Financial Frictions, Crises, Return Predictability, and Managerial Moral Hazard



Artur Doshchyn
Nuffield College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Trinity 2021

Contents

Acknowledgements **iv**

Abstract **v**

Introduction **1**

1 Long-run productivity growth, credit booms, and financial crises **4**

1.1 Introduction 4

1.2 Literature review 8

1.3 The model 10

1.3.1 The framework 10

1.3.2 The first best 13

1.4 Agency problems and financial contracts: equity vs. deposits . . . 13

1.4.1 Financial frictions 14

1.4.2 Equity financing: well-coordinated investors 15

1.4.3 Deposits: creating collective action problem among investors 16

1.5 Banks' choice of financing, equilibrium, and financial crises 18

1.5.1 Banker's problem 19

1.5.2 Equilibrium and financial crises 25

1.6 Welfare and policy implications 27

1.6.1 Ex-ante leverage restrictions 27

1.6.2 Cleaning-up ex post: government bailouts 28

1.6.3 Addressing financial frictions directly 30
1.7 Conclusion 32
Bibliography 33
Appendices 39
1.A Productivity, booms, and the 2008 Crisis 39
1.B Capital investment 41
1.C Proofs 44

2 Sinking ships: asset prices and the predictability of returns on shipping in recessions 49

2.1 Introduction 49
2.2 Dry bulk shipping: background and data 56
2.2.1 Basic facts about dry bulk shipping 56
2.2.2 Data 58
2.2.3 Time series properties of earnings, prices, and returns . . . 61
2.3 Countercyclical predictability of shipping returns 64
2.3.1 Return predictability in simple univariate regressions . . . 64
2.3.2 Regime-dependent predictability 66
2.3.3 Robustness 69
2.3.4 Risk-based explanations of return predictability 71
2.4 Liquidity constraints and industry slumps during recessions . . . 76
2.4.1 Financial frictions and return predictability in dry bulk shipping 76
2.4.2 The model 80
2.5 Conclusion 89
Bibliography 90
Appendices 94
2.A Alternative ship types 94
2.B NBER recessions 96

2.C Excluding the Financial Crisis 97
2.D Potential upward bias in returns during recessions 98
2.E Annual sampling 100
2.F Stambaugh bias and near-unit-root inference 101
2.G Measurement errors in earnings and price series 104
2.H Using unadjusted data 105

3 Executive compensation in high-growth no-dividend firms 107
3.1 Introduction 107
3.2 The model 113
3.2.1 Setup 114
3.2.2 Full information benchmark 118
3.3 Stock-based compensation and overinvestment under asymmetric information 121
3.3.1 Stock-based compensation with unobserved effort 121
3.3.2 Conceiling and overinvestment under stock based compensation and asymmetric information 123
3.3.3 Choice of effort under asymmetric information 127
3.4 Alternative compensation schemes 129
3.4.1 Flat wage 129
3.4.2 Deferred compensation: vesting 130
3.4.3 Deferred compensation: delayed payments 131
3.4.4 Option-based compensation 134
3.4.5 Cash flow-based (bonus) compensation 135
3.4.6 Mixed compensation 137
3.5 Quantitative analysis 139
3.6 Conclusion 142
Bibliography 143

Acknowledgements

I am deeply grateful to Martin Ellison and Steve Bond. Without their supervision and support throughout my DPhil, this thesis would not be possible. I would like to thank my confirmation assessors, Chris Bowdler and Michael McMahon, for invaluable feedback, comments, and suggestions. I also owe special thanks to Oren Sussman, who has been a wonderful mentor to me.

I am thankful for having been a part of the terrific Macroeconomics group at Oxford. I have benefited a lot from interacting with the faculty members and my fellow DPhil students. Among many others, I thank Alex Haas, Alistair Macaulay, Genevieve Nelson, Petr Sedláček, and Francesco Zanetti for helpful discussions and advice. I am grateful to all the seminar speakers I had a pleasure to meet and discuss my work. I also thank seminar participants at Saïd Business School and Department of Economics for invaluable feedback during and following my presentations.

I am very grateful to Nuffield College and the Department of Economics for their financial support, which enabled be to pursue my DPhil degree.

On a personal note, I would like to thank my parents, family members, and friends for their support, encouragement, and all the wonderful moments of joy. My deepest gratitude goes to Sarah. I cannot imagine my work and life without your love, patience, everlasting smile and positivity. I dedicate this thesis to you.

