(\overline{m}^* - m_G)} \\ 0 & 0 & r_1 & r_2 \end{bmatrix}, \quad v_2^G := \begin{bmatrix} 0 \\ \frac{\mu}{\rho+\lambda}(\gamma - 1) \\ \gamma - 1 \\ 1 - \gamma \end{bmatrix}.$$

B.1.4 Full information case

In the full information case, the optimal cash target m_G^* of the good bank simply equals the after-shock threshold given in Eq. (B.1). The optimal cash target m_B^* is determined by the high-contact condition around the payout threshold m_B:

$$\lim_{m\downarrow m_B^*} V_B''(m) = 0.$$

In general, there exists no closed-form solution for the optimal payout level m_B^*. In the case where the dividend threshold is smaller than shock size f, the bank is wiped out when the shock arrives. In this scenario, there exists a closed form solution for m_B^*:

$$m_B^* = \frac{2}{r_2 - r_1} \log\left(-\frac{r_1}{r_2}\right). \tag{B.4}$$

One can show that this value is smaller than m_G^*. This implies that shareholders hoard less cash to hedge against the Brownian liquidation risk in the presence of tail risk.

Furthermore, one can observe that the intrinsic value of the bank at optimal boundary m_G^* can be determined as follows:

$$
\begin{aligned}
&(\rho+\lambda)V_G(m_G^*) = \mu V_G'(m_G^*) + 0.5\sigma^2 V_G''(m_G^*) + \lambda W(m_G^*),\\
&(\rho+\lambda)V_G(m_G^*) = \mu + \lambda W(m_G^*) \implies V_G(m_G^*) = \frac{\mu}{\rho+\lambda} + \frac{\lambda}{\rho+\lambda} W(m_G^*),
\end{aligned}
$$

and analogously for the bad bank

$$
\begin{aligned}
&(\rho+\lambda)V_B(m_B^*) = \mu V_B'(m_B^*) + 0.5\sigma^2 V_B''(m_B^*) + \lambda W(m_B^* - f),\\
&(\rho+\lambda)V_B(m_B^*) = \mu + \lambda W(m_B^* - f) \implies V_B(m_B^*) = \frac{\mu}{\rho+\lambda} + \frac{\lambda}{\rho+\lambda} W(m_B^* - f).
\end{aligned}
$$

B.1.5 Optimal cash target in the presence of asymmetric information

We will first look at the optimal cash target of a good bank that is being mimicked by a bad bank. The objective function of the good bank becomes:

$$
\begin{aligned}
&k\left[\alpha V_G(m; m_D) + (1-\alpha)V_B(m; m_D)\right] + (1-k)V_G(m; m_D)\\
&= \left[1-(1-\alpha)k\right] V_G(m; m_D) + (1-\alpha)kV_B(m; m_D)
\end{aligned}
$$

The optimal barrier $m_{G,p}^*$ is the solution to the following high contact condition:

$$
\left[1-(1-\alpha)k\right] \left.\frac{\partial^2 V_G(m; m_{G,p}^*)}{\partial m^2}\right|_{m=m_{G,p}^*} + (1-\alpha)k \left.\frac{\partial^2 V_B(m; m_{G,p}^*)}{\partial m^2}\right|_{m=m_{G,p}^*} = 0.
$$

Similarly, the best pooling equilibrium strategy for the bad bank $m_{B,p}^*$ is the solution to:

$$
k\alpha \left.\frac{\partial^2 V_G(m; m_{B,p}^*)}{\partial m^2}\right|_{m=m_{B,p}^*} + (1-k\alpha) \left.\frac{\partial^2 V_B(m; m_{B,p}^*)}{\partial m^2}\right|_{m=m_{B,p}^*} = 0.
$$

We will now study the optimal cash target of a good bank that is considered to be bad. The objective function of the good bank becomes:

$$
kV_B(m; m_D) + (1-k)V_G(m; m_D).
$$

The optimal barrier $m_{G,B}^*$ is the solution to the following high contact condition:

$$
(1-k) \left.\frac{\partial^2 V_G(m; m_{G,B}^*)}{\partial m^2}\right|_{m=m_{G,B}^*} + k \left.\frac{\partial^2 V_B(m; m_{G,B}^*)}{\partial m^2}\right|_{m=m_{G,B}^*} = 0.
$$

B.2 Default probability

In this section we present the methodology used to find the bank's default probability. We first approximate the default probabilities after the liquidity shock has materialized and extend the numerical procedure to the setting before arrival of the shock.

B.2.1 Default probability after shock

Let $\overline{K}(m,t,T)$ be the probability that the bank did not default before time horizon T for a current time $t<T$ and a cash reserve m. It is assumed that after the shock, the bank plays its optimal dividend strategy $\overline{m}^*$ and denote the default time

$$\overline{K}(m,t,T)=\mathbb{P}(\tau^L>T\mid M_t^L=m).$$

For the ease of notation, we drop the argument T. Following Klimenko and Moreno–Bromberg (2016), one can show that the survival probability $\overline{K}(m,t)$ solves the following boundary problem described in Eq. (B.5) - (B.8). Note that the presence of a time dimension results in an additional partial derivative with respect to time. Furthermore, the survival probability is not discounted, so the term $\rho\overline{K}(m,t)$ is not showing up in the partial derivative.

$$\frac{\partial\overline{K}(m,t)}{\partial t}+\frac{1}{2}\sigma^2\frac{\partial^2\overline{K}(m,t)}{\partial m^2}+\mu\frac{\partial\overline{K}(m,t)}{\partial m}=0, \tag{B.5}$$

$$\overline{K}(0,t)=0,\quad \forall t\geq 0, \tag{B.6}$$

$$\overline{K}(m,T)=1,\quad \forall m>0, \tag{B.7}$$

$$\left.\frac{\partial\overline{K}(m,t)}{\partial m}\right|_{m=\overline{m}^*}=0. \tag{B.8}$$

Boundary condition (B.6) states that the probability of survival after having been liquidated is zero. Furthermore, boundary condition (B.7) says that the bank has survived until $t=T$ for all positive values of m. Lastly, Neumann condition (B.8) guarantees that the survival probability does not change anymore beyond $m>\overline{m}^*$, since the cash reserves do not increase beyond this level. Because of the time dimension in the above boundary system, there is no closed-form solution to this system. Instead, we solve it numerically using the Crank-Nicolson finite-difference method, see Crank and Nicolson (1947).

B.2.2 Computations of default probability before shock

Let $K_\ell(m,t,T)$ be the probability that the bank of type ℓ did not dfault before time horizon T for a current time $t<T$ and cash reserve m, before arrival of the liquidity shock. For the good bank, $K_G(m,t,T)$ is solved exactly as in $\overline{K}(m,t,T)$ but now with payout threshold m_G rather than $\overline{m}^*$. For the bad bank that is subject to the liquidity shock, the boundary system before arrival of the liquidity shock is given by:

$$\frac{\partial K_B(m,t)}{\partial t}+\frac{1}{2}\sigma^2\frac{\partial^2 K_B(m,t)}{\partial m^2}+\mu\frac{\partial K_B(m,t)}{\partial m}=\lambda\left[K_B(m,t)-\overline{K}(m-f,t)\right], \tag{B.9}$$

$$K_B(0,t)=0,\quad \forall t\geq 0,$$

$$K_B(m,T)=1,\quad \forall m>0,$$

$$\left.\frac{\partial K_B(m,t)}{\partial m}\right|_{m=m_B}=0.$$

Eq. (B.9) incorporates the jump of the liquidity reserves. We distinguish several cases.
Case i) $m_B-f\in(0,\overline{m}^*)$
In this scenario, the partial differential equation is given by:

$$\frac{\partial K_B(m,t)}{\partial t}+\frac{1}{2}\sigma^2\frac{\partial^2 K_B(m,t)}{\partial m^2}+\mu\frac{\partial K_B(m,t)}{\partial m}-\lambda K_B(m,t)=0,\qquad m<f,$$

$$\frac{\partial K_B(m,t)}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 K_B(m,t)}{\partial m^2} + \mu \frac{\partial K_B(m,t)}{\partial m} - \lambda \left[K_B(m,t) - \overline{K}(m-f,t) \right], \quad m \in [f, m_B].$$

Before applying the Crank Nicolson method, one discretizes the grid separately on the domains $(0,T) \times (0,f)$ and $(0,T) \times (f, m_B)$.

Case ii) $m_B - f \geq \overline{m}^*$
In this case, we differentiate three regions:

$$\begin{aligned}
&\frac{\partial K_B(m,t)}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 K_B(m,t)}{\partial m^2} + \mu \frac{\partial K_B(m,t)}{\partial m} - \lambda K_B(m,t) = 0, && m < f,\\
&\frac{\partial K_B(m,t)}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 K_B(m,t)}{\partial m^2} + \mu \frac{\partial K_B(m,t)}{\partial m} - \lambda \left[K_B(m,t) - \overline{K}(m-f,t) \right], && m \in [f, \overline{m}^*],\\
&\frac{\partial K_B(m,t)}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 K_B(m,t)}{\partial m^2} + \mu \frac{\partial K_B(m,t)}{\partial m} - \lambda \left[K_B(m,t) - \overline{K}(\overline{m}^* - f,t) \right], && m > \overline{m}^*,
\end{aligned}$$

Case iii) $m_B - f \leq 0$
The boundary system looks as follows:

$$\frac{\partial K_B(m,t)}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 K_B(m,t)}{\partial m^2} + \mu \frac{\partial K_B(m,t)}{\partial m} - \lambda K_B(m,t) = 0.$$

To conclude, the default probability PD_ℓ^T is given by $1 - K_\ell(m, 0, T)$.

B.3 Single-crossing

The bank's objective is to maximize the following function:

$$V_{\ell,\tilde{\ell}}(m) = kV_{\tilde{\ell}}(m) + (1-k)V_\ell(m),$$

where ℓ represents the bank's true type and $\tilde{\ell}$ the bank's perceived type. The single crossing condition states that lowering the dividend payout threshold can be considered a valid signal from the good bank when lowering dividend boundary m_D is less costly for the good bank than for the bad bank. This can be formalized as follows:

$$\frac{V_{G,G}(m) - V_{G,B}(m)}{\partial V_{G,\tilde{\ell}} / \partial m_D} > \frac{V_{B,G}(m) - V_{B,B}(m)}{\partial V_{B,\tilde{\ell}} / \partial m_D}. \tag{B.10}$$

Note that,

$$\begin{aligned}
V_{G,G}(m) - V_{G,B}(m) &= k(V_G(m) - V_B(m)),\\
V_{B,G}(m) - V_{B,B}(m) &= k(V_G(m) - V_B(m)).
\end{aligned}$$

Furthermore

$$\begin{aligned}
\frac{\partial V_{G,\tilde{\ell}}(m)}{\partial m_D} &= k\frac{\partial V_{\tilde{\ell}}(m)}{\partial m_D} + (1-k)\frac{\partial V_G(m)}{\partial m_D},\\
\frac{\partial V_{B,\tilde{\ell}}(m)}{\partial m_D} &= k\frac{\partial V_{\tilde{\ell}}(m)}{\partial m_D} + (1-k)\frac{\partial V_B(m)}{\partial m_D}.
\end{aligned}$$

Therefore, condition (B.10) can be simplified to

$$\frac{\partial V_G(m)}{\partial m_D} < \frac{\partial V_B(m)}{\partial m_D}.$$

We will show in the remainder of this section when this condition holds.

B.3.1 Derivative value function bad bank w.r.t. payout threshold

In this section we will compute the derivative of V_B with respect to m_D. First define the following quantities:

$$\begin{aligned}
\Delta &:= \Delta(m_D) = r_2 e^{-m_D r_1} - r_1 e^{-m_D r_2} > 0, \quad \forall m_D \geq 0, \\
\Delta' &:= \Delta'(m_D) = r_1 r_2 \left(e^{-m_D r_2} - e^{-m_D r_1}\right) \geq (>) 0, \quad \forall m_D \geq (>) 0.
\end{aligned} \tag{B.11}$$

Case i) $m_D - f \in (0, \overline{m}^*)$, $m \in (f, m_D)$:
In this case, the value of the bad bank equals (see Eq. (B.2)):

$$V_B(m; m_D) = \sum_{i=1}^{2} B_i^B e^{r_i(m-m_D)} + W(m-f).$$

One can show after many algebraic manipulations that

$$\begin{bmatrix} B_1^B \\ B_2^B \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} r_2 & -e^{-m_D r_2} \\ -r_1 & e^{-m_D r_1} \end{bmatrix} \begin{bmatrix} e^{-f r_1} - e^{-f r_2} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} (r_2 - r_1)^{-1} W'(0) \\ 1 - W'(m_D - f) \end{bmatrix}.$$

So that

$$\begin{aligned}
V_B &= \frac{1}{\Delta} \begin{bmatrix} e^{r_1(m-m_D)} \\ e^{r_2(m-m_D)} \end{bmatrix}^\top \begin{bmatrix} r_2 & -e^{-m_D r_2} \\ -r_1 & e^{-m_D r_1} \end{bmatrix} \begin{bmatrix} e^{-r_1 f} - e^{-r_2 f} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} (r_2 - r_1)^{-1} W'(0) \\ 1 - W'(m_D - f) \end{bmatrix} + W(m-f) \\
&= \frac{n_B}{\Delta} + W(m-f),
\end{aligned}$$

where n_B is defined by

$$\begin{aligned}
n_B := n_B(m_D) = \frac{W'(0)}{r_2 - r_1} &\left(e^{-r_1 f} - e^{-r_2 f}\right)\left(r_2 e^{r_1(m-m_D)} - r_1 e^{r_2(m-m_D)}\right) \\
&+ \left(1 - W'(m_D - f)\right) e^{-m_D(r_1+r_2)} \left(e^{r_2 m} - e^{r_1 m}\right).
\end{aligned}$$

and whose derivative with respect to m_D is given by:

$$\begin{aligned}
n_B' = \frac{\partial n_B(m_D)}{\partial m_D} &= W'(0) \frac{r_1 r_2}{r_2 - r_1} (e^{-r_1 f} - e^{-r_2 f})(e^{r_2(m-m_D)} - e^{r_1(m-m_D)}) \\
&+ e^{-m_D(r_1+r_2)} (e^{r_2 m} - e^{r_1 m}) \left(-(r_1 + r_2)(1 - W'(m_D - f) - W''(m_D - f)\right).
\end{aligned}$$

The derivative of V_B is then given by

$$\begin{aligned}
&\frac{\partial V_B(m)}{\partial m_D} = \frac{1}{\Delta^2}(n_B' \Delta - \Delta' n_B) \\
&= \frac{1}{\Delta^2}\left(\frac{W'(0)}{r_2 - r_1}\left(e^{-r_1 f} - e^{-r_2 f}\right)\left[r_1(r_2\Delta + \Delta')e^{r_2(m-m_D)} - r_2(r_1\Delta + \Delta')e^{-r_1(m-m_D)}\right]\right. \\
&\left. - (e^{r_2 m} - e^{r_1 m}) e^{-m_D(r_1+r_2)} \left[((r_1 + r_2)\Delta + \Delta')(1 - W'(m_D - f)) + W''(m_D - f)\Delta\right]\right).
\end{aligned}$$

where one can do the following simplifications

$$\left[r_1(r_2\Delta+\Delta')e^{r_2(m-m_D)} - r_2(r_1\Delta+\Delta')e^{-r_1(m-m_D)}\right] = r_1 r_2 e^{-m_D(r_1+r_2)}(r_2-r_1)(e^{r_2 m}-e^{r_1 m}),$$
$$(r_1+r_2)\Delta+\Delta' = r_2^2 e^{-m_D r_1} - r_1^2 e^{-m_D r_2}.$$

Define the following quantities

$$\Gamma := \Gamma(m_D) = (r_1+r_2)\Delta+\Delta' = r_2^2 e^{-r_1 m_D} - r_1^2 e^{-r_2 m_D}, \tag{B.12}$$
$$\Gamma' = -r_1 r_2 \Delta,$$

so that

$$\frac{\partial V_B}{\partial m_D} = \frac{e^{-m_D(r_1+r_2)}(e^{r_2 m}-e^{r_1 m})}{\Delta^2}$$
$$\Big(W'(0) r_1 r_2 (e^{-r_1 f}-e^{-r_2 f}) - (1-W'(m_D-f))\Gamma - W''(m_D-f)\Delta\Big). \tag{B.13}$$

Case ii) $m \le f,\ m_D - f < \overline{m}^*$:
In this scenario, one can show that the value function equals

$$V_B(m) = \frac{1}{\Delta}(e^{r_2 m}-e^{r_1 m})e^{-m_D(r_1+r_2)} \begin{bmatrix} r_2 e^{-r_2(f-m_D)} - r_1 e^{-r_1(f-m_D)} \\ 1 \end{bmatrix}^\top \begin{bmatrix} W'(0)(r_2-r_1)^{-1} \\ 1-W'(m_D-f) \end{bmatrix}$$
$$= \frac{\tilde{n}_B}{\Delta}(e^{r_2 m}-e^{r_1 m}),$$

where

$$\tilde{n}_B := \tilde{n}_B(m_D) = \left(\frac{W'(0)}{r_2-r_1}[r_2 e^{-r_2(f-m_D)} - r_1 e^{-r_1(f-m_D)}] + 1 - W'(m_D-f)\right) e^{-m_D(r_1+r_2)}$$

The derivative of interest is

$$\frac{\partial V_B(m)}{\partial m_D} = \frac{1}{\Delta^2}(\tilde{n}_B'\Delta - \Delta'\tilde{n}_B)(e^{r_2 m}-e^{r_1 m}),$$

where

$$\tilde{n}_B'\Delta - \Delta'\tilde{n}_B$$
$$= e^{-m_D(r_1+r_2)}\Big((W'(0) r_1 r_2 (e^{-r_1 f}-e^{-r_2 f}) - (1-W'(m_D-f))\Gamma - W''(m_D-f)\Delta\Big).$$

This corresponds to Eq. (B.13).

B.3.2 Derivative value function good bank w.r.t. payout threshold

For the high type the value function from Eq. (B.3) for $m \in [0, m_D)]$ can be rewritten as:

$$V_G(m;m_D) = \frac{1}{\Delta}\begin{bmatrix} e^{r_1(m-m_D)} \\ e^{r_2(m-m_D)} \end{bmatrix}^\top \begin{bmatrix} r_2 & -e^{-r_2 m_D} \\ -r_1 & e^{-r_1 m_D} \end{bmatrix} \begin{bmatrix} 0 \\ 1-W'(m_D) \end{bmatrix} + W(m)$$
$$= \frac{1}{\Delta}e^{-m_D(r_1+r_2)}(e^{r_2 m}-e^{r_1 m})(1-W'(m_D)) + W(m),$$

so that its derivative is given by

$$\frac{\partial V_G}{\partial m_D} = \frac{e^{-m_D(r_1+r_2)}(e^{r_2 m} - e^{r_1 m})}{\Delta^2}\left(-(1-W'(m_D))\Gamma - W''(m_D)\Delta\right).$$

Conditions for single-crossing to hold

For distortion to be more costly for the low type, it must be the case that

$$\frac{\partial V_G}{\partial m_D} < \frac{\partial V_B}{\partial m_D} \iff$$
$$(W'(m_D) - W'(m_D - f))\Gamma < \Delta(W''(m_D) - W''(m_D - f)) + W'(0) r_1 r_2 (e^{-r_1 f} - e^{-r_2 f}). \tag{B.14}$$

Note that for $m_D = \overline{m}_G^*$, $\partial V_G(m)/\partial m_D = 0$, since $W'(\overline{m}_G^*) = 1$ and $W''(\overline{m}_G^*) = 0$. As a result, the condition simplifies to $\partial V_B/\partial m_D > 0$.