Abstract

Chapter 1. *While financial crises tend to be preceded by credit booms, most credit booms do not end up in crises. Crises typically occur when there is also a persistent slowdown in productivity growth. I develop a model in which risk of crisis emerges endogenously during boom because of increased fragility of the banking sector. Banks raise financing from households to invest in long-term projects, but their ability to do so is limited by moral hazard. Demandable deposits create discipline by exposing misbehaving banks to runs, and thus help them increase external financing. Normally, banks finance themselves with a mix of equity and deposits that maximizes discipline, but ensures that they always remain solvent. When growth prospects become sufficiently strong, however, worsening moral hazard induces banks to rely exclusively on deposit financing, which enables a boom in credit, asset prices, and investment. If the anticipated growth fails to materialise, though, the excessive deposit financing leads to a systemic banking crisis. I also study policy implications of the model.*

Chapter 2. *I document that future returns on real assets in dry bulk shipping are predictable and negatively related to ship earnings, prices, and industry investment during global economic recessions, but are not forecastable in expansions. The predictability in recessions is statistically significant and economically large. The empirical evidence points against standard risk-based explanations of return predictability, and is also inconsistent with previously suggested behavioural explanations. I propose and formalise a theory of countercyclical return predictability arising due to liquidity constraints faced by the industry participants in downturns. When cash flows dry up in recessions, financial troubles force firms to sell assets to their liquidity constrained peers, resulting in plunges in ship prices and high expected returns to the buyers. Liquidity shortages also lead to slumps in industry investment.*

Chapter 3. *I study managerial compensation in financially constrained firms with high growth opportunities. Such firms can often profitably reinvest their operating incomes and pay no dividends. Absence of reliable performance indicators together with asymmetric information create acute agency problems. Stock-based compensation schemes, notoriously popular during the tech bubble of the 1990s, provide the necessary incentives for the manager to exert costly effort to pursue further growth. However, in line with the previous literature, I find that they may also induce the manager to conceal bad news about the future prospects of the firm, in order to maintain a high stock market valuation. This leads to a substantial overinvestment in the firm's assets, and expands a bubble-like surge in the firm's stock price, followed by an inevitable crash. I find that a stock-based compensation contract that delays payments by several years effectively deters the manager from concealing, while still providing incentives to work hard.*

Introduction

This thesis consists of three chapters in macroeconomics and finance. The common thread is an exploration into how models of asymmetric information and agency problems in corporate finance can help us understand macro-finance phenomena like boom-bust cycles and financial crises.

The first chapter, "Long-run productivity growth, credit booms, and financial crises" addresses the question of why booms make the economy susceptible to the risk of financial crises. Empirically, while financial crises tend to be preceded by credit booms and accelerated growth of productivity, most credit booms do not end up in crises. Crises typically occur when there is also a persistent slowdown in productivity growth. Many existing theories of financial crises place the origins of the financial boom-bust cycle within the financial sector, and cannot account for this evidence.

I build a theory of credit booms and banking crises driven by productivity that incorporates Diamond and Rajan's (2000, 2001) work on bank capital structure into a simple macroeconomic model with aggregate uncertainty. Booms may occasionally be followed by crises because optimal structure of banks' liabilities becomes increasingly fragile during periods of strong growth. Specifically, because banks invest in projects with uncertain long-run returns, in the first best they would raise funding using long-term state-contingent instruments, like equity. However, the ability of banks to issue equity is limited by moral hazard, as they can make it difficult for investors to obtain repayment in the future. Like in Dia-

mond and Rajan, the use of demandable deposits that subject misbehaving banks to runs creates discipline and allows banks to increase financing. Normally, banks fund themselves with a mix of equity and deposits that maximizes discipline, but ensures that they always remain solvent. When growth prospects become sufficiently strong, however, worsening moral hazard induces banks to rely exclusively on deposits, facilitating a boom in credit, asset prices, and investment. If the anticipated growth fails to materialise, though, the excessive deposit financing leads to a banking crisis. I consider implications of the model for welfare and government policy. For example, while widely discussed leverage restrictions may indeed eliminate crises, they hurt welfare by stifling asset prices and investment at the time when it is most valuable.

In the second paper, I focus on boom-bust cycles in investment and asset prices in competitive, capital-intensive industries. Specifically, I investigate the relation between investment, asset prices, and future returns in dry bulk shipping. It is an ideal setting with a large number of relatively small competitive firms, homogenous assets, substantial time to build new ships, and inelastic demand for shipping services that strongly depends on macroeconomic conditions. I present novel empirical evidence that future excess returns on owning ships are strongly predictable and negatively related to ship earnings, prices, and investments in new ships during global recessions. There appears to be no such predictability during expansions. The empirical evidence points against standard risk-based explanations of return predictability, and concentration of predictability during recessions is inconsistent with previously suggested behavioural explanations.

I propose a plausible theory according to which the return predictability in recessions is driven by forced ship sales by distressed firms when potential buyers are also financially constrained. This, in the spirit of Shleifer and Vishny (1992), results in the cash-in-the-market pricing and assets selling for below their fundamental value. It is only natural to expect financial constraints to bind in

industries like shipping, which are dominated by small players with specialist knowledge, often informal corporate structures, and traditional reliance on bank financing. I build a formal model of a competitive industry in which, due to limited commitment, firms are only able to borrow against the resale value of their capital. In a boom, when firms' cash flows are high, high liquidity and competition ensure that assets are traded at the fundamental value: buyers earn the required rate of return, and there is no predictability. In a recession, however, even firms with good idiosyncratic shocks have limited spare liquidity, and are unable to borrow. As the amount of capital that firms with bad shocks have to sell to repay their debts increases, the asset price has to fall below its fundamental value to equilibrate capital supply with the constrained capital demand. The worse is the aggregate shock, the more constrained the buyers are, and so the deeper asset price falls, implying high excess returns to the buyers. The theory thus provides a novel insight that return predictability on real assets in competitive industries is a manifestation of financial constraints these industries face during downturns.

The third chapter is concerned with managerial compensation in innovative, financially constrained firms with high growth opportunities. In a formal model of such a firm, I show that stock-based compensation can incentivize the manager to work hard to pursue further growth, but also to conceal from investors bad news about the firm's prospects in order to maintain high stock price growth. This leads to substantial overinvestment and bubble-like boom and bust in the stock price in equilibrium. This paper thus extends results of Benmelech et al. (2010) to high-growth, no-dividend firms. The results shed light on the tech boom of the 1990s, when stock-based compensation schemes increased dramatically in popularity, and telecom firms have arguably severely overinvested in the internet infrastructure, prior to the burst of the bubble. I also consider alternative compensation schemes that may perform better, notably contracts that delay payments from stock sales.

Chapter 1

Long-run productivity growth, credit booms, and financial crises

1.1 Introduction

It is well-documented that major financial crises are typically preceded by booms in credit and asset prices[1] and are followed by protracted periods of slow growth.[2] Existing literature often studies booms as the result of financial liberalisation, excessive risk taking, or over-optimistic expectations,[3] while slow recoveries are attributed to ensuing credit crunches and deleveraging during crises.[4] These mechanisms are undoubtedly important, but there are several empirical observations linking financial cycles to the real economy which they cannot account for. Firstly, starts of credit booms are often associated with accelerated trend growth rate of real productivity, and, importantly, most credit

[1] See Reinhart and Reinhart (2010); Mendoza and Terrones (2008, 2014); Schularick and Taylor (2012); Reinhart and Reinhart (2010); Krishnamurthy and Muir (2017).

[2] See Cerra and Saxena (2008); Reinhart and Reinhart (2010); Reinhart and Rogoff (2009, 2014); Krishnamurthy and Muir (2017); Queralto (2019).

[3] For example, different strands of literature emphasized either moral hazard due to implicit government guarantees (Farhi and Tirole, 2012; Chari and Kehoe, 2016), overoptimism (Bordalo et al., 2018), or pecuniary externalities operating through borrowing constraints (see e.g. Lorenzoni (2008)).

[4] See, for example, Queralto (2019); Anzoategui et al. (2019); Bianchi et al. (2019); Guerron-quintana and Jinnai (2014); Garcia-Macia (2017); Ennis and Keister (2003).

booms do not end up in crises. When they do, however, financial crises tend to follow, rather than lead, persistent slowdowns in productivity growth (Gorton and Ordoñez, 2019; Dell'Ariccia et al., 2012; Paul, 2020).[5] This suggests that slow economic growth cannot be attributed solely to crises, and, in fact, it is deteriorating economic prospects that could be among the factors that trigger crises in the first place. Strikingly, Cao and L'Huillier (2018) document that the three largest economic disasters in the developed world, – The Great Depression, The Japanese Slump, and The Great Recession, – were all preceded by periods of major technological innovations and rapid productivity growth which ran out of steam before the crises. The evidence thus suggests that the ingredients for a typical financial crisis include deteriorating macroeconomic fundamentals, but also a preceding period of economic prosperity. This leads to an important question which I study theoretically in this paper: why do booms make the economy susceptible to the risk of financial crises?