B.3.3 Uniqueness

Rewriting Eq.(B.14) as

$$(W'(m_D) - W'(m_D - f))\Gamma - \Delta(W''(m_D) - W''(m_D - f)) < W'(0) r_1 r_2 (e^{-r_1 f} - e^{-r_2 f})$$

the right hand side is constant with respect to m_D. If the left hand side is monotonic in m_D on the relevant range, there is at most one level of m_D for which the inequality holds with equality, defining a region where it holds and one where it does not. First, define the following function

$$h_1(m_D) := e^{\bar{r}_1(m_D - \bar{m}^*)}(1 - e^{-\bar{r}_1 f}),$$
$$h_2(m_D) := e^{\bar{r}_2(m_D - \bar{m}^*)}(1 - e^{-\bar{r}_2 f}).$$

Taking out the common constant in the after value functions $\frac{1}{\bar{r}_2 - \bar{r}_1} = \frac{\sigma^2}{2\sqrt{\bar{\Omega}}}$, the brackets on the left hand side can be rewritten as respectively:

$$\bar{r}_2 e^{\bar{r}_1(m_D - \bar{m}^*)} - \bar{r}_1 e^{\bar{r}_2(m_D - \bar{m}^*)} - \bar{r}_2 e^{\bar{r}_1(m_D - f - \bar{m}^*)} + \bar{r}_1 e^{\bar{r}_2(m_D - f - \bar{m}^*)} = \bar{r}_2 h_1(m_D) - \bar{r}_1 h_2(m_D),$$
$$\bar{r}_2 \bar{r}_1 \left(e^{\bar{r}_1(m_D - \bar{m}^*)} - e^{\bar{r}_2(m_D - \bar{m}^*)} - e^{\bar{r}_1(m_D - f - \bar{m}^*)} + e^{\bar{r}_2(m_D - f - \bar{m}^*)}\right) = \bar{r}_1 \bar{r}_2 [h_1(m_D) - h_2(m_D)],$$

and the left hand side can be summarized as

$$LHS = \frac{\sigma^2}{2\sqrt{\bar{\Omega}}}\{[\bar{r}_2 h_1(m_D) - \bar{r}_1 h_2(m_D)]\Gamma - \Delta \bar{r}_1 \bar{r}_2 [h_1(m_D) - h_2(m_D)]\}.$$

Taking a derivative with respect to m_D yields

$$\begin{aligned}\frac{\partial LHS}{\partial m_D} &= \frac{\sigma^2}{2\sqrt{\bar{\Omega}}}\Big\{[\bar{r}_1 h_2(m_D) - \bar{r}_2 h_1(m_D)] r_1 r_2 \Delta + \bar{r}_1 \bar{r}_2 [h_1(m_D) - h_2(m_D)]\Gamma \\ &\quad - \Delta' \bar{r}_1 \bar{r}_2 [h_1(m_D) - h_2(m_D)] - \Delta \bar{r}_1 \bar{r}_2 [\bar{r}_1 h_1(m_D) - \bar{r}_2 h_2(m_D)]\Big\} \\ &= \bar{r}_1 \bar{r}_2 \frac{\sigma^2}{2\sqrt{\bar{\Omega}}}\Big\{-\frac{r_1 r_2}{\bar{r}_1 \bar{r}_2}[\bar{r}_2 h_1(m_D) - \bar{r}_1 h_2(m_D)]\Delta \\ &\quad + (\Gamma - \Delta')[h_1(m_D) - h_2(m_D)] - \Delta[\bar{r}_1 h_1(m_D) - \bar{r}_2 h_2(m_D)]\Big\}\end{aligned}$$

$$
\begin{aligned}
&= \frac{\rho}{\sqrt{\bar{\Omega}}}\Delta\frac{\rho+\lambda}{\rho}\Bigg\{[\bar{r}_2 h_1(m_D)-\bar{r}_1 h_2(m_D)]+\frac{\rho}{\rho+\lambda}[\bar{r}_1 h_1(m_D)-\bar{r}_2 h_2(m_D)]\\
&\quad +2\frac{\mu}{\sigma^2}\frac{\rho}{\rho+\lambda}[h_1(m_D)-h_2(m_D)]\Bigg\}.
\end{aligned}
$$

using that $\Gamma-\Delta'=(r_1+r_2)\Delta=-\frac{2\mu}{\sigma^2}\Delta$, $\frac{r_1 r_2}{\bar{r}_1\bar{r}_2}=\frac{\rho+\lambda}{\rho}$ and $\bar{r}_1\bar{r}_2=-2\frac{\rho}{\sigma^2}$. Collecting $h_i(m_D)$, $\forall i\in\{1,2\}$ terms in the first line of the bracket yields

$$
\begin{aligned}
&\left(\bar{r}_2+\frac{\rho}{\rho+\lambda}\bar{r}_1\right)h_1(m_D)-\left(\bar{r}_1+\frac{\rho}{\rho+\lambda}\bar{r}_2\right)h_2(m_D)\\
&=\frac{1}{\sigma^2}\left\{-\mu\left(1+\frac{\rho}{\rho+\lambda}\right)[h_1(m_D)-h_2(m_D)]+\sqrt{\bar{\Omega}}\frac{\lambda}{\rho+\lambda}[h_1(m_D)+h_2(m_D)]\right\},
\end{aligned}
$$

which allows the following simplification of the derivative

$$
\begin{aligned}
\frac{\partial LHS}{\partial m_D} &= \frac{\rho}{\sqrt{\bar{\Omega}}}\Delta\frac{\rho+\lambda}{\rho\sigma^2}\Bigg\{\sqrt{\bar{\Omega}}\frac{\lambda}{\rho+\lambda}[h_1(m_D)+h_2(m_D)]\\
&\qquad -\mu\left(1+\frac{\rho}{\rho+\lambda}-2\frac{\rho}{\rho+\lambda}\right)[h_1(m_D)-h_2(m_D)]\Bigg\}\\
&=\frac{\lambda}{\sqrt{\bar{\Omega}}\sigma^2}\Delta\left\{\sqrt{\bar{\Omega}}[h_1(m_D)+h_2(m_D)]-\mu[h_1(m_D)-h_2(m_D)]\right\}\\
&=\frac{\lambda}{\sqrt{\bar{\Omega}}\sigma^2}\Delta\left\{\left(\sqrt{\bar{\Omega}}-\mu\right)h_1(m_D)+\left(\sqrt{\bar{\Omega}}+\mu\right)h_2(m_D)\right\}.
\end{aligned}
$$

Since both the constant $\rho\bar{\Omega}^{-1/2}$ and Δ are positive for all values of m_D, the sign of the derivative is determined by the bracket where both $\left(\sqrt{\bar{\Omega}}+\mu\right)\geq\left(\sqrt{\bar{\Omega}}-\mu\right)\geq 0$. The signs of the h-functions are mostly contrasting with $h_1(m_D)=e^{\bar{r}_1(m_D-\bar{m}^*)}(1-e^{-\bar{r}_1 f})\leq 0$ and $h_2(m_D)=e^{\bar{r}_2(m_D-\bar{m}^*)}(1-e^{-\bar{r}_2 f})\geq 0$ because $e^{-\bar{r}_1 f}\geq 1\geq e^{-\bar{r}_2 f}$, with the exception being at $f=0$ where the inequalities hold with equality and both functions are zero. For the range of $m_D\in(f,\bar{m}^*)$ a second set of relevant inequalities is $e^{\bar{r}_1(m_D-\bar{m}^*)}\geq 1\geq e^{\bar{r}_2(m_D-\bar{m}^*)}$ where equality holds at $m_D=\bar{m}^*$. The derivative of the bracket is increasing in m_D as both $h_1'(m_D)=r_1 h_1(m_D)\geq 0$ and $h_2'(m_D)=r_2 h_2(m_D)\geq 0$, so if $\frac{\partial LHS}{\partial m_D}\Big|_{m_D=\bar{m}^*}<0$ the derivative is negative over the full range. For $f>0$ this is indeed the case as

$$
\frac{\partial LHS}{\partial m_D}\bigg|_{m_D=\bar{m}^*}\left(\frac{\lambda}{\sqrt{\bar{\Omega}}\sigma^2}\Delta\right)^{-1}=\left(\sqrt{\bar{\Omega}}-\mu\right)(1-e^{-\bar{r}_1 f})+\left(\sqrt{\bar{\Omega}}+\mu\right)(1-e^{-\bar{r}_2 f}),
$$

which is zero at $f=0$ and decreasing in f which will be shown next. To establish the second point notice that

$$
\begin{aligned}
\frac{\partial^2 LHS}{\partial m_D\partial f}\bigg|_{m_D=\bar{m}^*}\left(\frac{\lambda}{\sqrt{\bar{\Omega}}\sigma^2}\Delta\right)^{-1}&=\left(\sqrt{\bar{\Omega}}-\mu\right)\bar{r}_1 e^{-\bar{r}_1 f}+\left(\sqrt{\bar{\Omega}}+\mu\right)\bar{r}_2 e^{-\bar{r}_2 f}\\
&=e^{-\bar{r}_2 f}\left[\left(\sqrt{\bar{\Omega}}-\mu\right)\bar{r}_1 e^{2\frac{\sqrt{\bar{\Omega}}}{\sigma^2}f}+\bar{r}_2\left(\sqrt{\bar{\Omega}}+\mu\right)\right].
\end{aligned}
$$

If the bracket in the last expression is negative, higher f means lower $\frac{\partial LHS}{\partial m_D}\Big|_{m_D=\bar{m}^*}$ since $e^{-\bar{r}_2 f}>0$. Because $\frac{\partial LHS}{\partial m_D}\Big|_{m_D=\bar{m}^*}$ is zero at $f=0$ it must be negative for $f>0$. Focusing on the square bracket, its derivative with respect to f is trivially negative as

$$\underbrace{\bar{r}_1}_{<0}\underbrace{\left(\sqrt{\bar{\Omega}}-\mu\right)2\frac{\sqrt{\bar{\Omega}}}{\sigma^2}e^{2\frac{\sqrt{\bar{\Omega}}}{\sigma^2}f}}_{>0}<0.$$

At $f=0$ the value of the square bracket is zero since

$$\begin{aligned}
&\left(\sqrt{\bar{\Omega}}-\mu\right)\bar{r}_1+\bar{r}_2\left(\sqrt{\bar{\Omega}}+\mu\right)=0 \iff \bar{r}_2\left(\sqrt{\bar{\Omega}}+\mu\right)<-\left(\sqrt{\bar{\Omega}}-\mu\right)\bar{r}_1\\
\iff &\frac{1}{\sigma^2}\left(\sqrt{\bar{\Omega}}+\mu\right)\left(-\mu+\sqrt{\bar{\Omega}}\right)=\frac{1}{\sigma^2}\left(\sqrt{\bar{\Omega}}-\mu\right)\left(\mu+\sqrt{\bar{\Omega}}\right)\\
\iff &\left(\mu+\sqrt{\bar{\Omega}}\right)\left(\sqrt{\bar{\Omega}}-\mu\right)=\left(\sqrt{\bar{\Omega}}-\mu\right)\left(\mu+\sqrt{\bar{\Omega}}\right),
\end{aligned}$$

so for $f>0$ the bracket is negative, which concludes the proof.

B.3.4 Existence

Define the following function

$$g(m_D)=\Gamma[W'(m_D)-W'(m_D-f)]-\Delta[W''(m_D)-W''(m_D-f)]+W'(0)\Delta'(f)$$

Evaluated at $m_D=f$

$$g(f)=\Gamma(f)[W'(f)-W'(0)]-\Delta(f)[W''(f)-W''(0)]+W'(0)\Delta'(f)=A(f)-B(f),$$

where

$$\begin{aligned}
A(f)&:=\Gamma(f)W'(f)-\Delta(f)W''(f),\\
B(f)&:=\Gamma(f)W'(0)-\Delta(f)W''(0)-W'(0)\Delta'(f).
\end{aligned}$$

Using that $\Gamma(f)-\Delta'(f)=(r_1+r_2)\Delta(f)$ and $r_1+r_2=\overline{r}_1+\overline{r}_2$, we show that $B(f)=0$:

$$\begin{aligned}
B(f)&=\Delta(f)\left[(r_1+r_2)W'(0)-W''(0)\right]\\
&=\frac{\Delta(f)}{\overline{r}_2-\overline{r}_1}\left[e^{-\overline{r}_1\overline{m}^*}\overline{r}_2(r_1+r_2-\overline{r}_1)+e^{-\overline{r}_2\overline{m}^*}\overline{r}_1(\overline{r}_2-r_1-r_2)\right]\\
&=\frac{\Delta(f)}{\overline{r}_2-\overline{r}_1}\left[\overline{r}_2^2e^{-\overline{r}_1\overline{m}^*}-\overline{r}_1^2e^{-\overline{r}_2\overline{m}^*}\right]=0.
\end{aligned}$$

We will show that $A(0)=0$:

$$\begin{aligned}
A(0)&=\Gamma(0)W'(0)-\Delta(0)W''(0)=(r_2^2-r_1^2)W'(0)-(r_2-r_1)W''(0)\\
&=(r_2-r_1)\left[(r_1+r_2)W'(0)-W''(0)\right]=0.
\end{aligned}$$

The derivative of $A(f)$ with respect to f is positive:

$$\begin{aligned}
\frac{\partial A}{\partial f}&=\Gamma'(f)W'(f)+\Gamma(f)W''(f)-\Delta'(f)W''(f)-\Delta(f)W'''(f)\\
&=-r_1r_2\Delta(f)W'(f)+(\Gamma(f)-\Delta'(f))W''(f)-\Delta(f)W'''(f)\\
&=\Delta(f)[W''(f)(r_1+r_2)-W'(f)r_1r_2-W'''(f)]\\
&=\frac{1}{\bar{r}_2-\bar{r}_1}\left[e^{\bar{r}_1(f-\bar{m}^*)}\bar{r}_2(r_1-\bar{r}_1)(\bar{r}_1-r_2)-e^{\bar{r}_2(f-\bar{m}^*)}\bar{r}_1(r_2-\bar{r}_2)(\bar{r}_2-r_1)\right]
\end{aligned}$$

$$= \frac{(r_1 - \bar{r}_1)(\bar{r}_1 - r_2)}{\bar{r}_2 - \bar{r}_1} \left[e^{\bar{r}_1(f - \bar{m}^*)}\bar{r}_2 - e^{\bar{r}_2(f-\bar{m}^*)}\bar{r}_1 \right] > 0.$$

Since $A(0) = 0$, and the derivative with respect to f is positive, we know that $A(f) > 0$ for $f > 0$.

Evaluated at $m_D = \overline{m}^*$

$$g(\overline{m}^*) = (1 - W'(\overline{m}^* - f))\Gamma(\overline{m}^*) + \Delta(\overline{m}^*)W''(\overline{m}^* - f) + W'(0)\Delta'(f)$$

Again $g(\overline{m}^*)|_{f=0} = 0$. The derivative is

$$\begin{aligned}
\frac{\partial g(\overline{m}^*)}{\partial f} &= \Gamma(\overline{m}^*)W''(\overline{m}^* - f)) - \Delta(\overline{m}^*)W'''(\overline{m}^* - f) + W'(0)\Delta''(f) \\
&= \frac{\bar{r}_1\bar{r}_2}{\bar{r}_2 - \bar{r}_1}\Big\{(e^{-\bar{r}_1 f} - e^{-\bar{r}_2 f})\Gamma(\overline{m}^*) - \Delta(\overline{m}^*)(\bar{r}_1 e^{-\bar{r}_1 f} - \bar{r}_2 e^{-\bar{r}_2 f}) \\
&\qquad - \frac{\rho+\lambda}{\rho}(\bar{r}_2 e^{-\bar{r}_1\overline{m}^*} - \bar{r}_1 e^{-\bar{r}_2\overline{m}^*})(r_2 e^{-r_2 f} - r_1 e^{-r_1 f})\Big\} \\
&= -\frac{\bar{r}_1\bar{r}_2}{\bar{r}_2 - \bar{r}_1}\Big\{(e^{-\bar{r}_2 f} - e^{-\bar{r}_1 f})\Gamma(\overline{m}^*) + \Delta(\overline{m}^*)(\bar{r}_1 e^{-\bar{r}_1 f} - \bar{r}_2 e^{-\bar{r}_2 f}) \\
&\qquad + \frac{\rho+\lambda}{\rho}\Delta(\overline{m}^*)(r_2 e^{-r_2 f} - r_1 e^{-r_1 f})\Big\}
\end{aligned}$$

Plugging in these quantities gives

$$\begin{aligned}
g(\overline{m}^*) &= \left(1 - \frac{1}{\overline{r}_2 - \overline{r}_1}\left[\overline{r}_2 e^{-\overline{r}_1 f} - \overline{r}_1 e^{-\overline{r}_2 f}\right]\right)\Gamma(\overline{m}^*) + \Delta(\overline{m}^*)\frac{\overline{r}_1\overline{r}_2}{\overline{r}_2 - \overline{r}_1}\left[e^{-\overline{r}_1 f} - e^{-\overline{r}_2 f}\right] \\
&\quad + \frac{1}{\overline{r}_2 - \overline{r}_1}\left[\overline{r}_2 e^{-\overline{r}_1\overline{m}^*} - \overline{r}_1 e^{-\overline{r}_2\overline{m}^*}\right]\Delta'(f) \\
(\bar{r}_2 - \bar{r}_1)g(\overline{m}^*) &= \underbrace{\left(\overline{r}_2 - \overline{r}_1 - \overline{\Delta}(f)\right)\Gamma(\overline{m}^*)}_{<0}\underbrace{-\Delta(\overline{m}^*)\overline{\Delta}'(f)}_{<0} + \underbrace{\overline{\Delta}(\overline{m}^*)\Delta'(f)}_{>0}.
\end{aligned}$$

Note that $\Gamma(\overline{m}^*) > 0$, since $\Gamma(m^*) = 0$, $\Gamma' > 0$ and $\overline{m}^* > m^*$.

$$\Delta(x) - \overline{\Delta}(x) = r_2 e^{-r_1 x} - \overline{r}_2 e^{-\overline{r}_1 x} - r_1 e^{-r_2 x} + \overline{r}_1 e^{-\overline{r}_2 x}$$

This expression is 0 for $x = 0$. Take the derivative:

$$\begin{aligned}
\Delta'(x) - \overline{\Delta}'(x) &= -r_1 r_2 e^{-r_1 x} + \overline{r}_1\overline{r}_2 e^{-\overline{r}_1 x} + r_1 r_2 e^{-r_2 x} - \overline{r}_1\overline{r}_2 e^{-\overline{r}_2 x} \\
&= \frac{2(\rho+\lambda)}{\sigma^2}\left[e^{-r_2 x} - e^{-r_1 x}\right] - \frac{2\rho}{\sigma^2}\left[e^{-\overline{r}_2 x} - e^{-\overline{r}_1 x}\right] < 0.
\end{aligned}$$

This expression is negative because $e^{-r_2 x} < e^{-\overline{r}_2 x}$, $e^{-r_1 x} > e^{-\overline{r}_1 x}$, so that $e^{-r_2 x} - e^{-r_1 x} < e^{-\overline{r}_2 x} - e^{-\overline{r}_1 x} < 0$.