To address it, I develop a model in which the key mechanism is a build-up of financial fragility in the banking sector during periods of booming productivity growth, when expected returns on banks' long-term investments are high. Building on the seminal work of Diamond and Rajan (2000, 2001b), part of banks' external financing comes in the form of short-term demandable deposits. Deposits discipline banks against trying to squeeze their external investors by exposing misbehaving banks to runs, and thus allow them to increase financing. Normally, banks effectively mix deposits and equity financing to maximize discipline, yet to avoid the risks of destructive runs. Crucially, however, bankers' moral hazard becomes worse when the long-run productivity growth prospects are strong. This prompts them to rely heavily on deposits, since their ability to raise financing through other means deteriorates. Consequently, it is deposit financing that en-

[5]For the Great Recession, Fernald (2014); Cette et al. (2016); Cao and L'Huillier (2018) provide evidence that productivity has slowed *well before* the onset of the Financial Crisis. An updated version of Kahn and Rich (2007) gives a similar conclusion – see 1.A for details.

ables a boom in credit, asset prices, and investment. If future aggregate growth prospects deteriorate prematurely, however, the value of banks' assets falls, rendering them insolvent, and instigating systemic bank runs. A long period of slow subsequent growth is then not just the result of the crisis, but also its cause.

More specifically, banks in the model have access to productive long-term projects that require capital investments. The long-run payoff of banks' projects is uncertain, and can be either high or low, depending on the state of aggregate productivity. Capital is supplied by households, who cannot use it as productively as banks. Banks purchase capital on a competitive market, using own funds and financing they raise from households.

While banks would prefer to finance their uncertain long-term investments with outside equity or similar flexible state-contingent instruments, their ability to issue equity is constrained by moral hazard and limited commitment. Specifically, banks can take actions that irreversibly reduce the ability of external investors to collect repayments from them in the future.[6] But since equity investors cannot commit to act against their best collective interest, they cannot credibly threaten to punish and liquidate opportunistic banks that try to squeeze them down to the liquidation value of assets.

Banks can, however, increase their financing by issuing deposits that are demandable at any time and feature a sequential service (aka first-come first-served) property, as in Diamond and Rajan (2001b). This creates a built-in collective action problem among depositors: when faced with a prospect of lower repayments than promised, each individual depositor finds it optimal to run and demand full repayment, hoping to be among the first in line. Deposits thus act as a disciplining device, since depositors will run and liquidate bank's assets whenever the bank is unable to guarantee them full repayments in the future, even if it is not in

[6]This is meant to capture that banks can create obscure and opaque corporate structure, reduce transparency about nature of their investments, maintain weak accounting standards etc.

their best collective interest.[7] Anticipating this, the bank has strong incentives to maintain high pledgeability of future returns.

In normal times, when low productivity growth is a fairly likely outcome, banks are careful: they limit the level of deposits to what they can repay in both high and low productivity states, and issue an incentive-compatible amount of risky outside equity against the high state. When the probability of the high state increases, however, the ability of safe banks to issue outside equity deteriorates, because banks have a greater incentive to squeeze the equity investors' share of the high state returns. Therefore, when the high productivity state becomes sufficiently likely, banks find it optimal to follow a risky strategy and finance themselves exclusively with deposits, which they would be unable to repay if the low aggregate state materialises. This allows banks to increase the total amount of financing in a boom, and leads to higher equilibrium asset prices and investment. On the dark side, however, high deposit financing introduces financial fragility, and there is a systemic banking crisis in the low productivity state.

I consider implications of the model for welfare and government policy. In contrast to much of the post-Great Recession focus on macroprudential regulation, the model presented here cautions against relying solely on ex-ante preventive policies like leverage restrictions. While a strict macroprudential limit on banks' depositary liabilities can eliminate crises, it also limits investment at times of high productivity growth, precisely when it is the most desirable, and thus reduces welfare. Ex-post government bailouts, on the other hand, can save the banking sector from a collapse and improve welfare *in a crisis*. However, anticipation of unrestrained bailouts can completely destroy market discipline of deposits, leading to an inefficient level of investment and hurting welfare. Ex-post interventions are thus only desirable if the government can commit to help banks in case of a

[7] 'Deposits' in this paper is a metaphor for a wide class of short-term demandable instruments that are capable of creating a collective action problem between the claimholders, and not only classic bank deposits. For example, Gorton and Metrick (2012) and Martin et al. (2014) argue that it was the run on the repo market that was at the heart of the 2008 Crisis.

systemic banking crisis driven by worsening fundamentals, but not to interfere when the market carries out discipline against opportunistic banks.