B.4 Reverse signaling

In this section we will consider the case where the good bank signals its quality by increasing its dividend policy. In this scenario, the single-crossing condition becomes:

$$\frac{\partial V_G(m; m_D)}{\partial m_D} > \frac{\partial V_B(m; m_D)}{\partial m_D}.$$

This condition implies that the cost of increasing m_D is lower for the good bank than for the bad bank.

B.4.1 Derivative value function good bank with respect to payout level

After some algebraic steps, one can show that the value function of the good bank for $m \in [m_G^*, m_D)$ is given by:

$$V_G(m; m_D) = \frac{1-\gamma}{(r_2 - r_1)\Delta} \begin{bmatrix} e^{r_1(m-m_G)} \\ e^{r_2(m-m_G)} \end{bmatrix}^\top \begin{bmatrix} r_2 & r_2 & e^{-m_D r_2} \\ -r_1 & -r_1 & -e^{-m_D r_1} \end{bmatrix}$$
$$\times \operatorname{diag}\left(\begin{bmatrix} r_2 e^{-m_G^* r_1} - r_1 e^{-m_G^* r_2} \\ e^{-m_G^* r_2} - e^{-m_G^* r_1} \\ r_1 - r_2 \end{bmatrix} \right) \begin{bmatrix} \frac{\mu}{\rho+\lambda} \\ 1 \\ 1 \end{bmatrix} + \beta_G + \gamma m.$$

This can be written as

$$V_G(m; m_D) = \frac{n_G}{(r_2 - r_1)\Delta} + \beta_G + \gamma m,$$

where

$$\begin{aligned} n_G &\coloneqq n_G(m_D) \\ &= C_G \left[r_2 e^{r_1(m-m_D)} - r_1 e^{r_2(m-m_D)} \right] + (1-\gamma)(e^{r_1 m} - e^{r_2 m}) e^{-(r_1+r_2)m_D} (r_1 - r_2), \\ n_G' &\coloneqq n_G'(m_D) \\ &= r_1 r_2 (e^{r_2(m-m_D)} - e^{r_1(m-m_D)}) C_G + (1-\gamma)(e^{r_1 m} - e^{r_2 m})(r_1 + r_2)(r_1 - r_2) e^{-(r_1+r_2)m_D}, \end{aligned}$$

and

$$\begin{aligned} C_G &\coloneqq (1-\gamma) \left[\left(\frac{\mu r_2}{\rho+\lambda} - 1 \right) e^{-\overline{m}^* r_1} - \left(\frac{\mu r_1}{\rho+\lambda} - 1 \right) e^{-\overline{m}^* r_2} \right] \\ &= \frac{1}{2}\sigma^2 \frac{1-\gamma}{\rho+\lambda} \left[r_2^2 e^{-\overline{m}^* r_1} - r_1^2 e^{-\overline{m}^* r_2} \right] = \frac{1}{2}\sigma^2 \frac{\rho}{(\rho+\lambda)^2} \Gamma(\overline{m}^*). \end{aligned}$$

Now consider

$$\frac{\partial V_G(m; m_D)}{\partial m_D} = \frac{1}{(r_2 - r_1)\Delta^2} (n_G' \Delta - \Delta' n_G),$$

where

$$n_G' \Delta - \Delta' n_G = (r_2 - r_1) e^{-m_G(r_1+r_2)} (e^{r_2 m} - e^{r_1 m}) \left[r_1 r_2 C_G - (1-\gamma)\Gamma \right].$$

Putting everything together gives the following expression for the derivative

$$\frac{\partial V_G(m; m_D)}{\partial m_G} = \frac{e^{-m_D(r_1+r_2)} (e^{r_2 m} - e^{r_1 m})}{\Delta^2} \left[r_1 r_2 C_G - (1-\gamma)\Gamma \right].$$

Observe that

$$r_1 r_2 C_G = -\frac{2\rho}{\sigma^2} \frac{1}{2}\sigma^2 \frac{\rho}{(\rho+\lambda)^2} \Gamma(\overline{m}^*) = -\left(\frac{\rho}{\rho+\lambda} \right)^2 \Gamma(\overline{m}^*).$$

B.4.2 Derivative value function bad bank with respect to payout level

The derivative of $V_B(m;m_D)$ with respect to m_D in the case that $m_D - f < \overline{m}^*$, is given by Eq. (B.13). In the case that $m_D - f > \overline{m}^*$, the value function for $m \in [f, m_D)$ is given by

$$V_B(m;m_D) = \frac{1}{\Delta(r_2-r_1)}\begin{bmatrix} e^{r_1(m-m_D)} \\ e^{r_2(m-m_D)} \end{bmatrix}^\top \begin{bmatrix} -r_2 & r_2 & -r_2 & e^{-m_D r_2} \\ r_1 & -r_1 & r_1 & -e^{-m_D r_1} \end{bmatrix}$$
$$\times \operatorname{diag}\left(\begin{bmatrix} e^{-fr_2} - e^{-fr_1} \\ r_1 e^{-r_2(\overline{m}^*+f)} - r_2 e^{-r_1(\overline{m}^*+f)} \\ e^{-r_2(\overline{m}^*+f)} - e^{-r_1(\overline{m}^*+f)} \\ r_1 - r_2 \end{bmatrix}\right)\begin{bmatrix} W'(0) \\ \frac{\mu}{\rho+\lambda}(\gamma-1) \\ \gamma-1 \\ 1-\gamma \end{bmatrix} + \beta_B + \gamma m.$$

So we can write the value function as

$$V_B(m;m_D) = \frac{n_B(m_D)}{\Delta(r_2-r_1)} + \beta_B + \gamma m,$$

where

$$n_B(m_D) = C_B\left[r_2 e^{r_1(m-m_D)} - r_1 e^{r_2(m-m_D)}\right] + (1-\gamma)(r_2-r_1)(e^{r_2 m} - e^{r_1 m})e^{-(r_1+r_2)m_D},$$

and

$$\begin{aligned} C_B &:= W'(0)(e^{-fr_1} - e^{-fr_2}) + (1-\gamma)\left(\left[\frac{\mu r_2}{\rho+\lambda} - 1\right]e^{-r_1(\overline{m}^*+f)} - \left[\frac{\mu r_1}{\rho+\lambda} - 1\right]e^{-r_2(\overline{m}^*+f)}\right) \\ &= W'(0)(e^{-fr_1} - e^{-fr_2}) + \frac{1}{2}\sigma^2\frac{\rho}{(\rho+\lambda)^2}\Gamma(\overline{m}^*+f). \end{aligned}$$

The derivative of n_B with respect to m_D is

$$n_B' = C_B r_1 r_2\left[e^{r_2(m-m_D)} - e^{r_1(m-m_D)}\right] - (1-\gamma)(r_2-r_1)(r_1+r_2)(e^{r_2 m} - e^{r_1 m})e^{-(r_1+r_2)m_D}.$$

Then the derivative of $V_B(m)$ with respect to m_D is

$$\frac{\partial V_B(m;m_D)}{\partial m_D} = \frac{1}{(r_2-r_1)\Delta^2}(n_B'\Delta - \Delta' n_B).$$

where the term in brackets is

$$n_B'\Delta - \Delta' n_B = (r_2-r_1)e^{-m_D(r_1+r_2)}(e^{r_2 m} - e^{r_1 m})\left[C_B r_1 r_2 - (1-\gamma)\Gamma\right].$$

Then we get

$$\frac{\partial V_B(m;m_D)}{\partial m_D} = \frac{e^{-m_D(r_1+r_2)}(e^{r_2 m} - e^{r_1 m})}{\Delta^2}\left[C_B r_1 r_2 - (1-\gamma)\Gamma\right].$$

B.4.3 Single crossing condition

When $m_D - f > \overline{m}^*$:

$$\frac{\partial V_G(m;m_D)}{\partial m_D} - \frac{\partial V_B(m;m_D)}{\partial m_D} = \frac{e^{-m_D(r_1+r_2)}(e^{r_2 m} - e^{r_1 m})}{\Delta^2} r_1 r_2 (C_G - C_B).$$

where

$$
\begin{aligned}
r_1 r_2 (C_G - C_B) &= W'(0) r_1 r_2 (e^{-f r_1} - e^{-f r_2}) + \left(\frac{\rho}{\rho+\lambda}\right)^2 [\Gamma(\overline{m}^* + f) - \Gamma(\overline{m}^*)] \\
&= -\Delta'(f) W'(0) + \left(\frac{\rho}{\rho+\lambda}\right)^2 [\Gamma(\overline{m}^* + f) - \Gamma(\overline{m}^*)].
\end{aligned}
$$

Note that the first term $-\Delta'(f)W'(0) < 0$ and the remaining term is positive, since $\Gamma'(x) = -r_1 r_2 \Delta(x) > 0$ for $x > 0$. For all m_D for which this expression is positive, increasing payout threshold can be used as a signaling device.

B.4.4 Separating equilibrium

The incentive compatibility constraint of the bad bank in this scenario becomes

$$
V_{B,B}^{FI}(\tilde{m}^S; m_B^*) \geq V_{B,G}(\tilde{m}^S; \tilde{m}^S). \tag{B.15}
$$

When this condition does not hold at $\tilde{m}^S = m_G^*$, the good bank will have to deviate from its privately optimal strategy m_G^* by choosing a higher payout boundary. The ICC of the good bank is

$$
V_{G,G}(\tilde{m}^S; \tilde{m}^S) \geq V_{G,B}(\tilde{m}^S; m_B^*). \tag{B.16}
$$

For $\tilde{m}^S$ to be a PBE, it is sufficient that the good bank does not have an incentive to defect to a different strategy under the pessimistic belief that the good bank is a bad bank instead. The corresponding condition becomes:

$$
V_{G,G}(\tilde{m}^S; \tilde{m}^S) \geq V_{G,B}(\tilde{m}^S; m_{G,B}^*). \tag{B.17}
$$

A separating equilibrium exists when there is a $\tilde{m}^S$ for which the three conditions Eq. (B.15), (B.16) and (B.17) are jointly satisfied.

B.4.5 Pooling equilibrium

The incentive compatibility constraint of the bad bank:

$$
V_{B,p}(\tilde{m}^P; \tilde{m}^P) \geq V_{B,B}(\tilde{m}^P; m_B^*).
$$

Restriction that the value of the good bank in the pooling equilibrium is larger than in the least-cost separating equilibrium:

$$
V_{G,p}(\tilde{m}^P; \tilde{m}^P) \geq \mathbb{1}_{\{m_G^* \geq \overline{m}^S\}} V_{G,G}(\tilde{m}^P; m_G^*) + \mathbb{1}_{\{m_G^* < \overline{m}^S\}} V_{G,G}(m^P; \overline{m}^S).
$$

where $\overline{m}^S$ is now the solution to the ICC of the bad bank.

Appendix to Chapter 3

C.1 Model solution

C.1.1 Proof of Proposition 12

The conditonal variance of the informed is unaffected by her own communication and the derivative of the certainty equivalent is proportional to the variance of the difference between informed beliefs and price

$$\frac{\partial CE_I}{\partial \sigma_{\eta_I}} = \frac{1}{2\alpha} \frac{Var[y|s_I]}{Var[E[y|s_I] - p]} \frac{\partial Var[E[y|s_I] - p]}{\partial \sigma_{\eta_I}}.$$

Using explicit expressions for uninformed beliefs, (see Appendix C.1.3) the variance of informed per dollar profit simplifies to

$$Var[E[y|s_I] - p] = \sigma_y^2 \beta_I^2 \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \left(\frac{|\mathbf{\Sigma}_\eta|}{|\mathbf{\Sigma}_{s_U}|} Var[s_I] + \lambda_p^2 \right),$$

see Appendix C.1.5. Since only the bracket depends on informed messaging noise σ_{η_I}, the derivative of the variance is proportional to the derivative of the bracket

$$\frac{\partial Var[E[y|s_I] - p]}{\partial \sigma_{\eta_I}} \propto Var[s_I] \frac{\partial}{\partial \sigma_{\eta_I}} \left\{ \frac{|\mathbf{\Sigma}_\eta|}{|\mathbf{\Sigma}_{s_U}|} \right\} + 2\lambda_p \frac{\partial \lambda_p}{\partial \sigma_{\eta_I}}. \tag{C.1}$$

Both derivatives in (C.1) are proportional to $\alpha\sigma_\varepsilon^2 \sigma_{\eta_Z} + \rho_\eta \sigma_{\eta_I}$, (see Appendix C.1.5) and the derivative is zero at

$$\sigma_{\eta_I}^* = -\frac{\alpha\sigma_\varepsilon^2 \sigma_{\eta_Z}}{\rho_\eta} \operatorname*{arg\,max}_{\sigma_{\eta_I}} CE_I.$$

This is a unique maximum since, applying the envelope theorem,

$$\left. \frac{\partial^2 CE_I}{\partial \sigma_{\eta_I}^2} \right|_{\sigma_{\eta_I}^*} \propto \frac{\partial \left\{ \alpha\sigma_\varepsilon^2 \sigma_{\eta_Z} + \rho_\eta \sigma_{\eta_I} \right\}}{\partial \sigma_{\eta_I}} = \rho_\eta < 0.$$

C.1.2 Informed certainty equivalent

By definition, the utility of the certainty equivalent must equal the expected utility of the risky gamble

$$E[U(w_I)] = U(CE_I) \implies CE_I = U^{-1}(E[U(w_I)]) = -\alpha^{-1} \ln(E[e^{-\alpha w_I}]).$$

Using the tower property of conditional expectations and the log-normality of the exponential of the negative of scaled wealth

$$E[e^{-aw_I}] = E\left[E[e^{-aw_I}|s_I, p]\right] = E\left[e^{-\alpha\{E[w_I|s_I,p]-\alpha\frac{1}{2}Var[w_I|s_I,p]\}}\right].$$

Substituting in the optimal demand $\delta_I = (E[y|s_I]-p)/(\alpha Var[y|s_I])$ the moments of wealth are

$$E[w_I|s_I, p] = \frac{(E[y|s_I]-p)E[(y-p)|s_I,p]}{\alpha Var[y|s_I]} = \frac{(E[y|s_I]-p)^2}{\alpha Var[y|s_I]},$$
$$Var[w_I|s_I, p] = \frac{(E[y|s_I]-p)^2 Var[y|s_I]}{\alpha^2 Var[y|s_I]^2} = \frac{(E[y|s_I]-p)^2}{\alpha^2 Var[y|s_I]},$$

and the negative expected utility is

$$E[e^{-\alpha w_I}] = E\left[e^{-\alpha\left\{\frac{(E[y|s_I]-p)^2}{\alpha Var[y|s_I]}-\alpha\frac{1}{2}\frac{(E[y|s_I]-p)^2}{\alpha^2 Var[y|s_I]}\right\}}\right] = E\left[e^{-\frac{1}{2}\frac{(E[y|s_I]-p)^2}{Var[y|s_I]}}\right].$$

It follows from of properties of squared normal variable (see Vives (2010), chapter 10.2.4) that for a mean zero normally distributed x the expected exponential of a scaled negative second degree polynomial of x is

$$E[e^{-d(ax^2+bx+c)}] = Var[x]^{-\frac{1}{2}}\left\{Var[x]^{-1}+2da\right\}^{-\frac{1}{2}} e^{-d\left[c-\frac{d}{2}\frac{b^2}{Var[x]^{-1}+2da}\right]}.$$

Letting $x = E[y|s_I]-p$, which is normally distributed with mean zero, pattern matching to the negative expected utility requires setting $a = 1, b, c = 0$ and $d = \frac{1}{2}Var[y|s_I]^{-1}$. Substituting in and rewriting yields

$$E[e^{-\alpha w_I}] = \left\{Var[E[y|s_I]-p]\left(Var[E[y|s_I]-p]^{-1}+Var[y|s_I]^{-1}\right)\right\}^{-\frac{1}{2}}$$
$$= \left(1+\frac{Var[E[y|s_I]-p]}{Var[y|s_I]}\right)^{-\frac{1}{2}}.$$

Substituting in and taking the natural log provides the result for the certainty equivalent

$$CE_I = -\alpha^{-1}\left[-\frac{1}{2}\ln\left(1+\frac{Var[E[y|s_I]-p]}{Var[y|s_I]}\right)\right] = \frac{1}{2\alpha}\ln\left(1+\frac{Var[E[y|s_I]-p]}{Var[y|s_I]}\right).$$

C.1.3 Uninformed beliefs

Since the covariance between the pay-off y and signal vector $\boldsymbol{s}_U$ simply is

$$Cov[\,y, \boldsymbol{s}_U] = \sigma_y^2\begin{pmatrix}1 & 1 & 0\end{pmatrix},$$

the conditional expectation and variance of the uninformed can be understood through the coefficients

$$\boldsymbol{\beta}_U^\top = Cov[y, \boldsymbol{s}_U]Var[s_U]^{-1} =: \begin{pmatrix}\beta_\zeta & \beta_{c_I} & \beta_{c_Z}\end{pmatrix}.$$

The conditional moments can be written as

$$E[y|\boldsymbol{s}_U] = \boldsymbol{\beta}_U^\top \boldsymbol{s}_U = (\beta_\zeta+\beta_{c_I})s_I - (\alpha\sigma_\varepsilon^2\beta_\zeta+\beta_{c_Z})z + \beta_{c_I}c_I + \beta_{c_Z}c_Z, \text{ and}$$
$$Var[y|\boldsymbol{s}_U] = \sigma_y^2[1-(\beta_\zeta+\beta_{c_I})].$$

Variance-covariance matrix of signal vector $\boldsymbol{s}_U$ is

$$\boldsymbol{\Sigma}_{s_U} = Var[\boldsymbol{s}_U] = \begin{pmatrix} Var[\zeta] & Var[s_I] & \alpha\sigma_\varepsilon^2\sigma_z^2 \\ Var[s_I] & Var[s_I]+\sigma_{\eta_I}^2 & \rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z} \\ \alpha\sigma_\varepsilon^2\sigma_z^2 & \rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z} & \sigma_z^2+\sigma_{\eta_Z}^2 \end{pmatrix}.$$

Define the vector $\boldsymbol{\kappa} = \begin{pmatrix}1 & \alpha\sigma_\varepsilon^2\end{pmatrix}^\top$, then the inverse of the variance-covariance matrix is

$$\boldsymbol{\Sigma}_{s_U}^{-1} = |\boldsymbol{\Sigma}_{s_U}|^{-1}\times$$
$$\begin{pmatrix} |\boldsymbol{\Sigma}_m| & -[Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)-\alpha\sigma_\varepsilon^2\sigma_z^2\rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z}] & Var[s_I]\rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z}-\alpha\sigma_\varepsilon^2\sigma_z^2(Var[s_I]+\sigma_{\eta_I}^2) \\ - & Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)+\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2 & -[Var[\zeta]\rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z}-\alpha\sigma_\varepsilon^2\sigma_z^2 Var[s_I]] \\ - & - & \alpha^2\sigma_\varepsilon^4(Var[s_I]+\sigma_{\eta_I}^2)+Var[s_I]\sigma_{\eta_I}^2 \end{pmatrix}$$

where $|\boldsymbol{\Sigma}_m| = |\boldsymbol{\Sigma}_\eta| + Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)+\sigma_z^2\sigma_{\eta_I}^2$, and

$|\boldsymbol{\Sigma}_{s_U}| = Var[\zeta|\boldsymbol{m}]|\boldsymbol{\Sigma}_m| = Var[\zeta]|\boldsymbol{\Sigma}_\eta| + Var[s_I]\sigma_z^2\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}$.

Explicitly, the coefficients $\boldsymbol{\beta}_U$ are given by

$$\begin{aligned}
\beta_\zeta &= \frac{\sigma_y^2}{|\boldsymbol{\Sigma}_{s_U}|}\left\{|\boldsymbol{\Sigma}_m| - [Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)-\alpha\sigma_\varepsilon^2\sigma_z^2\rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z}]\right\} \\
&= \frac{\sigma_y^2}{|\boldsymbol{\Sigma}_{s_U}|}\left\{|\boldsymbol{\Sigma}_\eta| + (\sigma_{\eta_I}^2+\alpha\sigma_\varepsilon^2\rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z})\sigma_z^2\right\} \\
\beta_{c_I} &= \frac{\sigma_y^2}{|\boldsymbol{\Sigma}_{s_U}|}\left\{Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)+\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2 - [Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)-\alpha\sigma_\varepsilon^2\sigma_z^2\rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z}]\right\} \\
&= \frac{\sigma_y^2}{|\boldsymbol{\Sigma}_{s_U}|}\alpha\sigma_\varepsilon^2\sigma_z^2\left\{\alpha\sigma_\varepsilon^2\sigma_{\eta_Z}^2+\rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z}\right\} \\
\beta_{c_Z} &= -\frac{\sigma_y^2}{|\boldsymbol{\Sigma}_{s_U}|}\alpha\sigma_\varepsilon^2\sigma_z^2\left\{\sigma_{\eta_I}^2+\alpha\sigma_\varepsilon^2\rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z}\right\},
\end{aligned}$$

and the sums of the conditional moments are

$$\begin{aligned}
\beta_\zeta+\beta_{c_I} &= \sigma_y^2\frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|}+\sigma_y^2\frac{\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}}{|\boldsymbol{\Sigma}_{s_U}|}\sigma_z^2, \\
\text{and } \alpha\sigma_\varepsilon^2\beta_\zeta+\beta_{c_Z} &= \sigma_y^2\frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|}\alpha\sigma_\varepsilon^2.
\end{aligned}$$

Finally, the conditional moments can be re-written as

$$\begin{aligned}
E[y|\boldsymbol{s}_U] &= \boldsymbol{\beta}_U^\top\boldsymbol{s}_U = \sigma_y^2\left(\frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|}+\frac{\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}\sigma_z^2}{|\boldsymbol{\Sigma}_{s_U}|}\right)s_I - \sigma_y^2\frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|}\alpha\sigma_\varepsilon^2 z + \beta_{c_I}\eta_I+\beta_{c_Z}\eta_Z, \text{ and} \\
Var[y|\boldsymbol{s}_U] &= \sigma_y^2\left[1-\sigma_y^2\left(\frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|}+\frac{\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}\sigma_z^2}{|\boldsymbol{\Sigma}_{s_U}|}\right)\right]
\end{aligned}$$

C.1.4 Uninformed beliefs under optimal communication strategy

Substituting in the optimal posting noise of Proposition 12 yields the following expressions

$$|\boldsymbol{\Sigma}_\eta||_{\sigma_{\eta_I}^*} = \frac{1-\rho_\eta^2}{\rho_\eta^2}\alpha^2\sigma_\varepsilon^4\sigma_{\eta_Z}^4, \quad \boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}|_{\sigma_{\eta_I}^*} = \left(\frac{1}{\rho_\eta^2}+1-2\right)\alpha^2\sigma_\varepsilon^4\sigma_{\eta_Z}^2 = \frac{1-\rho_\eta^2}{\rho_\eta^2}\alpha^2\sigma_\varepsilon^4\sigma_{\eta_Z}^2,$$

$$|\boldsymbol{\Sigma}_{s_U}||_{\sigma^*_{\eta_I}} = \frac{1-\rho_\eta^2}{\rho_\eta^2}\alpha^2\sigma_\varepsilon^4\sigma_{\eta_Z}^2\left(Var[\zeta]\sigma_{\eta_Z}^2 + Var[s_I]\sigma_z^2\right),$$

which result in the following coefficients

$$\beta_\zeta|_{\sigma^*_{\eta_I}} = \sigma_y^2\frac{\frac{1-\rho_\eta^2}{\rho_\eta^2}\alpha^2\sigma_\varepsilon^4\sigma_{\eta_Z}^4 + \frac{1-\rho_\eta^2}{\rho_\eta^2}\alpha^2\sigma_\varepsilon^4\sigma_{\eta_Z}^2\sigma_z^2}{\frac{1-\rho_\eta^2}{\rho_\eta^2}\alpha^2\sigma_\varepsilon^4\sigma_{\eta_Z}^2\left(Var[\zeta]\sigma_{\eta_Z}^2 + Var[s_I]\sigma_z^2\right)} = \sigma_y^2\frac{\sigma_z^2+\sigma_{\eta_Z}^2}{Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)+\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2}$$

$$\beta_{c_I}|_{\sigma^*_{\eta_I}} = 0$$

$$\beta_{c_Z}|_{\sigma^*_{\eta_I}} = -\sigma_y^2\frac{\alpha\sigma_\varepsilon^2\sigma_z^2}{Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)+\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2},$$

relevant sums of coefficients

$$\alpha\sigma_\varepsilon^2\beta_\zeta|_{\sigma^*_{\eta_I}} + \beta_{c_Z}|_{\sigma^*_{\eta_I}} = \sigma_y^2\frac{\alpha\sigma_\varepsilon^2\sigma_{\eta_Z}^2}{Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)+\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2},$$

and conditional variance

$$Var[y|\boldsymbol{s}_U]|_{\sigma^*_{\eta_I}} = \sigma_y^2(1-\beta_\zeta|_{\sigma^*_{\eta_I}}) = \sigma_y^2\sigma_\varepsilon^2\frac{\{1+\alpha^2\sigma_\varepsilon^2\sigma_z^2\}\sigma_{\eta_Z}^2+\sigma_z^2}{Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)+\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2}.$$

Optimal silence

Carrying out the calculations of Appendix C.1.3 for a signal vector that only contains noise trader communication $\{\boldsymbol{s}_U \setminus c_I\}$ yields a determinant of

$$|\boldsymbol{\Sigma}_{\boldsymbol{s}_U\setminus c_I}| = Var[\zeta](\sigma_z^2+\sigma_{\eta_Z}^2) - \alpha^2\sigma_\varepsilon^4\sigma_z^4 = Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)+\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2,$$

coefficients

$$\beta_{\zeta,\boldsymbol{s}_U\setminus c_I} = \sigma_y^2\frac{\sigma_z^2+\sigma_{\eta_Z}^2}{Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)+\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2} = \beta_\zeta|_{\sigma^*_{\eta_I}},\text{ and} \tag{C.2}$$

$$\beta_{c_Z,\boldsymbol{s}_U\setminus c_I} = -\sigma_y^2\frac{\alpha\sigma_\varepsilon^2\sigma_z^2}{Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)+\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2} = \beta_{c_Z}|_{\sigma^*_{\eta_I}}$$

and conditional variance is the same as under optimal communication noise $\sigma^*_{\eta_I}$ because of (C.2).

C.1.5 Variance of informed per dollar profit

The variance of the difference between informed conditional expectation and price is

$$Var[E[y|s_I]-p] = (1-\lambda_p)^2 Var[E[y|s_I]-E[y|\boldsymbol{s}_U]] + \lambda_p^2\beta_I^2\alpha^2\sigma_\varepsilon^4\sigma_z^2$$
$$-2\lambda_p(1-\lambda_p)\beta_I\alpha\sigma_\varepsilon^2 Cov[E[y|\boldsymbol{s}_U],z],$$

where $Cov[E[y|\boldsymbol{s}_U],z] = -\sigma_y^2\frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|}\alpha\sigma_\varepsilon^2\sigma_z^2$ and the variance of the difference of beliefs is

$$Var[E[y|s_I]-E[y|\boldsymbol{s}_U]]$$
$$= \beta_I^2 Var[s_I] + (\sigma_y^2 - Var[y|\boldsymbol{s}_U]) - 2\beta_I(\beta_\zeta+\beta_{c_I})Var[s_I]$$

$$
\begin{aligned}
&= \sigma_y^2 \left\{ \beta_I + \sigma_y^2 \left(\frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|} + \frac{\boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa} \sigma_z^2}{|\boldsymbol{\Sigma}_{s_U}|} \right) - 2\sigma_y^2 \left(\frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|} + \frac{\boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa} \sigma_z^2}{|\boldsymbol{\Sigma}_{s_U}|} \right) \right\} \\
&= \frac{\sigma_y^2}{|\boldsymbol{\Sigma}_{s_U}|} \beta_I \left(|\boldsymbol{\Sigma}_{s_U}| - Var[s_I] |\boldsymbol{\Sigma}_\eta| - Var[s_I] \boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa} \sigma_z^2 \right) \\
&= \sigma_y^2 \frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|} \beta_I \alpha^2 \sigma_\varepsilon^4 \sigma_z^2,
\end{aligned}
$$

using $Var[y|\boldsymbol{s}_U] = \sigma_y^2 - Var[E[y|\boldsymbol{s}_U]]$ and expressions for uninformed beliefs that can be found in Appendix C.1.3. Explicitly, the variance between informed conditional expectation and price can be written as

$$
\begin{aligned}
Var[E[y|s_I] - p] &= (1-\lambda_p)^2 \sigma_y^2 \frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|} \beta_I \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 + \lambda_p^2 \beta_I^2 \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \\
&\quad + 2\lambda_p (1-\lambda_p) \sigma_y^2 \frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|} \beta_I \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \\
&= \beta_I^2 \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \left(\left[(1-\lambda_p)^2 + 2\lambda_p(1-\lambda_p) \right] \frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|} Var[s_I] + \lambda_p^2 \right) \\
&= \beta_I^2 \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \left(\frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|} Var[s_I](1-\lambda_p^2) + \lambda_p^2 \right) \\
&= \beta_I^2 \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \left(\frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|} Var[s_I] + \lambda_p^2 \frac{|\boldsymbol{\Sigma}_{s_U}| - |\boldsymbol{\Sigma}_\eta| Var[s_I]}{|\boldsymbol{\Sigma}_{s_U}|} \right) \\
&= \beta_I^2 \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \left(\frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|} Var[s_I] + \lambda_p^2 \left\{ \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|} + Var[s_I] \sigma_z^2 \frac{\boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa}}{|\boldsymbol{\Sigma}_{s_U}|} \right\} \right).
\end{aligned}
$$

Derivative

The derivative of informed disagreement with price is

$$
\begin{aligned}
\frac{\partial Var[E[y|s_I] - p]}{\partial \sigma_{\eta_I}} &= \sigma_y^2 \beta_I^2 \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \frac{1}{\partial \sigma_{\eta_I}} \left\{ \frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|} Var[s_I] + \lambda_p^2 \right\} \\
&\propto Var[s_I] \frac{\partial}{\partial \sigma_{\eta_I}} \left\{ \frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|} \right\} + 2\lambda_p \frac{\partial \lambda_p}{\partial \sigma_{\eta_I}} \left(\alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|} + Var[s_I] \sigma_z^2 \frac{\boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa}}{|\boldsymbol{\Sigma}_{s_U}|} \right) \\
&\quad + \lambda_p^2 \left(\alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \frac{\partial}{\partial \sigma_{\eta_I}} \left\{ \frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|} \right\} + Var[s_I] \sigma_z^2 \frac{\partial}{\partial \sigma_{\eta_I}} \left\{ \frac{\boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa}}{|\boldsymbol{\Sigma}_{s_U}|} \right\} \right)
\end{aligned}
$$

The first derivative is

$$
\begin{aligned}
\frac{\partial}{\partial \sigma_{\eta_I}} \left\{ \frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|} \right\} &= \frac{1}{|\boldsymbol{\Sigma}_{s_U}|^2} \left(\frac{\partial |\boldsymbol{\Sigma}_\eta|}{\partial \sigma_{\eta_I}} |\boldsymbol{\Sigma}_{s_U}| - |\boldsymbol{\Sigma}_\eta| \frac{\partial |\boldsymbol{\Sigma}_{s_U}|}{\partial \sigma_{\eta_I}} \right) \\
&\propto 2 \frac{|\boldsymbol{\Sigma}_\eta|}{\sigma_{\eta_I}} |\boldsymbol{\Sigma}_{s_U}| - |\boldsymbol{\Sigma}_\eta| \left(2Var[\zeta] \frac{|\boldsymbol{\Sigma}_\eta|}{\sigma_{\eta_I}} + Var[s_I] \sigma_z^2 \frac{\partial \boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa}}{\partial \sigma_{\eta_I}} \right) \\
&= |\boldsymbol{\Sigma}_\eta| Var[s_I] \sigma_z^2 \left(2 \frac{\boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa}}{\sigma_{\eta_I}} - \frac{\partial \boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa}}{\partial \sigma_{\eta_I}} \right) \\
&= 2|\boldsymbol{\Sigma}_\eta| Var[s_I] \sigma_z^2 \frac{\alpha \sigma_\varepsilon^2}{\sigma_{\eta_I}} (\alpha \sigma_\varepsilon^2 \sigma_{\eta_Z} + \rho_\eta \sigma_{\eta_I}) \\
&\propto \alpha \sigma_\varepsilon^2 \sigma_{\eta_Z} + \rho_\eta \sigma_{\eta_I} \qquad \text{(C.3)}
\end{aligned}
$$

where the last equality follows from $\frac{\partial \boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa}}{\partial \sigma_{\eta_I}} = 2\sigma_{\eta_I} + 2\alpha\sigma_\varepsilon^2 \rho_\eta$. The second derivative is

$$\frac{\partial \lambda_p}{\partial \sigma_{\eta_I}} = -\lambda_p^2 \frac{\partial}{\partial \sigma_{\eta_I}} \left\{ \frac{Var[y|s_I]}{Var[y|\boldsymbol{s}_U]} \right\} = \lambda_p^2 \frac{Var[y|s_I]}{Var[y|\boldsymbol{s}_U]^2} \frac{\partial Var[y|\boldsymbol{s}_U]}{\partial \sigma_{\eta_I}},$$

where

$$\frac{\partial Var[y|\boldsymbol{s}_U]}{\partial \sigma_{\eta_I}} = -\sigma_y^2 \left(\frac{\partial}{\partial \sigma_{\eta_I}} \left\{ \frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|} \right\} + \frac{\partial}{\partial \sigma_{\eta_I}} \left\{ \frac{\boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa}}{|\boldsymbol{\Sigma}_{s_U}|} \right\} \sigma_z^2 \right).$$

The first of these two derivatives is the one shown above to be proportional to (C.3) and the second is

$$\begin{aligned}
\frac{\partial}{\partial \sigma_{\eta_I}} \left\{ \frac{\boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa}}{|\boldsymbol{\Sigma}_{s_U}|} \right\} &= \frac{1}{|\boldsymbol{\Sigma}_{s_U}|^2} \left(\frac{\partial \boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa}}{\partial \sigma_{\eta_I}} |\boldsymbol{\Sigma}_{s_U}| - \boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa} \frac{\partial |\boldsymbol{\Sigma}_{s_U}|}{\partial \sigma_{\eta_I}} \right) \\
&\propto Var[\zeta] |\boldsymbol{\Sigma}_\eta| \left(\frac{\partial \boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa}}{\partial \sigma_{\eta_I}} - 2 \frac{\boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa}}{\sigma_{\eta_I}} \right) \\
&= -2 Var[\zeta] |\boldsymbol{\Sigma}_\eta| \frac{\alpha \sigma_\varepsilon^2}{\sigma_{\eta_I}} \left(\rho_\eta \sigma_{\eta_I} + \alpha \sigma_\varepsilon^2 \sigma_{\eta_Z} \right) \\
&\propto -(\rho_\eta \sigma_{\eta_I} + \alpha \sigma_\varepsilon^2 \sigma_{\eta_Z}).
\end{aligned}$$

The two remaining derivatives are repetitions and the derivative of the variance of the difference between informed expectation and price is proportional to

$$\frac{\partial Var[E[y|s_I] - p]}{\partial \sigma_{\eta_I}} \propto \rho_\eta \sigma_{\eta_I} + \alpha \sigma_\varepsilon^2 \sigma_{\eta_Z}.$$

C.2 Model predictions

Rearranging price for weights on base shocks yields

$$p = [\lambda_p \beta_I + (1-\lambda_p)\beta_\zeta] s_I - \left\{ \alpha\sigma_\varepsilon^2 [\lambda_p \beta_I + (1-\lambda_p)\beta_\zeta] + (1-\lambda_p)\beta_{c_Z} \right\} z + (1-\lambda_p)\beta_{c_Z} \eta_Z$$

where

$$\begin{aligned}
\lambda_p \beta_I + (1-\lambda_p)\beta_\zeta &= \frac{\beta_I \left\{ \lambda_p [Var[s_I](\sigma_z^2 + \sigma_{\eta_Z}^2) + \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \sigma_{\eta_Z}^2] + (1-\lambda_p) Var[s_I](\sigma_z^2 + \sigma_{\eta_Z}^2) \right\}}{Var[s_I](\sigma_z^2 + \sigma_{\eta_Z}^2) + \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \sigma_{\eta_Z}^2} \\
&= \beta_I \frac{Var[s_I](\sigma_z^2 + \sigma_{\eta_Z}^2) + \lambda_p \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \sigma_{\eta_Z}^2}{Var[s_I](\sigma_z^2 + \sigma_{\eta_Z}^2) + \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \sigma_{\eta_Z}^2} = \beta_I \gamma_{s_I},
\end{aligned}$$

and

$$\begin{aligned}
&\left\{ \alpha\sigma_\varepsilon^2 [\lambda_p \beta_I + (1-\lambda_p)\beta_\zeta] + (1-\lambda_p)\beta_{c_Z} \right\} \\
&= \beta_I \frac{Var[s_I](\sigma_z^2 + \sigma_{\eta_Z}^2) + \lambda_p \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \sigma_{\eta_Z}^2 - (1-\lambda_p) Var[s_I] \sigma_z^2}{Var[s_I](\sigma_z^2 + \sigma_{\eta_Z}^2) + \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \sigma_{\eta_Z}^2} \alpha \sigma_\varepsilon^2 \\
&= \beta_I \frac{Var[s_I](\lambda_p \sigma_z^2 + \sigma_{\eta_Z}^2) + \lambda_p \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \sigma_{\eta_Z}^2}{Var[s_I](\sigma_z^2 + \sigma_{\eta_Z}^2) + \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \sigma_{\eta_Z}^2} \alpha \sigma_\varepsilon^2 = \beta_I \gamma_z \alpha \sigma_\varepsilon^2.
\end{aligned}$$