Of course, even well-targeted ex-post bailouts deal with symptoms rather than the underlying problem: inability of banks to credibly pledge future returns to investors. This motivates the question: should the banks be monitored during booms to mitigate agency problems and enforce repayments to outside investors? Under simple assumptions about costs and benefits of supervision, I find that the regulator will indeed sometimes supervise banks, increasing their financing and eliminating crises, thus improving welfare. However, the regulator may optimally choose to withdraw and let banks rely on discipline provided by deposits when the confidence in high productivity growth is particularly strong. This potentially sheds light on why, for example, regulation and supervision of the financial sector in the US appear particularly lax in the years preceding the 2008 Financial Crisis.

1.2 Literature review

This paper builds on Diamond and Rajan (2001b) who, in a similar vein to the seminal paper of Calomiris and Kahn (1991), strive to explain banks' reliance on short-term demandable deposits that expose them to runs. In these papers, demandable deposits with the first-come first-served feature align incentives and preclude banks from acting against depositors' interests. Similarly to the present work, Diamond and Rajan (2000) also analyse optimal bank capital structure under uncertainty. The contribution of my paper is incorporation of this mechanism in the context of a simple macroeconomic model with aggregate uncertainty. I then demonstrate how systemic fragility of the banking sector increases in order to facilitate financing of investments during periods of strong economic growth, thus explaining why financial crises tend to follow credit booms.

The underlying idea that productivity underpins credit booms and financial

crises can be traced back to at least Fisher (1933). Closely related recent studies include Gorton and Ordoñez (2019), L'Huillier et al. (2020), Boissay et al. (2016). While in these models crises are certain and perfectly predictable during booms, in my model crises are low probability events, consistent with evidence on agents' expectations.[8]

My paper also contributes to the literature on bank runs. Following Thakor (2018), the literature can be loosely divided into "illiquidity" and "insolvency" camps, although the line between the two issues is thin in practice. According to the first camp, rooted in the seminal work of Diamond and Dybvig (1983), runs are panic-based phenomena, and are essentially coordination failures in a world with multiple equilibria. The present paper belongs to the second camp, which seeks to explain bank runs as driven by the insolvency risk of banks due to weak economic fundamentals, consistent with a number of empirical findings.[9] One prominent example of this literature is Allen and Gale (1998), who build a model in which partial bank runs are an efficient way to improve aggregate risk sharing between early and late consumers. In a series of recent notable papers by Gertler and Kiyotaki (2015); Gertler et al. (2019a); Martin et al. (2014), while bank runs are still mostly panic-based, run equilibria are only possible when the fundamentals are sufficiently weak. In contrast to the present paper, however, the above papers do not account for why banking crises tend to follow credit booms, or why banks finance themselves with deposits or repos.

Gertler et al. (2019b) develop a model in which banking panics are preceded by credit booms, but not all credit booms are "bad", i.e. result in crises. While

[8]See also Cao and L'Huillier (2018), who similarly relate major economic booms and slumps to low frequency variations in the trend productivity growth. They abstract from financial crises altogether, however, and instead focus on consumption behaviour over the medium run in a permanent income framework with learning about permanent productivity shocks.

[9]For example, Calomiris and Gorton (1991) document that banking crises are not random events, but are accounted for by bad macroeconomic news and high insolvency risk of banks; Calomiris and Mason (2003) show that most bank failures during the Great Depression can be well explained by fundamentals, leaving only limited role to panics; Thakor (2018) surveys empirical evidence and concludes that the 2008 Financial Crisis was likely an insolvency risk crisis, rather than an illiquidity crisis.

this is similar to my mechanism, I differ from them in that runnable deposits constitute part of banks' optimal capital structure in my model, and deposit financing endogenously rises in booms. The advantage of this is that banks cannot eliminate the problem by making suspension of convertibility (temporary ban on withdrawals) in case of a run part of the deposit contract.

Starting with Aguiar and Gopinath (2007), there has been a growing literature on small open emerging economies that associates Sudden Stops with changes in trend productivity growth in these countries due to frequent policy regime changes, e.g. Flemming et al. (2019), Seoane and Yurdagul (2019), and Akıncı and Chahrour (2018). These models feature pecuniary externalities due to occasionally binding collateral constraints, and show that trend, rather than transitory shocks, produce realistic Sudden Stops dynamics, and can account for both preceding credit booms and slow recoveries. In this paper I leave out the pecuniary externalities and financial accelerator effect that are central to these papers, and specifically focus on bank runs and insolvency crises.[10]

1.3 The model

1.3.1 The framework

There are three periods: $t \in \{0, 1, 2\}$. I shall also refer to period 2 as the 'long run'. There are two types of agents in the economy: households and bankers. There are two types of goods: consumption good, which is the numeraire, and physical capital. Capital is used to produce consumption good, as described below, but cannot be consumed itself. Lastly, all agents are risk-neutral, and, for clarity, I