C.2.1 Return predictability

Covariance between sentiment and price and returns

Covariance between the two sources of communication and price and returns respectively are given by

$$\begin{aligned}
Cov[p, c_Z] &= \beta_I \gamma_z \alpha \sigma_\varepsilon^2 \sigma_z^2 + (1-\lambda_p)\beta_{c_Z}\sigma_{\eta Z}^2 \\
&= \beta_I \frac{Var[s_I](\lambda_p \sigma_z^2 + \sigma_{\eta Z}^2) + \lambda_p \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \sigma_{\eta Z}^2 - (1-\lambda_p)\sigma_{\eta Z}^2 Var[s_I]}{Var[s_I](\sigma_z^2 + \sigma_{\eta Z}^2) + \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \sigma_{\eta Z}^2} \alpha \sigma_\varepsilon^2 \sigma_z^2 \\
&= \beta_I \lambda_p \alpha \sigma_\varepsilon^2 \sigma_z^2 = \lambda_p \alpha \sigma_z^2 Var[y|s_I] \implies Cov[r, c_Z] = -\lambda_p \alpha \sigma_z^2 Var[y|s_I] \leq 0,
\end{aligned}$$

and

$$Cov[p, c_I] = \beta_I \gamma_{s_I} Var[s_I] \implies Cov[r, c_I] = \sigma_y^2 - \beta_I \gamma_{s_I} Var[s_I] = \sigma_y^2 (1 - \gamma_{s_I}) \geq 0.$$

Alternatively, using that

$$\lambda_p = \left\{1 + \frac{Var[y|s_I]}{Var[y|\boldsymbol{s}_U]}\right\}^{-1} \text{ and } Var[y|\boldsymbol{s}_U] = \frac{\sigma_y^2 \sigma_\varepsilon^2 \left(\{1 + \alpha^2 \sigma_\varepsilon^2 \sigma_z^2\} \sigma_{\eta Z}^2 + \sigma_z^2\right)}{Var[s_I](\sigma_z^2 + \sigma_{\eta Z}^2) + \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \sigma_{\eta Z}^2},$$

covariances with returns can be written as

$$\begin{aligned}
Cov[r, c_Z] &= -\alpha \sigma_z^2 Var[y|s_I] \left\{1 + \frac{Var[y|s_I]}{Var[y|\boldsymbol{s}_U]}\right\}^{-1}, \text{ and} \\
Cov[r, c_I] &= \sigma_y^2 \frac{(1-\lambda_p)\alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \sigma_{\eta Z}^2}{Var[s_I](\sigma_z^2 + \sigma_{\eta Z}^2) + \alpha^2 \sigma_\varepsilon^4 \sigma_z^2 \sigma_{\eta Z}^2} \\
&= \frac{\alpha^2 \sigma_\varepsilon^2 \sigma_z^2 \sigma_{\eta Z}^2}{\{1 + \alpha^2 \sigma_\varepsilon^2 \sigma_z^2\} \sigma_{\eta Z}^2 + \sigma_z^2} Var[y|s_I] \left\{1 + \frac{Var[y|s_I]}{Var[y|\boldsymbol{s}_U]}\right\}^{-1} \\
&= -\frac{\alpha \sigma_\varepsilon^2 \sigma_{\eta Z}^2}{\{1 + \alpha^2 \sigma_\varepsilon^2 \sigma_z^2\} \sigma_{\eta Z}^2 + \sigma_z^2} Cov[r, c_Z].
\end{aligned}$$

Positive relation between predicted returns and mixed sentiment

To have a positive relationship between mixed sentiment and predicted returns, the covariance between the two must be positive. Using expressions for covariances found in Appendix C.2.1, such positivity depends on the following condition

$$\begin{aligned}
Cov[r, c_s] = Cov[r, c_I] + Cov[r, c_Z] &= \left(\frac{\alpha \sigma_\varepsilon^2 \sigma_{\eta Z}^2}{\{1 + \alpha^2 \sigma_\varepsilon^2 \sigma_z^2\} \sigma_{\eta Z}^2 + \sigma_z^2} - 1\right) Cov[p, c_Z] > 0 \\
\iff \alpha \sigma_\varepsilon^2 \sigma_{\eta Z}^2 &> \{1 + \alpha^2 \sigma_\varepsilon^2 \sigma_z^2\} \sigma_{\eta Z}^2 + \sigma_z^2,
\end{aligned}$$

for which the following conditions are necessary

$$\alpha \sigma_\varepsilon^2 > 1 \iff \sigma_\varepsilon^2 > \frac{1}{\alpha}, \quad \alpha \sigma_\varepsilon^2 > \alpha^2 \sigma_\varepsilon^2 \sigma_z^2 \iff \frac{1}{\alpha} > \sigma_z^2.$$

Taking limits of $Cov[r, c_S]$ yield

$$\lim_{\sigma_z \to 0} Cov[r, c_S] \propto \alpha \sigma_\varepsilon^2 - 1, \quad \lim_{\sigma_\varepsilon \to \infty} Cov[r, c_S] \propto 1 - \alpha \sigma_z^2.$$

T-stats

The t-stat of regressing target r on communication c is given by

$$t\{E[r|c]\} = \frac{Cov[r,c]}{\sqrt{Var[r]Var[c] - Cov[r,c]^2}},$$

therefore, the respective t-stats with the sentiment of the two equilibria are

$$\begin{aligned} t\{E[r|c_Z]\} &= \frac{Cov[r,c_Z]}{\sqrt{Var[r]Var[c_Z] - Cov[r,c_Z]^2}} \propto -\lambda_p \alpha \sigma_z^2 Var[y|s_I], \text{ and} \\ t\{E[r|c_s]\} &= \frac{Cov[r,c_I] + Cov[r,c_Z]}{\sqrt{Var[r]Var[c_s] - Cov[r,c_s]^2}} \propto \sigma_y^2(1-\gamma_{s_I}) - \lambda_p \alpha \sigma_z^2 Var[y|s_I] \end{aligned}$$

using covariances derived in Appendix C.2.1.

Proof of Proposition 14

The magnitude of the t-stat on the sentiment of only noise trader communication is larger than the t-stat with mixed communication if

$$\begin{aligned} |t\{E[r|c_Z]\}| &> |t\{E[r|c_s]\}| \\ \frac{|Cov[r,c_Z]|}{\sqrt{Var[r]Var[c_Z] - Cov[r,c_Z]^2}} &> \frac{|Cov[r,c_s]|}{\sqrt{Var[r]Var[c_s] - Cov[r,c_s]^2}} \\ Cov[r,c_Z]^2(Var[r]Var[c_s] - Cov[r,c_s]^2) &> Cov[r,c_s]^2(Var[r]Var[c_Z] - Cov[r,c_Z]^2) \\ Cov[r,c_Z]^2 Var[c_s] &> Cov[r,c_s]^2 Var[c_Z] \\ Cov[r,c_Z]^2 Var[c_s] &> (Cov[r,c_Z] + Cov[r,c_I])^2 Var[c_Z] \\ Cov[r,c_Z]^2(Var[c_s] - Var[c_Z]) &> (Cov[r,c_I]^2 + 2Cov[r,c_Z]Cov[r,c_I])Var[c_Z] \end{aligned}$$

where the difference in variance is positive

$$Var[c_s] - Var[c_Z] = Var[s_I] + \sigma_{\eta_Z}^2(\alpha^2\sigma_\varepsilon^4/\rho_\eta^2 + 2\alpha\sigma_\varepsilon^2) \geq 0.$$

Writing $Cov[r,c_I]$ in terms of $Cov[r,c_Z]$ as in Appendix C.2.1, the inequality simplifies to

$$\begin{aligned} & Var[s_I] + \sigma_{\eta_Z}^2\left(\frac{\alpha^2\sigma_\varepsilon^4}{\rho_\eta^2} + 2\alpha\sigma_\varepsilon^2\right) \\ & > \left\{\left(\frac{\alpha\sigma_\varepsilon^2\sigma_{\eta_Z}^2}{\{1+\alpha^2\sigma_\varepsilon^2\sigma_z^2\}\sigma_{\eta_Z}^2 + \sigma_z^2}\right)^2 - \frac{\alpha\sigma_\varepsilon^2\sigma_{\eta_Z}^2}{\{1+\alpha^2\sigma_\varepsilon^2\sigma_z^2\}\sigma_{\eta_Z}^2 + \sigma_z^2}\right\} Var[c_Z], \end{aligned}$$

and, since $\rho_\eta^2 \in (0,1)$, a sufficient condition for the inequality to hold is

$$\begin{aligned} 1 &\geq \frac{\sigma_{\eta_Z}^2 Var[c_Z]}{\left(\{1+\alpha^2\sigma_\varepsilon^2\sigma_z^2\}\sigma_{\eta_Z}^2 + \sigma_z^2\right)^2} = \frac{\sigma_{\eta_Z}^2(\sigma_z^2 + \sigma_{\eta_Z}^2)}{\left(\{1+\alpha^2\sigma_\varepsilon^2\sigma_z^2\}\sigma_{\eta_Z}^2 + \sigma_z^2\right)^2} \\ \sigma_{\eta_Z}^2(\sigma_z^2 + \sigma_{\eta_Z}^2) &\leq \{1+\alpha^2\sigma_\varepsilon^2\sigma_z^2\}\sigma_{\eta_Z}^4 + \sigma_z^4 + 2\{1+\alpha^2\sigma_\varepsilon^2\sigma_z^2\}\sigma_{\eta_Z}^2\sigma_z^2 \\ 0 &\leq \alpha^2\sigma_\varepsilon^2\sigma_z^2\sigma_{\eta_Z}^4 + \sigma_z^4 + 2\alpha^2\sigma_\varepsilon^2\sigma_z^2\sigma_{\eta_Z}^2\sigma_z^2, \end{aligned}$$

which is trivially satisfied.

C.3 Model variations

C.3.1 Noise trader communication decision

Following Han and Yang (2013), the certainty equivalent of the noise trader is captured by their expected opportunity cost

$$
\begin{aligned}
CE_Z &= E[-z(y-p)] = E[zp] = -\lambda_p\beta_I\alpha\sigma_\varepsilon^2\sigma_z^2 - (1-\lambda_p)(\beta_\zeta\alpha\sigma_\varepsilon^2 + \beta_{c_Z})\sigma_z^2 \\
&= -\left\{\lambda_p\beta_I + (1-\lambda_p)\sigma_y^2\frac{|\mathbf{\Sigma}_\eta|}{|\mathbf{\Sigma}_{s_U}|}\right\}\alpha\sigma_\varepsilon^2\sigma_z^2 = -\left\{\lambda_p + (1-\lambda_p)Var[s_I]\frac{|\mathbf{\Sigma}_\eta|}{|\mathbf{\Sigma}_{s_U}|}\right\}\beta_I\alpha\sigma_\varepsilon^2\sigma_z^2
\end{aligned}
$$

with

$$
\frac{\partial CE_Z}{\partial\sigma_{\eta_Z}} = -\frac{\partial\lambda_p}{\partial\sigma_{\eta_Z}}\left\{1 - Var[s_I]\frac{|\mathbf{\Sigma}_\eta|}{|\mathbf{\Sigma}_{s_U}|}\right\}\beta_I\alpha\sigma_\varepsilon^2\sigma_z^2 - (1-\lambda_p)Var[s_I]\frac{\partial}{\partial\sigma_{\eta_Z}}\left\{\frac{|\mathbf{\Sigma}_\eta|}{|\mathbf{\Sigma}_{s_U}|}\right\}\beta_I\alpha\sigma_\varepsilon^2\sigma_z^2
$$

so

$$
\frac{\partial CE_Z}{\partial\sigma_{\eta_Z}} = 0 \iff \frac{\partial\lambda_p}{\partial\sigma_{\eta_Z}}\left\{1 - \frac{Var[s_I]|\mathbf{\Sigma}_\eta|}{|\mathbf{\Sigma}_{s_U}|}\right\} = -(1-\lambda_p)Var[s_I]\frac{\partial}{\partial\sigma_{\eta_Z}}\left\{\frac{|\mathbf{\Sigma}_\eta|}{|\mathbf{\Sigma}_{s_U}|}\right\} \tag{C.4}
$$

where $\frac{Var[s_I]|\mathbf{\Sigma}_\eta|}{|\mathbf{\Sigma}_{s_U}|} \leq 1$, so either the signs of the derivatives must be opposite at the extremum or they must both be zero. It turns out that they are proportional to the same sum (see Appendix C.3.1)

$$
\frac{\partial\lambda_p}{\partial\sigma_{\eta_Z}} \propto \frac{\partial}{\partial\sigma_{\eta_Z}}\left\{\frac{|\mathbf{\Sigma}_\eta|}{|\mathbf{\Sigma}_{s_U}|}\right\} \propto \sigma_{\eta_I} + \alpha\sigma_\varepsilon^2\rho_\eta\sigma_{\eta_Z},
$$

so, if $\rho_\eta < 0$, an extremum exists at

$$
\sigma_{\eta_Z} = -\frac{\sigma_{\eta_I}}{\alpha\sigma_\varepsilon^2\rho_\eta}.
$$

However, in contrast to the optimal communication strategy of the informed investor described in Proposition 12, the extremum is a minimum since, applying the envelope theorem

$$
\left.\frac{\partial^2 CE_Z}{\partial\sigma_{\eta_Z}^2}\right|_{\sigma_{\eta_Z}^*} =\propto -\frac{\partial\{\sigma_{\eta_I} + \alpha\sigma_\varepsilon^2\rho_\eta\sigma_{\eta_Z}\}}{\partial\sigma_{\eta_Z}} = -\rho_\eta > 0.
$$

Instead the noise traders prefer a limiting case. The certainty equivalent of the two limiting cases are (see Appendix C.3.1 and Appendix C.3.1)

$$
\begin{aligned}
&\lim_{\sigma_{\eta_Z}\to 0} CE_Z = -\frac{1}{2}\beta_I\alpha\sigma_\varepsilon^2\sigma_z^2, \text{ and} \\
&\lim_{\sigma_{\eta_Z}\to\infty} CE_Z = -\frac{Var[s_I](1-\rho_\eta^2)\sigma_{\eta_I}^2 + \lambda_p\{(1-\rho_\eta^2)\sigma_{\eta_I}^2 + Var[s_I]\}\alpha^2\sigma_\varepsilon^4\sigma_z^2}{Var[\zeta](1-\rho_\eta^2)\sigma_{\eta_I}^2 + Var[s_I]\sigma_z^2\alpha^2\sigma_\varepsilon^4}\beta_I\alpha\sigma_\varepsilon^2\sigma_z^2.
\end{aligned}
$$

In the limiting case of noise trader posting noise going to zero the uninformed investor can perfectly extract the informed investors' signal from price. This is the limiting case one would expect the noise traders to prefer as it provides more volume to execute against. This is indeed the case since

$$
\begin{aligned}
&\frac{1}{2} < \frac{Var[s_I](1-\rho_\eta^2)\sigma_{\eta_I}^2 + \lambda_p\{(1-\rho_\eta^2)\sigma_{\eta_I}^2 + Var[s_I]\}\alpha^2\sigma_\varepsilon^4\sigma_z^2}{Var[\zeta](1-\rho_\eta^2)\sigma_{\eta_I}^2 + Var[s_I]\sigma_z^2\alpha^2\sigma_\varepsilon^4} \\
&0 < 2Var[s_I](1-\rho_\eta^2)\sigma_{\eta_I}^2 + 2\lambda_p\{(1-\rho_\eta^2)\sigma_{\eta_I}^2 + Var[s_I]\}\alpha^2\sigma_\varepsilon^4\sigma_z^2
\end{aligned}
$$

$$
\begin{aligned}
&- Var[\zeta](1-\rho_\eta^2)\sigma_{\eta_I}^2 - Var[s_I]\sigma_z^2\alpha^2\sigma_\varepsilon^4 \\
&= Var[s_I](1-\rho_\eta^2)\sigma_{\eta_I}^2 + (2\lambda_p - 1)\{(1-\rho_\eta^2)\sigma_{\eta_I}^2 + Var[s_I]\}\alpha^2\sigma_\varepsilon^4\sigma_z^2
\end{aligned}
$$

and a sufficient condition for positivity of the right hand side is $\lambda_p \geq \frac{1}{2}$, which must hold generally as (3.2) in Section 3.2.1.

Derivatives with respect to noise trader posting noise

Basic derivatives are

$$
\begin{aligned}
\frac{\partial|\boldsymbol{\Sigma}_\eta|}{\partial\sigma_{\eta_Z}} &= 2\frac{|\boldsymbol{\Sigma}_\eta|}{\sigma_{\eta_Z}}, \\
\frac{\partial\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}}{\partial\sigma_{\eta_Z}} &= 2\left(\alpha^2\sigma_\varepsilon^4\sigma_{\eta_Z} + \alpha\sigma_\varepsilon^2\rho_\eta\sigma_{\eta_I}\right) = 2\alpha\sigma_\varepsilon^2\left(\alpha\sigma_\varepsilon^2\sigma_{\eta_Z} + \rho_\eta\sigma_{\eta_I}\right) \\
\frac{\partial|\boldsymbol{\Sigma}_{s_U}|}{\partial\sigma_{\eta_Z}} &= Var[\zeta]\frac{\partial|\boldsymbol{\Sigma}_\eta|}{\partial\sigma_{\eta_Z}} + Var[s_I]\sigma_z^2\frac{\partial\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}}{\partial\sigma_{\eta_Z}} \\
&= 2\left\{Var[\zeta]|\boldsymbol{\Sigma}_\eta| + Var[s_I]\sigma_z^2\alpha\sigma_\varepsilon^2\left(\alpha\sigma_\varepsilon^2\sigma_{\eta_Z} + \rho_\eta\sigma_{\eta_I}\right)\right\}.
\end{aligned}
$$

Derivative of the price weight is

$$
\frac{\partial\lambda_p}{\partial\sigma_{\eta_Z}} = \lambda_p^2\frac{Var[y|s_I]}{Var[y|\boldsymbol{s}_U]^2}\frac{\partial Var[y|\boldsymbol{s}_U]}{\partial\sigma_{\eta_Z}} \propto \sigma_{\eta_I} + \alpha\sigma_\varepsilon^2\rho_\eta\sigma_{\eta_Z}
$$

since

$$
\begin{aligned}
\frac{\partial Var[y|\boldsymbol{s}_U]}{\partial\sigma_{\eta_Z}} &= -\sigma_y^4\frac{\partial}{\partial\sigma_{\eta_Z}}\left\{\frac{|\boldsymbol{\Sigma}_\eta| + \boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}\sigma_z^2}{|\boldsymbol{\Sigma}_{s_U}|}\right\} \\
&= \frac{\sigma_y^4}{|\boldsymbol{\Sigma}_{s_U}|^2}\left\{-\left(\frac{\partial|\boldsymbol{\Sigma}_\eta|}{\partial\sigma_{\eta_Z}} + \frac{\partial\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}}{\partial\sigma_{\eta_Z}}\sigma_z^2\right)|\boldsymbol{\Sigma}_{s_U}| + \left(|\boldsymbol{\Sigma}_\eta| + \boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}\sigma_z^2\right)\frac{\partial|\boldsymbol{\Sigma}_{s_U}|}{\partial\sigma_{\eta_Z}}\right\} \\
&\propto \frac{\partial|\boldsymbol{\Sigma}_\eta|}{\partial\sigma_{\eta_Z}}\left\{-(Var[\zeta]|\boldsymbol{\Sigma}_\eta| + Var[s_I]\sigma_z^2\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}) + Var[\zeta](|\boldsymbol{\Sigma}_\eta| + \sigma_z^2\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa})\right\} \\
&+ \sigma_z^2\frac{\partial\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}}{\partial\sigma_{\eta_Z}}\left\{-(Var[\zeta]|\boldsymbol{\Sigma}_\eta| + Var[s_I]\sigma_z^2\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}) + Var[s_I](|\boldsymbol{\Sigma}_\eta| + \sigma_z^2\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa})\right\} \\
&= \frac{\partial|\boldsymbol{\Sigma}_\eta|}{\partial\sigma_{\eta_Z}}\alpha^2\sigma_\varepsilon^4\sigma_z^4\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa} - \frac{\partial\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}}{\partial\sigma_{\eta_Z}}\alpha^2\sigma_\varepsilon^4\sigma_z^4|\boldsymbol{\Sigma}_\eta| \\
&= |\boldsymbol{\Sigma}_\eta|\alpha^2\sigma_\varepsilon^4\sigma_z^4\left(2\frac{\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}}{\sigma_{\eta_Z}} - \frac{\partial\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}}{\partial\sigma_{\eta_Z}}\right) \\
&= 2|\boldsymbol{\Sigma}_\eta|\frac{\alpha^2\sigma_\varepsilon^4\sigma_z^4}{\sigma_{\eta_Z}}\left(\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa} - \alpha^2\sigma_\varepsilon^4\sigma_{\eta_Z}^2 - \alpha\sigma_\varepsilon^2\rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z}\right) \\
&= 2|\boldsymbol{\Sigma}_\eta|\alpha^2\sigma_\varepsilon^4\sigma_z^4\frac{\sigma_{\eta_I}}{\sigma_{\eta_Z}}\left(\sigma_{\eta_I} + \alpha\sigma_\varepsilon^2\rho_\eta\sigma_{\eta_Z}\right).
\end{aligned}
$$