[10]There has been a large post-crisis literature on the role of BGG/KM-like financial accelerator effect during the financial crisis. Gertler and Karadi (2011) build a macro model where the banking sector is subject to balance sheet constraints and thus the classic financial accelerator. Brunnermeier and Sannikov (2014); Mendoza (2010); He and Krishnamurthy (2019) emphasize asymmetries and nonlinearities of models with financial frictions away from steady state. Lorenzoni (2008); Stein (2012) theoretically study the ex ante overborrowing and inefficiency associated with pecuniary externalities interacting with borrowing constraints.

assume that there is no discounting and everyone consumes in period 2.

In period 0, banks can invest in projects that mature and pay off in the long run, i.e. period 2. Banks are scalable, and perfectly diversify away any idiosyncratic projects' risks. Each project requires an investment of one unit of physical capital, which bankers need to purchase on a competitive market. For simplicity, I assume that all capital is owned by households in period 0, and the aggregate supply of capital is fixed and normalised to 1. Later, in Appendix 1.B, I relax the assumption of fixed capital stock and allow for variable aggregate investment.

There is undiversifiable aggregate uncertainty about the long-run productivity that pertains to all banks' projects. Specifically, there are two possible aggregate productivity states $s \in \{H, L\}$, with respective ex-ante probabilities π and $1 - \pi$. In the high productivity state H, the average return per project is R^H in period 2, whereas in the low productivity state L it is R^L, with $R^L < R^H$. The state is realised in period 1 and is observed by all agents. However, it takes time until period 2 to formally verify the aggregate state, and thus contracts cannot be state-contingent in period 1.

Projects are bank-specific, and other banks or households cannot run them after the investment is made in period 0.[11] This reflects the idea that banks are relationship lenders who possess superior information about their borrowers, although in this model I assume for simplicity that banks manage their projects directly.[12]

Banks can, however, liquidate their projects for Λ per project in period 1, before they mature. Specifically, if liquidated, a project's capital is extracted in full and sold on a competitive and liquid secondary capital market back to the

[11] This assumption of extreme project illiquidity ensures that no banks bet on buying projects cheaply off distressed counterparties in a crisis, and considerably simplifies the analysis.

[12] The setup can also be interpreted as one in which there are entrepreneurs who run the projects, but (a) entrepreneurs are penniless and need to raise external financing; (b) households are unable to finance the entrepreneurs directly due to (unmodelled) agency problems, whereas banks have the expertise to resolve them; (c) entrepreneurs compete for bank funding, so all project returns go to banks.

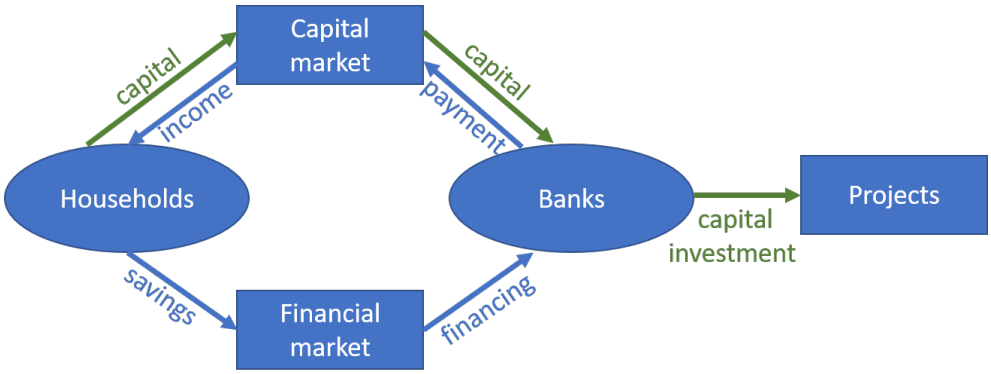


Figure 1.1: Flow of resources in period 0

households. Households employ this capital in a traditional sector that yields Λ per unit of investment. Liquidation is always less profitable than leaving capital with the bank and letting the project mature: $\Lambda < R^L$.[13]

Each bank starts with a positive endowment of consumption good, which we will call net worth.[14] Banks finance their capital purchases with their net worth and financing they obtain from households. In what follows, I assume that aggregate net worth of the banking sector N is sufficiently scarce such that banks indeed raise additional financing from households. Households, on the other hand, are not constrained in their ability to provide financing to banks in period 0, or purchase banking sector's capital in case of liquidation in period 1.[15] Financial market is competitive, and banks' investors break-even in expectation. Figure 1.1 illustrates the flow of resources in period 0.