Derivative of the ratio in (C.4) is

$$
\begin{aligned}
\frac{\partial}{\partial\sigma_{\eta_Z}}\left\{\frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|}\right\} &= \boldsymbol{\Sigma}_{s_U}^{-2}\left(\frac{\partial|\boldsymbol{\Sigma}_\eta|}{\partial\sigma_{\eta_Z}}|\boldsymbol{\Sigma}_{s_U}| - |\boldsymbol{\Sigma}_\eta|\frac{\partial|\boldsymbol{\Sigma}_{s_U}|}{\partial\sigma_{\eta_Z}}\right) \\
&\propto \frac{\partial|\boldsymbol{\Sigma}_\eta|}{\partial\sigma_{\eta_Z}}\left(Var[\zeta]|\boldsymbol{\Sigma}_\eta| + Var[s_I]\sigma_z^2\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa} - Var[\zeta]|\boldsymbol{\Sigma}_\eta|\right)
\end{aligned}
$$

$$
\begin{aligned}
&- |\boldsymbol{\Sigma}_\eta| Var[s_I]\sigma_z^2 \frac{\partial \boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa}}{\partial \sigma_{\eta_Z}} \\
&= |\boldsymbol{\Sigma}_\eta| Var[s_I]\sigma_z^2 \left(2\frac{\boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa}}{\sigma_{\eta_Z}} - \frac{\partial \boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa}}{\partial \sigma_{\eta_Z}} \right) \\
&= 2|\boldsymbol{\Sigma}_\eta| Var[s_I]\sigma_z^2 \frac{\sigma_{\eta_I}}{\sigma_{\eta_Z}} \left(\sigma_{\eta_I} + \alpha\sigma_\varepsilon^2 \rho_\eta \sigma_{\eta_Z} \right).
\end{aligned}
$$

Disappearing posting noise

Basic limits are

$$
\lim_{\sigma_{\eta_Z}\to 0} |\boldsymbol{\Sigma}_\eta| = 0, \quad \lim_{\sigma_{\eta_Z}\to 0} \boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa} = \sigma_{\eta_I}^2, \quad \lim_{\sigma_{\eta_Z}\to 0} |\boldsymbol{\Sigma}_{s_U}| = Var[s_I]\sigma_z^2 \sigma_{\eta_I}^2
$$

so

$$
\lim_{\sigma_{\eta_Z}\to 0} Var[y|\boldsymbol{s}_U] = \sigma_y^2 \left[1 - \frac{\sigma_y^2}{Var[s_I]} \right] = Var[y|s_I], \quad \lim_{\sigma_{\eta_Z}\to 0} \lambda_p = \frac{1}{2}
$$

and ultimately

$$
\implies \lim_{\sigma_{\eta_Z}\to 0} CE_Z = -\frac{1}{2}\beta_I \alpha \sigma_\varepsilon^2 \sigma_z^2.
$$

Exploding posting noise

Basic limits are

$$
\begin{aligned}
\lim_{\sigma_{\eta_Z}\to\infty} \frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|} &= \frac{(1-\rho_\eta^2)\sigma_{\eta_I}^2}{Var[\zeta](1-\rho_\eta^2)\sigma_{\eta_I}^2 + Var[s_I]\sigma_z^2\alpha^2\sigma_\varepsilon^4} \\
\lim_{\sigma_{\eta_Z}\to\infty} \frac{\boldsymbol{\kappa}^\top \boldsymbol{\Sigma}_\eta \boldsymbol{\kappa}}{|\boldsymbol{\Sigma}_{s_U}|} &= \frac{\alpha^2\sigma_\varepsilon^4}{Var[\zeta](1-\rho_\eta^2)\sigma_{\eta_I}^2 + Var[s_I]\sigma_z^2\alpha^2\sigma_\varepsilon^4}
\end{aligned}
$$

so

$$
\begin{aligned}
&\lim_{\sigma_{\eta_Z}\to\infty} Var[y|\boldsymbol{s}_U] = \sigma_y^2 \left[1 - \sigma_y^2 \frac{(1-\rho_\eta^2)\sigma_{\eta_I}^2 + \sigma_z^2\alpha^2\sigma_\varepsilon^4}{Var[\zeta](1-\rho_\eta^2)\sigma_{\eta_I}^2 + Var[s_I]\sigma_z^2\alpha^2\sigma_\varepsilon^4} \right] > Var[y|s_I] \\
&\implies \lim_{\sigma_{\eta_Z}\to\infty} \lambda_p > \frac{1}{2}
\end{aligned}
$$

and ultimately

$$
\lim_{\sigma_{\eta_Z}\to\infty} CE_Z = -\frac{Var[s_I](1-\rho_\eta^2)\sigma_{\eta_I}^2 + \lambda_p\{(1-\rho_\eta^2)\sigma_{\eta_I}^2 + Var[s_I]\}\alpha^2\sigma_\varepsilon^4\sigma_z^2}{Var[\zeta](1-\rho_\eta^2)\sigma_{\eta_I}^2 + Var[s_I]\sigma_z^2\alpha^2\sigma_\varepsilon^4} \beta_I \alpha \sigma_\varepsilon^2 \sigma_z^2.
$$

C.3.2 Uninformed beliefs when informed message is only noise

The relevant moments of the uninformed signal vector $\boldsymbol{s}_U$ with $c_I = \eta_I$ are

$$
Cov[y, \boldsymbol{s}_U] = \sigma_y^2 \begin{pmatrix} 1 & 0 & 0 \end{pmatrix},
$$

$$\boldsymbol{\Sigma}_{s_U} = \begin{pmatrix} Var[\zeta] & 0 & \alpha\sigma_\varepsilon^2\sigma_z^2 \\ 0 & \sigma_{\eta_I}^2 & \rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z} \\ \alpha\sigma_\varepsilon^2\sigma_z^2 & \rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z} & \sigma_z^2+\sigma_{\eta_Z}^2 \end{pmatrix}$$

and

$$\boldsymbol{\Sigma}_{s_U}^{-1} = |\boldsymbol{\Sigma}_{s_U}|^{-1} \begin{pmatrix} |\boldsymbol{\Sigma}_m| & \alpha\sigma_\varepsilon^2\sigma_z^2\rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z} & -\alpha\sigma_\varepsilon^2\sigma_z^2\sigma_{\eta_I}^2 \\ \alpha\sigma_\varepsilon^2\sigma_z^2\rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z} & Var[\zeta](\sigma_z^2+\sigma_{\eta_Z}^2)-\alpha^2\sigma_\varepsilon^4\sigma_z^4 & -Var[\zeta]\rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z} \\ -\alpha\sigma_\varepsilon^2\sigma_z^2\sigma_{\eta_I}^2 & -Var[\zeta]\rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z} & Var[\zeta]\sigma_{\eta_I}^2 \end{pmatrix}$$

where

$$\begin{aligned} |\boldsymbol{\Sigma}_{s_U}| &= Var[\zeta|\boldsymbol{m}]||\boldsymbol{\Sigma}_m| \\ &= Var[\zeta]|\boldsymbol{\Sigma}_m| - \begin{pmatrix} 0 & \alpha\sigma_\varepsilon^2\sigma_z^2 \end{pmatrix} \begin{pmatrix} \sigma_z^2+\sigma_{\eta_Z}^2 & -\rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z} \\ -\rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z} & \sigma_{\eta_I}^2 \end{pmatrix} \begin{pmatrix} 0 \\ \alpha\sigma_\varepsilon^2\sigma_z^2 \end{pmatrix} \\ &= (Var[s_I]+\alpha^2\sigma_\varepsilon^4\sigma_z^2)(|\boldsymbol{\Sigma}_\eta|+\sigma_z^2\sigma_{\eta_I}^2) - \alpha^2\sigma_\varepsilon^4\sigma_z^4\sigma_{\eta_I}^2 \\ &= Var[\zeta]|\boldsymbol{\Sigma}_\eta| + Var[s_I]\sigma_z^2\sigma_{\eta_I}^2 = |\boldsymbol{\Sigma}_\eta|Var[\zeta] + |\boldsymbol{\Sigma}_{x_\zeta}|\sigma_{\eta_I}^2 \end{aligned}$$

so

$$\begin{aligned} \beta_\zeta &= \sigma_y^2|\boldsymbol{\Sigma}_{s_U}|^{-1}|\boldsymbol{\Sigma}_m| \\ \beta_{m_I} &= \sigma_y^2|\boldsymbol{\Sigma}_{s_U}|^{-1}|\alpha\sigma_\varepsilon^2\sigma_z^2\rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z} \\ \beta_{m_Z} &= -\sigma_y^2|\boldsymbol{\Sigma}_{s_U}|^{-1}\alpha\sigma_\varepsilon^2\sigma_z^2\sigma_{\eta_I}^2 \end{aligned}$$

and there is no interior optimum.

C.4 Model extension

C.4.1 Uninformed investor trading on aggregate sentiment

The additional uninformed investor is signified by the subscript U_2 whereas the subscript for the uninformed investor of the model in the main text will be U_1, similarly versions of variables related to the extension are subscripted 2 and 1 for the main model, see Section 3.4 for more details.

Conditional mean and variance are given by

$$\begin{aligned} E[y|\boldsymbol{s}_{U_2}] &= \sigma_y^2 \begin{pmatrix} 1+\frac{\sigma_\varepsilon^2(\beta_{\zeta_1}+\beta_{c_I})}{Var[y|\boldsymbol{s}_{U_1}]} & 1 \end{pmatrix} \boldsymbol{\Sigma}_{s_{U_2}}^{-1}\boldsymbol{s}_{U_2} \\ \boldsymbol{\Sigma}_{s_{U_2}} &= \begin{pmatrix} Var[\zeta_2] & Cov[\zeta_2,c_S] \\ Cov[\zeta_2,c_S] & Var[s_I]+\sigma_z^2+\boldsymbol{1}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{1} \end{pmatrix} \\ Var[\zeta_2] &= Var[\zeta_1] + \frac{\sigma_\varepsilon^4\sigma_y^4}{Var[y|\boldsymbol{s}_{U_1}]^2} \begin{pmatrix} 1 & 1 & 0 \end{pmatrix} \boldsymbol{\Sigma}_{s_{U_1}}^{-1} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \\ &\quad +2\frac{\sigma_\varepsilon^2\left(\beta_{\zeta_1}Var[\zeta_1]+\beta_{c_I}Var[s_I]-\beta_{c_Z}\alpha\sigma_\varepsilon^2\sigma_z^2\right)}{Var[y|\boldsymbol{s}_{U_1}]} \\ Cov[\zeta_2,c_S] &= Var[s_I]+\alpha\sigma_\varepsilon^2\sigma_z^2+\frac{\sigma_\varepsilon^2}{Var[y|\boldsymbol{s}_{U_1}]}\beta_{\zeta_1}(Var[s_I]+\alpha\sigma_\varepsilon^2\sigma_z^2) \\ &\quad +\frac{\sigma_\varepsilon^2}{Var[y|\boldsymbol{s}_{U_1}]}\left(\beta_{c_I}(Var[s_I]+\sigma_{\eta_I}^2+\rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z})+\beta_{c_Z}(\sigma_z^2+\sigma_{\eta_Z}^2+\rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z})\right) \end{aligned}$$

$$Var[y|\boldsymbol{s}_{U_2}] = \sigma_y^2 \left(1 - \sigma_y^2 \begin{pmatrix} 1 + \frac{\sigma_\varepsilon^2(\beta_{\zeta_1}+\beta_{c_I})}{Var[y|\boldsymbol{s}_{U_1}]} & 1 \end{pmatrix} \boldsymbol{\Sigma}_{s_{U_2}}^{-1} \begin{pmatrix} 1 + \frac{\sigma_\varepsilon^2(\beta_{\zeta_1}+\beta_{c_I})}{Var[y|\boldsymbol{s}_{U_1}]} \\ 1 \end{pmatrix} \right)$$

where

$$\begin{aligned}
\frac{\sigma_\varepsilon^2\left(\beta_{\zeta_1}+\beta_{c_I}\right)}{Var[y|\boldsymbol{s}_{U_1}]} &= \frac{\sigma_\varepsilon^2\sigma_y^2\left(\frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|}+\frac{\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}}{|\boldsymbol{\Sigma}_{s_U}|}\sigma_z^2\right)}{\sigma_y^2\left[1-\sigma_y^2\left(\frac{|\boldsymbol{\Sigma}_\eta|}{|\boldsymbol{\Sigma}_{s_U}|}+\frac{\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}\sigma_z^2}{|\boldsymbol{\Sigma}_{s_U}|}\right)\right]} = \frac{\sigma_\varepsilon^2\left(|\boldsymbol{\Sigma}_\eta|+\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}\sigma_z^2\right)}{|\boldsymbol{\Sigma}_{s_U}|-\sigma_y^2\left(|\boldsymbol{\Sigma}_\eta|+\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}\sigma_z^2\right)} \\
&= \frac{\sigma_\varepsilon^2\left(|\boldsymbol{\Sigma}_\eta|+\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}\sigma_z^2\right)}{Var[\zeta_1]|\boldsymbol{\Sigma}_\eta|+Var[s_I]\sigma_z^2\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}-\sigma_y^2\left(|\boldsymbol{\Sigma}_\eta|+\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}\sigma_z^2\right)} \\
&= \frac{\sigma_\varepsilon^2\left(|\boldsymbol{\Sigma}_\eta|+\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}\sigma_z^2\right)}{\sigma_\varepsilon^2\left(1+\alpha^2\sigma_\varepsilon^2\sigma_z^2\right)|\boldsymbol{\Sigma}_\eta|+\sigma_\varepsilon^2\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}\sigma_z^2} = \frac{|\boldsymbol{\Sigma}_\eta|+\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}\sigma_z^2}{\left(1+\alpha^2\sigma_\varepsilon^2\sigma_z^2\right)|\boldsymbol{\Sigma}_\eta|+\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\kappa}\sigma_z^2} < 1.
\end{aligned}$$

Denoting the price of the main text by p_1, market clearing with the additional investor yields price p_2

$$\begin{aligned}
z &= \frac{E[y|s_I]-p_2}{\alpha Var[y|s_I]} + \frac{E[y|\boldsymbol{s}_{U_1}]-p_2}{\alpha Var[y|\boldsymbol{s}_{U_1}]} + \frac{E[y|\boldsymbol{s}_{U_2}]-p_2}{\alpha Var[y|\boldsymbol{s}_{U_2}]} \\
\implies p_2 &= \frac{\frac{E[y|s_I]-\alpha Var[y|s_I]z}{Var[y|s_I]} + \frac{E[y|\boldsymbol{s}_{U_1}]}{Var[y|\boldsymbol{s}_{U_1}]} + \frac{E[y|\boldsymbol{s}_{U_2}]}{Var[y|\boldsymbol{s}_{U_2}]}}{\frac{1}{Var[y|s_I]} + \frac{1}{Var[y|\boldsymbol{s}_{U_1}]} + \frac{1}{Var[y|\boldsymbol{s}_{U_2}]}} \\
&= \lambda_{p_2}\left\{\lambda_{p_1}(E[y|s_I]-\alpha Var[y|s_I]z) + (1-\lambda_{p_1})E[y|\boldsymbol{s}_{U_1}]\right\} + (1-\lambda_{p_2})E[y|\boldsymbol{s}_{U_2}] \\
&= \lambda_{p_2}p_1 + (1-\lambda_{p_2})E[y|\boldsymbol{s}_{U_2}]
\end{aligned}$$

where λ_{p_1} is the price weight of the main text and the additional price weight λ_{p_2} is given by

$$\lambda_{p_2} = \frac{\frac{1}{Var[y|s_I]} + \frac{1}{Var[y|\boldsymbol{s}_{U_1}]}}{\frac{1}{Var[y|s_I]} + \frac{1}{Var[y|\boldsymbol{s}_{U_1}]} + \frac{1}{Var[y|\boldsymbol{s}_{U_2}]}}.$$

Variance of the dollar profit margin is given by

$$\begin{aligned}
Var\left[E[y|s_I]-p_2\right] = \lambda_{p_2}^2 Var\left[E[y|s_I]-p_1\right] &+ (1-\lambda_{p_2})^2 Var\left[E[y|s_I]-E[y|\boldsymbol{s}_{U_2}]\right] \\
&+ 2\lambda_{p_2}(1-\lambda_{p_2})Cov\left[E[y|s_I]-p_1, E[y|s_I]-E[y|\boldsymbol{s}_{U_2}]\right],
\end{aligned}$$

where

$$\begin{aligned}
&Var[E[y|s_I]-E[y|\boldsymbol{s}_{U_2}]] \\
&= \beta_I^2 Var[s_I] + (\sigma_y^2 - Var[y|\boldsymbol{s}_{U_2}]) - 2\beta_I\boldsymbol{\beta}_{U_2}^\top \begin{pmatrix} Var[s_I] + \frac{\sigma_\varepsilon^2}{Var[y|\boldsymbol{s}_{U_1}]}Cov[s_I, E[y|\boldsymbol{s}_{U_1}]] \\ Var[s_I] \end{pmatrix} \\
&= \beta_I^2 Var[s_I] + \sigma_y^4 \begin{pmatrix} 1 + \frac{\sigma_\varepsilon^2(\beta_{\zeta_1}+\beta_{c_I})}{Var[y|\boldsymbol{s}_{U_1}]} & 1 \end{pmatrix} \boldsymbol{\Sigma}_{s_{U_2}}^{-1} \begin{pmatrix} 1 + \frac{\sigma_\varepsilon^2(\beta_{\zeta_1}+\beta_{c_I})}{Var[y|\boldsymbol{s}_{U_1}]} \\ 1 \end{pmatrix} \\
&\quad - 2\beta_I\boldsymbol{\beta}_{U_2}^\top \begin{pmatrix} 1 + \frac{\sigma_\varepsilon^2(\beta_{\zeta_1}+\beta_{c_I})}{Var[y|\boldsymbol{s}_{U_1}]} \\ 1 \end{pmatrix} Var[s_I] \\
&= \sigma_y^4 \left(\frac{1}{Var[s_I]} - \begin{pmatrix} 1 + \frac{\sigma_\varepsilon^2(\beta_{\zeta_1}+\beta_{c_I})}{Var[y|\boldsymbol{s}_{U_1}]} & 1 \end{pmatrix} \boldsymbol{\Sigma}_{s_{U_2}}^{-1} \begin{pmatrix} 1 + \frac{\sigma_\varepsilon^2(\beta_{\zeta_1}+\beta_{c_I})}{Var[y|\boldsymbol{s}_{U_1}]} \\ 1 \end{pmatrix} \right)
\end{aligned}$$

and

$$
\begin{aligned}
&Cov\left[E[y|s_I]-p_1, E[y|s_I]-E[y|\boldsymbol{s}_{U_2}]\right] \\
&\quad= Cov\left[(1-\lambda_{p_1})(E[y|s_I]-E[y|\boldsymbol{s}_U])+\lambda_{p_1}\alpha Var[y|s_I]z, \beta_I s_I-\beta_{\zeta_2}\zeta_2-\beta_{c_S}c_S\right] \\
&\quad= Cov\Bigg[(1-\lambda_{p_1})(\beta_I s_I-\beta_{\zeta_1}\zeta_1-\beta_{c_I}c_I-\beta_{c_Z}c_Z)+\lambda_{p_1}\alpha Var[y|s_I]z, \\
&\qquad\qquad \beta_I s_I-\beta_{\zeta_2}\left(\zeta_1+\frac{\sigma_\varepsilon^2}{Var[y|\boldsymbol{s}_{U_1}]}E[y|\boldsymbol{s}_{U_1}]\right)-\beta_{c_S}(c_I+c_Z)\Bigg] \\
&\quad= Cov\Bigg[(1-\lambda_{p_1})\{\beta_I-\beta_{\zeta_1}-\beta_{c_I}\}s_I \\
&\qquad\qquad +\left\{(1-\lambda_{p_1})(\beta_{\zeta_1}\alpha\sigma_\varepsilon^2+\beta_{c_Z})+\lambda_{p_1}\alpha Var[y|s_I]\right\}z \\
&\qquad\qquad -(1-\lambda_{p_1})\beta_{c_I}\eta_I-(1-\lambda_{p_1})\beta_{c_Z}\eta_Z, \\
&\qquad\qquad \left\{\beta_I-\beta_{\zeta_2}\left(1+\frac{\sigma_\varepsilon^2(\beta_{\zeta_1}+\beta_{c_I})}{Var[y|\boldsymbol{s}_{U_1}]}\right)-\beta_{c_S}\right\}s_I \\
&\qquad\qquad +\left\{\beta_{\zeta_2}\left(\alpha\sigma_\varepsilon^2+\frac{\sigma_\varepsilon^2(\beta_{\zeta_1}\alpha\sigma_\varepsilon^2+\beta_{c_Z})}{Var[y|\boldsymbol{s}_{U_1}]}\right)+\beta_{c_S}\right\}z \\
&\qquad\qquad -\left(\beta_{\zeta_2}\beta_{c_I}+\beta_{c_S}\right)\eta_I-\left(\beta_{\zeta_2}\beta_{c_Z}+\beta_{c_S}\right)\eta_Z\Bigg] \\
&\quad= (1-\lambda_{p_1})\{\beta_I-\beta_{\zeta_1}-\beta_{c_I}\}\left\{\beta_I-\beta_{\zeta_2}\left(1+\frac{\sigma_\varepsilon^2(\beta_{\zeta_1}+\beta_{c_I})}{Var[y|\boldsymbol{s}_{U_1}]}\right)-\beta_{c_S}\right\}Var[s_I] \\
&\qquad +\left\{(1-\lambda_{p_1})(\beta_{\zeta_1}\alpha\sigma_\varepsilon^2+\beta_{c_Z})+\lambda_{p_1}\alpha Var[y|s_I]\right\}\left\{\beta_{\zeta_2}\left(\alpha\sigma_\varepsilon^2+\frac{\sigma_\varepsilon^2(\beta_{\zeta_1}\alpha\sigma_\varepsilon^2+\beta_{c_Z})}{Var[y|\boldsymbol{s}_{U_1}]}\right)+\beta_{c_S}\right\}\sigma_z^2 \\
&\qquad\qquad +(1-\lambda_{p_1})Cov\left[\beta_{c_I}\eta_I+\beta_{c_Z}\eta_Z, \left(\beta_{\zeta_2}\beta_{c_I}+\beta_{c_S}\right)\eta_I+\left(\beta_{\zeta_2}\beta_{c_Z}+\beta_{c_S}\right)\eta_Z\right]
\end{aligned}
$$

where

$$
\begin{aligned}
&Cov\left[\beta_{c_I}\eta_I+\beta_{c_Z}\eta_Z, \left(\beta_{\zeta_2}\beta_{c_I}+\beta_{c_S}\right)\eta_I+\left(\beta_{\zeta_2}\beta_{c_Z}+\beta_{c_S}\right)\eta_Z\right] \\
&\quad= \beta_{c_I}\left(\beta_{\zeta_2}\beta_{c_I}+\beta_{c_S}\right)\sigma_{\eta_I}^2+\beta_{c_Z}\left(\beta_{\zeta_2}\beta_{c_Z}+\beta_{c_S}\right)\sigma_{\eta_Z}^2 \\
&\qquad +\left\{\beta_{c_I}\left(\beta_{\zeta_2}\beta_{c_Z}+\beta_{c_S}\right)+\beta_{c_Z}\left(\beta_{\zeta_2}\beta_{c_I}+\beta_{c_S}\right)\right\}\rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z} \\
&\quad= \beta_{c_I}\left(\beta_{\zeta_2}\beta_{c_I}+\beta_{c_S}\right)\sigma_{\eta_I}^2+\beta_{c_Z}\left(\beta_{\zeta_2}\beta_{c_Z}+\beta_{c_S}\right)\sigma_{\eta_Z}^2+\left\{2\beta_{\zeta_2}\beta_{c_Z}\beta_{c_I}+\beta_{c_S}\left(\beta_{c_I}+\beta_{c_Z}\right)\right\}\rho_\eta\sigma_{\eta_I}\sigma_{\eta_Z}.
\end{aligned}
$$

Price variance in the no communication equilibrium is

$$
\begin{aligned}
Var[E[y|s_I]-p_{2,s_U\setminus c_I}] = {}& \lambda_{p_{2,s_U\setminus c_I}}^2 Var[E[y|s_I]-E[y|\zeta_1,c_Z]]+(1-\lambda_{p_{2,s_U\setminus c_I}})^2\beta_I^2\alpha^2\sigma_\varepsilon^4\sigma_z^2 \\
&+2\lambda_{p_{2,s_U\setminus c_I}}(1-\lambda_{p_{2,s_U\setminus c_I}})\beta_I\alpha\sigma_\varepsilon^2 Cov[E[y|\zeta_1,c_Z],z],
\end{aligned}
$$

where

$$
\begin{aligned}
&Var[E[y|s_I]-E[y|\zeta_1,c_Z]] \\
&= \beta_I^2 Var[s_I]+\sigma_y^2\beta_{\zeta_1,\boldsymbol{s}_U\setminus c_I}-2\beta_I\beta_{\zeta_1,\boldsymbol{s}_U\setminus c_I}Var[s_I] \\
&= \frac{\sigma_y^4}{Var[s_I]}-\sigma_y^2\beta_{\zeta_1,\boldsymbol{s}_U\setminus c_I} = \frac{\sigma_y^4}{Var[s_I]}-\sigma_y^4\frac{\sigma_z^2+\sigma_{\eta_Z}^2}{Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)+\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2} \\
&= \sigma_y^4\frac{Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)+\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2-Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)}{Var[s_I]\left(Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)+\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2\right)}
\end{aligned}
$$

$$= \frac{\sigma_y^4}{Var[s_I]} \frac{\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2}{Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)+\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2} = \beta_I^2 \frac{\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2 Var[s_I]}{Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)+\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2}$$

and

$$Cov[E[y|\zeta_1, c_Z], z] = -(\alpha\sigma_\varepsilon^2\beta_{\zeta_1,\boldsymbol{s}_U\backslash c_I} + \beta_{c_Z,\boldsymbol{s}_U\backslash c_I})\sigma_z^2 = -\sigma_y^2 \frac{\alpha\sigma_\varepsilon^2\sigma_z^2\sigma_{\eta_Z}^2}{Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)+\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2}.$$

Therefore, price variance can be written as

$$\begin{aligned} Var[E[y|s_I] - p_{2,s_U\backslash c_I}] &= (1-\lambda_{p_2,s_U\backslash c_I})^2\beta_I^2 \frac{\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2 Var[s_I]}{Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)+\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2} \\ &\quad + \lambda_{p_2,s_U\backslash c_I}^2\beta_I^2\alpha^2\sigma_\varepsilon^4\sigma_z^2 \\ &\quad + 2\lambda_{p_2,s_U\backslash c_I}(1-\lambda_{p_2,s_U\backslash c_I})\beta_I\sigma_y^2 \frac{\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2}{Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)+\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2} \\ &= \beta_I^2\alpha^2\sigma_\varepsilon^4\sigma_z^2 \left\{ \frac{(1-\lambda_{p_2,s_U\backslash c_I}^2)\sigma_{\eta_Z}^2 Var[s_I]}{Var[s_I](\sigma_z^2+\sigma_{\eta_Z}^2)+\alpha^2\sigma_\varepsilon^4\sigma_z^2\sigma_{\eta_Z}^2} + \lambda_{p_2,s_U\backslash c_I}^2 \right\}. \end{aligned}$$

Bibliography

Acharya, V. V., Cooley, T., Richardson, M., Walter, I., 2009. Manufacturing Tail Risk: A Perspective on the Financial Crisis of 2007-2009. Foundations and Trends in Finance 4, 247–325.

Acharya, V. V., Gujral, I., Kulkarni, N., Shin, H. S., 2011. Dividends and Bank Capital in the Financial Crisis of 2007-2009. NBER Working Paper Series 16896.

Acharya, V. V., Le, H. T., Shin, H. S., 2017. Bank Capital and Dividend Externalities. The Review of Financial Studies 30, 988–1018.

Admanti, A. R., Pfleiderer, P., 1991. Sunshine trading and financial market equilibrium. The Review of Financial Studies 4, 443–481.

Admati, A. R., 1985. A Noisy Rational Expectations Equilibrium for Multi-Asset Securities Markets. Econometrica 53, 629.

Admati, A. R., DeMarzo, P. M., Hellwig, M. F., Pfleiderer, P., 2013. Fallacies, Irrelevant Facts, and Myths in the Discussion of Capital Regulation: Why Bank Equity is Not Socially Expensive. Rock Center for Corporate Governance 161.

Al-Najjar, 2009. Decision Makers as Statisticians: Diversity, Ambiguity, and Learning. Econometrica 77, 1371–1401.

Antweiler, W., Frank, M. Z., 2004. Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. The Journal of Finance 59, 1259–1294.

Aragones, E., Gilboa, I., Postlewaite, A., Schmeidler, D., 2005. Fact-Free Learning. American Economic Review 95, 1355–1368.

Araujo, A., Moreira, H., Tsuchida, M., 2011. Do dividend changes signal future earnings? Journal of Financial Intermediation 20, 117–134.

Avramov, D., Cheng, S., Metzker, L., 2022. Machine Learning Versus Economic Restrictions: Evidence from Stock Return Predictability. Management Science .

Bai, J., Philippon, T., Savov, A., 2016. Have financial markets become more informative? Journal of Financial Economics 122, 625–654.

Baker, M., Wurgler, J., 2007. Investor Sentiment in the Stock Market. Journal of Economic Perspectives 21, 129–151.

Bandara, W., 2016. The Information Content of Pre-Open Social Media. SSRN Electronic Journal .

Banerjee, S., 2011. Learning from Prices and the Dispersion in Beliefs. Review of Financial Studies 24, 3025–3068.

Banerjee, S., Kaniel, R., Kremer, I., 2009. Price Drift as an Outcome of Differences in Higher-Order Beliefs. Review of Financial Studies 22, 3707–3734.

Barth, A., Moreno–Bromberg, S., Reichmann, O., 2016. A Non-stationary Model of Dividend Distribution in a Stochastic Interest-Rate Setting. Computational Economics 47, 447–472.

Bartov, E., Faurel, L., Mohanram, P. S., 2018. Can Twitter Help Predict Firm-Level Earnings and Stock Returns? The Accounting Review 93, 25–57.

Beck, T., Mazzaferro, F., Portes, R., Quin, J., Schett, C., 2020. Preserving capital in the financial sector to weather the storm. VOXEU.org .

Behrendt, S., Schmidt, A., 2018. The Twitter myth revisited: Intraday investor sentiment, Twitter activity and individual-level stock return volatility. Journal of Banking & Finance 96, 355–367.

Bessler, W., Nohel, T., 2000. Asymmetric information, dividend reductions, and contagion effects in bank stock returns. Journal of Banking and Finance 24, 1831–1848.

Bhattacharya, S., 1979. Imperfect Information, Dividend Policy, and "The Bird in the Hand" Fallacy. The Bell Journal of Economics 10, 259–270.

Biais, B., Bossaerts, P., Spatt, C., 2010. Equilibrium Asset Pricing and Portfolio Choice Under Asymmetric Information. Review of Financial Studies 23, 1503–1543.

Bianchi, J., Bigio, S., 2014. Banks, Liquidity Management and Monetary Policy. NBER Working Paper Series 20490.

Boldin, R., Leggett, K., 1995. Bank dividend policy as a signal of bank quality. Financial Services Review 4, 1–8.

Bollen, J., Mao, H., Zeng, X., 2011. Twitter mood predicts the stock market. Journal of Computational Science 2, 1–8.

Breon-Drish, B., 2015. On Existence and Uniqueness of Equilibrium in a Class of Noisy Rational Expectations Models. The Review of Economic Studies 82, 868–921.

Brunnermeier, M., Farhi, E., Koijen, R. S. J., Krishnamurthy, A., Ludvigson, S. C., Lustig, H., Nagel, S., Piazzesi, M., 2021. Review Article: Perspectives on the Future of Asset Pricing. The Review of Financial Studies 34, 2126–2160.

Brunnermeier, M. K., Parker, J. A., 2005. Optimal Expectations. The American Economic Review 95, 1092–1118.

Buncic, D., Tischhauser, M., 2017. Macroeconomic factors and equity premium predictability. International Review of Economics & Finance 51, 621–644.

Campbell, J., Grossman, S. J., Wang, J., 1993. Trading volume and serial correlation in stock returns. The Quarterly Journal of Economics 108, 905–939.

Campbell, J. Y., Thompson, S. B., 2008. Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? Review of Financial Studies 21, 1509–1531.

Campbell, T., Kracaw, W., 1980. Information Production, Market Signalling, and the Theory of Financial Intermediation. The Journal of Finance 35, 863–882.

Cao, H. H., Coval, J. D., Hirshleifer, D., 2002. Sidelined Investors, Trading-Generated News, and Security Returns. The Review of Financial Studies 15.

Cejnek, G., Randl, O., Zechner, J., 2020. The Covid-19 Pandemic and Corporate Dividend Policy 1, 1–21.

Checkley, M., Higón, D. A., Alles, H., 2017. The hasty wisdom of the mob: How market sentiment predicts stock market behavior. Expert Systems with Applications 77, 256–263.

Chen, H., De, P., Hu, Y. J., Hwang, B.-H., 2014. Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media. Review of Financial Studies 27, 1367–1403.

Cho, I.-K., Kreps, D. M., 1987. Signaling Games and Stable Equilibria. The Quarterly Journal of Economics 102, 179–222.

Cochrane, J. H., 2011. Presidential Address: Discount Rates. The Journal of Finance 66, 1047–1108.

Colla, P., Mele, A., 2010. Information Linkages and Correlated Trading. Review of Financial Studies 23, 203–246.

Cookson, J. A., Niessner, M., 2020. Why Don't We Agree? Evidence from a Social Network of Investors. The Journal of Finance 75, 173–228.

Crank, B. Y. J., Nicolson, P., 1947. A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type. Mathematical Proceedings of the Cambridge Philosophical Society 43, 50–67.

Crawford, S., Gray, W., Johnson, B. R., Price, R. A., 2018. What Motivates Buy-Side Analysts to Share Recommendations Online? Management Science 64, 2574–2589.

Das, S. R., Chen, M. Y., 2007. Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. Management Science 53, 1375–1388.

de Prado, M. L., 2018. THE 10 REASONS MOST MACHINE LEARNING FUNDS FAIL. The Journal of Portfolio Management 44, 120–133.

DeAngelo, H., DeAngelo, L., Skinner, D. J., 2007. Corporate payout policy. Foundations and Trends in Finance 3, 95–287.

Décamps, J. P., Mariotti, T., Rochet, J. C., Villeneuve, S., 2011. Free Cash Flow, Issuance Costs, and Stock Prices. The Journal of Finance 66, 1501–1544.

Demarzo, P. M., Vayanos, D., Zwiebel, J., 2003. Persuasion bias, social influence, and unidimensional opinions. The Quarterly journal of economics 118, 909–968.

Dessaint, O., Foucault, T., Frésard, L., 2020. Does Big Data Improve Financial Forecasting? The Horizon Effect .

Dickens, R., Casey, K. M., Newman, J., 2002. Bank Dividend Policy: Explanatory Factors. Quarterly Journal of Business and Economics 41, 3–12.

Didisheim, A., Ke, S., Kelly, B. T., Malamud, S., 2023. Complexity in Factor Pricing Models. SSRN Electronic Journal .

Duffie, D., Malamud, S., Manso, G., 2009. Information Percolation With Equilibrium Search Dynamics. Econometrica 77, 1513–1574.

Dugast, J., Foucault, T., 2020. Equilibrium Data Mining and Data Abundance. SSRN Electronic Journal .

Edlin, A. S., Shannon, C., 1998. Strict Single Crossing and the Strict Spence-Mirrlees Condition: A Comment on Monotone Comparative Statics. Econometrica 66, 1417–1425.

Eierle, B., Klamer, S., Muck, M., 2022. Does it really pay off for investors to consider information from social media? International Review of Financial Analysis 81, 102074.

Emerson, S., Kennedy, R., O'Shea, L., O'Brien, J., 2019. Trends and Applications of Machine Learning in Quantitative Finance. In: *8th International Conference on Economics and Finance Research.*

European Central Bank, 2020. Recommendation of the European Central Bank of 15 December 2020 on dividend distributions during the COVID-19 pandemic and repealing Recommendation ECB/2020/35 .

European Systemic Risk Board, 2020a. Liquidity risks arising from margin calls.

European Systemic Risk Board, 2020b. System-wide restraints on dividend payments, share buybacks and other pay-outs.

Farboodi, M., Matray, A., Veldkamp, L., Venkateswaran, V., 2020. Where Has All the Data Gone? NBER .

Farboodi, M., Veldkamp, L., 2020. Long-Run Growth of Financial Data Technology. American Economic Review 110, 2485–2523.

Farboodi, M., Veldkamp, L., 2021. A Growth Model of the Data Economy p. 52.

Farrell, M., Green, T. C., Jame, R., Markov, S., 2022. The democratization of investment research and the informativeness of retail investor trading. Journal of Financial Economics 145, 616–641.

Farrell, M., Jame, R., Qiu, T., 2020. The Cross-Section of Non-Professional Analyst Skill. SSRN Electronic Journal .

Federal Reserve System, 2020. Federal Reserve Board releases results of stress tests for 2020 and additional sensitivity analyses conducted in light of the coronavirus event. Press Release June 25.