[13]The assumption that the liquidation value does not depend on the state is not critical for the results, but simplifies the exposition.

[14]The results would be similar if, instead, we assumed that banks start with a stock of physical capital and liabilities to households.

[15]This, of course, will be always true if households have sufficiently large endowments of consumption good. This would also be the case, however, even if households start with zero endowments if we assume that all transactions happen instantaneously. In that case, household sector effectively ends up lending capital to banks in exchange of promises of future payments (or, conversely, accepting physical capital in lieu of repayments when banks liquidate projects).

1.3.2 The first best

It is instructive to describe the first-best outcome in this model. Absent agency problems we shall introduce below, it is never desirable to liquidate banks' projects in period 1. Any financing banks raise from households, therefore, takes the form of long-term, state-contingent liabilities that are repaid in period 2 when projects mature.

Consider a bank with net worth n. If the bank raises additional funds b, it can start $\frac{n+b}{Q}$ projects. Since bank's investors break-even in expectation, the expected value of future repayments is simply b, and the bank's expected return is:

$$\frac{n+b}{Q}\mathbb{E}[R^s] - b,$$

where $\mathbb{E}[R^s] = \pi R^H + (1-\pi)R^L$ is the expected return on banks' projects. Clearly, as long as $Q \neq \mathbb{E}[R^s]$, banks are either willing to expand borrowing to infinity, or are not willing to invest into projects at all, neither of which can be the equilibrium. The first-best equilibrium thus requires $Q^* = \mathbb{E}[R^s]$. There is no capital invested in the traditional sector. Intermediation is costless and efficient, as banks earn zero expected profits.

1.4 Agency problems and financial contracts: equity vs. deposits

In this section we outline the agency problems that banks suffer from, and explore the instruments that they can use to raise financing from households. As we shall see, banks financed by well-coordinated equity investors are never inefficiently liquidated. However, ability of banks to finance their capital investments using only outside equity is limited. This is because, after the investments are made,

bankers can reduce the ability of outsiders to collect repayments out of future returns, while equity investors cannot commit to punish misbehaving banks. As in Diamond and Rajan (2001b), banks can overcome this by issuing demand deposits, which effectively commit depositors to run and liquidate banks that cannot repay them in full, even when liquidation is not in their best collective interest. While deposits bring about the needed discipline, they can also lead to socially undesirable bank failures.

1.4.1 Financial frictions

The first best described in the previous section may be unattainable in equilibrium due to the following financial frictions. Banks suffer from limited pledgeability of future returns: at most fraction $\theta \in (0,1)$ of banks' returns is collectable by external claimholders in period 2. What is worse, after the investments in period 0 are made, banks can also take actions that further irreversibly reduce the ability of investors to collect from them in the future. Formally, a bank can set $\tilde{\theta} \in [0,\theta]$, such that at most fraction $\tilde{\theta}$ of its projects' output can be collected by external claimants in period 2. This assumption is meant to capture the ability of banks to create obscure and opaque corporate structures, limit transparency about the nature of their investments, stick with weak accounting standards etc. Bank's choice of $\tilde{\theta}$ is observable by its investors in period 1, but not verifiable and cannot be included in a contract. Bankers cannot commit to maintain high $\tilde{\theta}$ if it is not in their best interest. In addition, there is also limited commitment on behalf of households, precluding banks from being able to purchase insurance against the aggregate state.

Financial contracts can give investors control rights to liquidate banks' assets. I assume it is always possible to force banks to liquidate projects and pay proceeds to the investors in period 1, irrespective of chosen $\tilde{\theta}$. Without the liquidation rights, investors would not be able to get any repayments at all, since banks would

always set $\tilde{\theta} = 0$; banks thus would not obtain any financing from the households in the first place. To keep the model interesting, I assume $\theta R^L > \Lambda$, so even in the low state investors may receive more than the liquidation value of assets if bank maintains high $\tilde{\theta}$.

1.4.2 Equity financing: well-coordinated investors

One financing option banks have is to issue long-term state-contingent liabilities, which, for concreteness, we will call (outside) equity. The general equity contract we consider specifies control rights to liquidate assets in period 1, as well as promised state-contingent payouts in period 2. If investors are not repaid in full in period 2, they are entitled to everything that can be recovered from the bank.[16] Crucially, equity investors are assumed to be well-coordinated and to always act as one to maximize their collective payoff. Whenever there is a conflict of interest between individual investors, they are always able to efficiently negotiate among themselves. The distribution of claims among equity investors is therefore irrelevant, and we can treat them as one representative financier.