Federal Reserve System, 2021. Federal Reserve announces temporary and additional restrictions on bank holding company dividends and share repurchases (...). Press Release March 25.

Fekrazad, A., Harun, S. M., Sardar, N., 2022. Social media sentiment and the stock market. Journal of Economics and Finance 46, 397–419.

Ferson, W. E., Harvey, C. R., 1999. Conditioning Variables and the Cross Section of Stock Returns. The Journal of Finance 54, 1325–1360.

Financial Times, 2021. Outlook for US airlines remains cloudy despite further $15bn aid injection. March 3.

Flannery, M. J., Kwan, S. H., Nimalendran, M., 2004. Market evidence on the opaqueness of banking firms' assets. Journal of Financial Economics 71, 419–460.

Floyd, E., Li, N., Skinner, D. J., 2015. Payout policy through the financial crisis: The growth of repurchases and the resilience of dividends. Journal of Financial Economics 118, 299–316.

Forti, C., Schiozer, R. F., 2015. Bank dividends and signaling to information-sensitive depositors. Journal of Banking and Finance 56, 1–11.

Fulghieri, P., García, D., Hackbarth, D., 2020. Asymmetric Information and the Pecking (Dis)Order. Review of Finance 24, 961–996.

Gabaix, X., 2014. A Sparsity-Based Model of Bounded Rationality *. The Quarterly Journal of Economics 129, 1661–1710.

Gagliardini, P., Ma, H., 2019. Extracting Statistical Factors When Betas are Time-Varying. SSRN Electronic Journal .

Gambacorta, L., Oliviero, T., Shin, H. S., 2020. Low price-to-book ratios and bank dividend payout policies. BIS Working Papers 907.

Garrett, I., Priestley, R., 2012. Dividend Growth, Cash Flow, and Discount Rate News. The Journal of Financial and Quantitative Analysis 47, 1003–1028.

Gentzkow, M., Kelly, B., Taddy, M., 2019. Text as Data. Journal of Economic Literature 57, 535–574.

Gerber, H. U., Shiu, E. S., 2004. Optimal Dividends: Analysis with Brownian Motion. North American Actuarial Journal 8.

Giannini, R., Irvine, P., Shu, T., 2018. Nonlocal Disadvantage: An Examination of Social Media Sentiment. The Review of Asset Pricing Studies 8, 293–336.

Giannini, R., Irvine, P., Shu, T., 2019. The convergence and divergence of investors' opinions around earnings news: Evidence from a social network. Journal of Financial Markets 42, 94–120.

Gilboa, I., Schmeidler, D., 1989. Maxmin expected utility with non-unique prior. Journal of mathematical economics 18, 141–153.

Gillet, R., Lapointe, M. A., Raimbourg, P., 2008. Dividend Policy and Reputation. Journal of Business Finance and Accounting 35, 516–540.

Goldman, E., Martel, J., Schneemeier, J., 2022. A theory of financial media. Journal of Financial Economics 145, 239–258.

Goldstein, I., Spatt, C. S., Ye, M., 2021. Big Data in Finance. The Review of Financial Studies 34, 3213–3225.

Goldstein, I., Xiong, Y., Yang, L., 2020. Whoever Has Will Be Given More: Information Sharing in Financial Markets. SSRN Electronic Journal .

Goodhart, C. A., Peiris, M. U., Tsomocos, D. P., Vardoulakis, A. P., 2010. On dividend restrictions and the collapse of the interbank market. Annals of Finance 6, 455–473.

Green, E. J., Marshall, R. C., Marx, L. M., 2014. Tacit Collusion in Oligopoly. Oxford University Press.

Grossman, S. J., Stiglitz, J. E., 1980. On the Impossibility of Informationally Efficient Markets. The American Economic Review 70, 393–408.

Gu, C., Kurov, A., 2020. Informational role of social media: Evidence from Twitter sentiment. Journal of Banking & Finance 121, 105969.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical Asset Pricing via Machine Learning. The Review of Financial Studies 33, 2223–2273.

Gu, S., Kelly, B., Xiu, D., 2021. Autoencoder asset pricing models. Journal of Econometrics 222, 429–450.

Guntay, L., Jacewitz, S., Pogach, J., 2015. Proving Approval: Bank Dividends, Regulation, and Runs. FDIC CFR Working Paper Series 2015-05.

Guttman, I., Kadan, O., Kandel, E., 2010. Dividend Stickiness and Strategic Pooling. The Review of Financial Studies 23, 4455–4495.

Hammerschmid, R., Lohre, H., 2018. Regime shifts and stock return predictability. International Review of Economics & Finance 56, 138–160.

Han, B., Yang, L., 2013. Social Networks, Information Acquisition, and Asset Prices. Management Science 59, 1444–1457.

Han, Y., Yang, K., Zhou, G., 2013. A New Anomaly: The Cross-Sectional Profitability of Technical Analysis. Journal of Financial and Quantitative Analysis 48, 1433–1461.

Hardy, B., 2021. Covid-19 bank dividend payout restrictions: effects and trade-offs. BIS Bulletin 38.

Harford, J., Klasa, S., Maxwell, W. F., 2014. Refinancing Risk and Cash Holdings. The Journal of Finance 69.

Harvey, C. R., 2017. Presidential Address: The Scientific Outlook in Financial Economics: Scientific Outlook in Finance. The Journal of Finance 72, 1399–1440.

Hastie, T., Tibshirani, R., Friedman, J. H., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, vol. 2. New York: springer.

Hellwig, M. F., 1980. On the aggregation of information in competitive markets. Journal of Economic Theory 22, 477–498.

Hendershott, T., Livdan, D., Schürhoff, N., 2015. Are institutions informed about news? Journal of Financial Economics 117, 249–287.

Hendershott, T., Menkveld, A. J., 2014. Price pressures. Journal of Financial Economics 114, 405–423.

Heyden, K. J., Heyden, T., 2021. Market reactions to the arrival and containment of COVID-19: An event study. Finance Research Letters 38.

Hosseini, A., Jostova, G., Philipov, A., Savickas, R., 2020. The Social Media Risk Premium. SSRN Electronic Journal .

Hou, K., Xue, C., Zhang, L., 2020. Replicating Anomalies. The Review of Financial Studies 33.

Houlihan, P., Creamer, G. G., 2017. Risk premium of social media sentiment. The Journal of Investing 26, 21–28.

Hugonnier, J., Morellec, E., 2017. Bank capital, liquid reserves, and insolvency risk. Journal of Financial Economics 125, 266–285.

Indjejikian, R., Lu, H., Yang, L., 2014. Rational Information Leakage. Management Science 60, 2762–2775.

Jame, R., Johnston, R., Markov, S., Wolfe, M. C., 2016. The Value of Crowdsourced Earnings Forecasts: THE VALUE OF CROWDSOURCED EARNINGS FORECASTS. Journal of Accounting Research 54, 1077–1110.

Jeanblanc-Picqué, M., Shiryaev, A. N., 1995. Optimization of the flow of dividends. Russian Mathematical Surveys 50, 257–277.

Jensen, T. I., Kelly, B. T., Pedersen, L. H., 2021. Is there a replication crisis in finance? NBER .

Kale, J. R., Noe, T. H., 1990. Dividends, Uncertainty, and Underwriting Costs Under Asymmetric Information. Journal of Financial Research 13, 265–277.

Kelly, B. T., Malamud, S., Zhou, K., 2022. The Virtue of Complexity in Return Prediction. NBER .

Kelly, B. T., Pedersen, L. H., Malamud, S., 2020. Prinicpal Portfolios. National Bureau of Economic Research. .

Kelly, B. T., Pruitt, S., Su, Y., 2017. Instrumented Principal Component Analysis. SSRN Electronic Journal .

Kelly, B. T., Pruitt, S., Su, Y., 2019. Characteristics are covariances: A unified model of risk and return. Journal of Financial Economics 134, 501–524.

Khadjeh Nassirtoussi, A., Aghabozorgi, S., Ying Wah, T., Ngo, D. C. L., 2014. Text mining for market prediction: A systematic review. Expert Systems with Applications 41, 7653–7670.

Kim, S.-H., Kim, D., 2014. Investor sentiment from internet message postings and the predictability of stock returns. Journal of Economic Behavior & Organization 107, 708–729.

Klimenko, N., Moreno–Bromberg, S., 2016. The shadow costs of repos and bank liability structure. Journal of Economic Dynamics and Control 65, 1–29.

Kose, J., Williams, J., 1985. Dividends, Dilution, and Taxes: A Signalling Equilibrium. The Journal of Finance 40, 1053–1070.

Koukaras, P., Nousi, C., Tjortjis, C., 2022. Stock Market Prediction Using Microblogging Sentiment Analysis and Machine Learning. Telecom 3, 358–378.

Kubler, F., Scheidegger, S., 2021. Uniformly Self-Justified Equilibria. arXiv:2112.14054 [econ] .

Kyle, A. S., 1985. Continuous Auctions and Insider Trading. Econometrica 53, 1315.

Kyle, A. S., 1989. Informed Speculation with Imperfect Competition. The Review of Economic Studies 56, 317.

Lachanski, M., Pav, S., 2017. Shy of the Character Limit: "Twitter Mood Predicts the Stock Market" Revisited 14.

Leland, H. E., 1998. Agency costs, risk management, and capital structure. The Journal of Finance 53, 1213–1243.

Leung, H., Ton, T., 2015. The impact of internet stock message boards on cross-sectional returns of small-capitalization stocks. Journal of Banking & Finance 55, 37–55.

Lewellen, J., Shanken, J., 2002. Learning, Asset-Pricing Tests, and Market Efficiency. The Journal of Finance 57, 1113–1145.

Liu, Y., 2017. Why Do Large Investors Disclose Their Information? .

Ljungqvist, A., Qian, W., 2016. How Constraining Are Limits to Arbitrage? Review of Financial Studies 29, 1975–2028.

Lucas, R. E., 1976. Econometric Policy Evaluation: A Critique. In: *The Phillips Curve and Labor Markets*, American Elsevier, New York, vol. 1 of *Carnegie-Rochester Conference Series on Public Policy*, pp. 19–46.

Lucas, R. E., 1978. Asset Prices in an Exchange Economy. Econometrica 46, 1429–1445.

Ma, H., 2021. Conditional Latent Factor Models Via Econometrics-Based Neural Networks. SSRN Electronic Journal .

Martin, I. W., Nagel, S., 2021. Market efficiency in the age of big data. Journal of Financial Economics .

Maskin, E., Tirole, J., 1992. The Principal-Agent Relationship with an Informed Principal, II: Common Values. Econometrica 60, 1–42.

McGurk, Z., Nowak, A., Hall, J. C., 2020. Stock returns and investor sentiment: Textual analysis and social media. Journal of Economics and Finance 44, 458–485.

Michaely, R., Rossi, S., Weber, M., 2021. Signaling safety. Journal of Financial Economics 139, 405–427.

Milgrom, P., Shannon, C., 1994. Monotone Comparative Statics. Econometrica 62, 157–180.

Miller, M. H., Modigliani, F., 1961. Dividend Policy, Growth, and the Valuation of Shares. The Journal of Business 34, 411–433.

Miller, M. H., Rock, K., 1985. Dividend Policy under Asymmetric Information. The Journal of Finance 40, 1031–1051.

Milne, A., Whalley, A. E., 2005. Bank Capital Regulation and Incentives for Risk-Taking. WBS Finance Group Research Paper 15.

Molavi, P., Tahbaz-Salehi, A., Vedolin, A., 2021. Model Complexity, Expectations, and Asset Prices. NBER p. 44.

Morellec, E., Schürhoff, N., 2011. Corporate investment and financing under asymmetric information. Journal of Financial Economics 99, 262–288.

Moreno–Bromberg, S., Rochet, J.-C., 2018. Continuous-Time Models in Corporate Finance, Banking, and Insurance. Princeton University Press.

Muñoz, M., 2019. Rethinking capital regulation: the case for a dividend prudential target. ECB Working Paper Series 97.

Myers, S. C., Majluf, N. S., 1984. Corporate Financing and Investment Decisions when Firms have Information that Investors do not have. Journal of Financial Economics 13, 187–221.

Nadkarni, P. M., Ohno-Machado, L., Chapman, W. W., 2011. Natural language processing: An introduction. Journal of the American Medical Informatics Association 18, 544–551.

Nardo, M., Petracco-Giudici, M., Naltsidis, M., 2016. Walking down wall street with a tablet: A survey of stock market predictions using the web. Journal of Economic Surveys 30, 356–369.

Neely, C. J., Rapach, D. E., Tu, J., Zhou, G., 2014. Forecasting the Equity Risk Premium: The Role of Technical Indicators. Management Science 60, 1772–1791.

O'Hara, M., 2003. Presidential Address: Liquidity and Price Discovery. The Journal of Finance 58, 1335–1354.

Oliveira, N., Cortez, P., Areal, N., 2013. On the Predictability of Stock Market Behavior Using StockTwits Sentiment and Posting Volume. In: Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., Weikum, G., Correia, L., Reis, L. P., Cascalho, J. (eds.), *Progress in Artificial Intelligence*, Springer Berlin Heidelberg, Berlin, Heidelberg, vol. 8154, pp. 355–365.

Oliveira, N., Cortez, P., Areal, N., 2017a. The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices. Expert Systems with Applications 73, 125–144.

Oliveira, N., Cortez, P., Areal, N., 2017b. The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices. Expert Systems with Applications 73, 125–144.

Ozsoylev, H. N., Walden, J., 2011a. Asset pricing in large information networks. Journal of Economic Theory 146, 2252–2280.

Ozsoylev, H. N., Walden, J., 2011b. Asset pricing in large information networks. Journal of Economic Theory 146, 2252–2280.

Pastor, L., Veronesi, P., 2009. Learning in Financial Markets. Annual Review of Financial Economics 1, 361–381.

Pedersen, L. H., 2021. Game On: Social Networks and Markets. SSRN Electronic Journal .

Pettenuzzo, D., Sabbatucci, R., Timmermann, A., 2020. Dividend Suspensions and Cash Flow Risk during the COVID-19 Pandemic: A Dynamic Econometric Model. Swedish House of Finance Research Paper 20-18.

Polyzos, E., Samitas, A., Kampouris, I., 2022. Quantifying Market Efficiency: Information Dissemination Through Social Media. SSRN Electronic Journal .

Radner, R., Shepp, L., 1996. Risk vs. profit potential: A model for corporate strategy. Journal of Economic Dynamics and Control 20, 1373–1393.

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., Mozetič, I., 2015. The Effects of Twitter Sentiment on Stock Price Returns. PLOS ONE 10, e0138441.

Rapach, D. E., Strauss, J. K., Zhou, G., 2010. Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. Review of Financial Studies 23, 821–862.

Refaeilzadeh, P., Tang, L., Liu, H., 2009. Cross-validation. In: *Encyclopedia of Database Systems*, vol. 5, pp. 532–538.

Renault, T., 2017. Intraday online investor sentiment and return patterns in the U.S. stock market. Journal of Banking & Finance 84, 25–40.

Reuters, 2021. Lufthansa to pay no dividend for 2019 as airline seeks bailout. May 5.

Ross, M. P., 1998. Dynamic optimal risk management and dividend policy under optimal capital structure and maturity. Institute of Business and Economic Research .

Rundo, Francesco, Trenta, Francesca, di Stallo, Agatino Luigi, Battiato, Sebastiano, 2019. Machine Learning for Quantitative Finance Applications: A Survey. Applied Sciences 9, 5574.

Sabherwal, S., Sarkar, S. K., Zhang, Y., 2011. Do Internet Stock Message Boards Influence Trading? Evidence from Heavily Discussed Stocks with No Fundamental News: DO INTERNET STOCK MESSAGE BOARDS INFLUENCE TRADING? Journal of Business Finance & Accounting 38, 1209–1237.

Schmidt, D., 2020. Stock Market Rumors and Credibility. The Review of Financial Studies 33, 3804–3853.

Shi, Y., Mitra, G., Arbex-Valle, C., Yu, X., 2017. Using Social Media and News Sentiment Data to Construct a Momentum Strategy. SSRN Electronic Journal .

Shiller, R., 1980. Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? Tech. Rep. w0456, National Bureau of Economic Research, Cambridge, MA.

Shleifer, A., Vishny, R., 2011. Fire Sales in Finance and Macroeconomics. Journal of Economic Perspectives 25, 29–48.

Siganos, A., Vagenas-Nanos, E., Verwijmeren, P., 2014. Facebook's daily sentiment and international stock markets. Journal of Economic Behavior & Organization 107, 730–743.

Sims, C. A., 2003. Implications of rational inattention. Journal of Monetary Economics 50, 665–690.

Sprenger, T. O., Sandner, P. G., Tumasjan, A., Welpe, I. M., 2014a. News or Noise? Using Twitter to Identify and Understand Company-specific News Flow: UNDERSTAND COMPANY-SPECIFIC NEWS FLOW. Journal of Business Finance & Accounting 41, 791–830.

Sprenger, T. O., Tumasjan, A., Sandner, P. G., Welpe, I. M., 2014b. Tweets and Trades: The Information Content of Stock Microblogs: Tweets and Trades. European Financial Management 20, 926–957.

Sun, A., Lachanski, M., Fabozzi, F. J., 2016. Trade the tweet: Social media text mining and sparse matrix factorization for stock market prediction. International Review of Financial Analysis 48, 272–281.

Tetlock, P. C., 2007a. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. The Journal of Finance 62, 1139–1168.

Tetlock, P. C., 2007b. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. The Journal of Finance 62, 1139–1168.

Tetlock, P. C., 2010. Does Public Financial News Resolve Asymmetric Information? Review of Financial Studies 23, 3520–3557.

The Economist, 2017. Tech firms hoard huge cash piles. June 3.

The Economist, 2021a. Archegos, a family office, brings Nomura and Credit Suisse big losses. April 3.

The Economist, 2021b. The war against money-laundering is being lost. April 17.

Tripathy, N., Wu, D., Zheng, Y., 2021. Dividends and financial health: Evidence from U.S. bank holding companies. Journal of Corporate Finance 66.

Tumarkin, R., Management, M., Whitelaw, R. F., 2001. News or Noise? Internet Message Board Activity and Stock Prices .

Van Bommel, J., 2003. Rumors. The Journal of Finance 58, 1499–1520.

Van Nieuwerburgh, S., Veldkamp, L., 2010. Information Acquisition and Under-Diversification. Review of Economic Studies 77, 779–805.

Vissers, S., 2021. Bank Regulation and Market Discipline in the Presence of Risk-Shifting Incentives. Working Paper.

Vives, X., 2008. Information and Learning in Markets: The Impact of Market Microstructure. Princeton University Press, Princeton.

Wall Street Journal, 2021. Anti-Money-Laundering Prosecution Deals Setbacks to European Banks. April 19.

Welch, I., Goyal, A., 2008. A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. Review of Financial Studies 21, 1455–1508.

Wolpert, D. H., 1996. The Lack of A Priori Distinctions Between Learning Algorithms. Neural Computation 8, 1341–1390.

Yang, S. Y., Mo, S. Y. K., Liu, A., 2015. Twitter financial community sentiment and its predictive relationship to stock market movement. Quantitative Finance 15, 1637–1656.

Zhang, D., Wu, C., Zhou, C., 2013. Optimal liquidity reserve with funding liquidity risk. Applied Economics Letters 20, 1449–1452.

Zheng, Y., 2018. Do Banks' Dividends Signal Their Financial Health? Working Paper.