Equity investors cannot commit ex ante to act against their best interest ex post. They will thus never liquidate assets of a bank if they stand to gain more from letting it continue running until period 2, even if the bank is not able to repay the promised amount in full. This lenience is a double-edged sword. While it assures that equity-financed banks will never get liquidated while they generate value for their investors, it also allows banks to squeeze their investors by reducing their ability to collect from them in the future. As a consequence, the ability of banks to finance themselves with equity is limited:

Lemma 1. *The amount of financing that a bank can raise by only issuing equity is at most a fraction* $\frac{\Lambda}{R^L}$ *of the expected returns on its assets.*

[16] While this contract looks more like state-contingent long-term debt rather than equity, the results are the same if instead we consider a more traditional but restrictive equity contact that entitles investors to a fixed share of bank's returns in period 2.

This result follows because investors anticipate that the bank can always set $\tilde{\theta} = \Lambda/R^L < \theta$ after selling equity, without risking liquidation. As investors can still collect Λ per project in period 2 even if the low state materialises, they will never liquidate.[17]

1.4.3 Deposits: creating collective action problem among investors

The central problem with equity is that coordinated investors cannot make credible threats to punish and liquidate banks that try to squeeze them, since this may not be in their best interest ex post. In the seminal paper, Diamond and Rajan (2001b) suggest that the use of 'harder' claims like demandable deposit contracts disciplines the banks against trying to be opportunistic. Their idea is that deposit contracts that promise repayments on demand (forcing banks to liquidate assets in period 1, if necessary), and feature sequential service (aka first come, first served property), can break the coordination between investors and make it impossible for the banks to reduce their liabilities on deposits. This is because there is a built-in collective action problem among depositors: whenever a bank is unable to guarantee full repayment to all of them, each individual depositor has an incentive to run on the bank hoping to be among the first in line and be made whole, rather than accept lower repayment. An attempt to squeeze depositors will thus trigger a run in which all depositors try to withdraw, even when it is in their best collective interest to agree to lower payments.[18] To put differently, deposit contracts commit bank creditors to punish and liquidate opportunistic banks ex ante, even if they would not choose to do so ex post. Anticipating this, banks that are funded by deposits to a sufficiently high extent will have stronger incentives to maintain

[17] Note that the bank may have incentives to set $\tilde{\theta}$ even lower, thus risking liquidation in the low state and further limiting financing available ex ante.

[18] For the detailed formal development of this argument, see section III.D, and also the proof of Lemma 2 in the Appendix of Diamond and Rajan (2001b).

high pledgeability of their future returns, allowing them to borrow more.

I assume that deposit contracts with sequential service constraint are available to agents in the economy, and it is the only mechanism capable of creating a collective action problem among banks' investors. Because the state cannot be verified until period 2, however, face value of deposits in period 1 cannot be made state-contingent. In fact, I assume without loss of generality that banks issue only simple deposits that have fixed face value across both periods and states; banks can then issue equity against the states in which they are able to repay more than the face value of deposits.[19] Lastly, by virtue of their instant demandability, deposits are always senior to equity: presuming that depositors can run and liquidate assets quicker than equity investors in period 1, equity always gets only the residual value, if any, after all depositors are served.[20]

Crucially, deposit funding also has a dark side: a bank may be inefficiently run and liquidated even when it does not try to extract rents from its creditors, if the pledgeable value of its assets falls below the value of its depositary liabilities due to unfavourable economic conditions. In that case it would be desirable for depositors to agree to lower repayments, but deposit contracts are explicitly designed to prevent that from happening. I assume, however, that banks cannot fall prey to classic panic bank runs as in Diamond and Dybvig (1983), and fundamentally solvent banks are always able to obtain financing in period 1.

Figure 2.6 summarizes the timing of events. Let us now turn to how banks combine equity and deposit financing in equilibrium.

[19] Firstly, it would be suboptimal for a bank to issue deposits that promise lower payments in period 2 than in period 1 in some states, since depositors would then always run in period 1 in those states. Secondly, while it is in principle possible to issue state-contingent deposits that pay more in period 2 than period 1 in some states, it does not make a difference to total financing that a bank is able to raise. Since a prospect of reduction in future repayments would not trigger a run as long as depositors get at least as much as the current face value of their deposits, they cannot commit not to negotiate period 2 repayments that are in excess of the face value of deposits in period 1. But then they are not different from equity investors with respect to those excess repayments.

[20] This is because in period 1 equity investors will always agree to a necessary reduction in their period 2 claims to forestall a run by depositors.

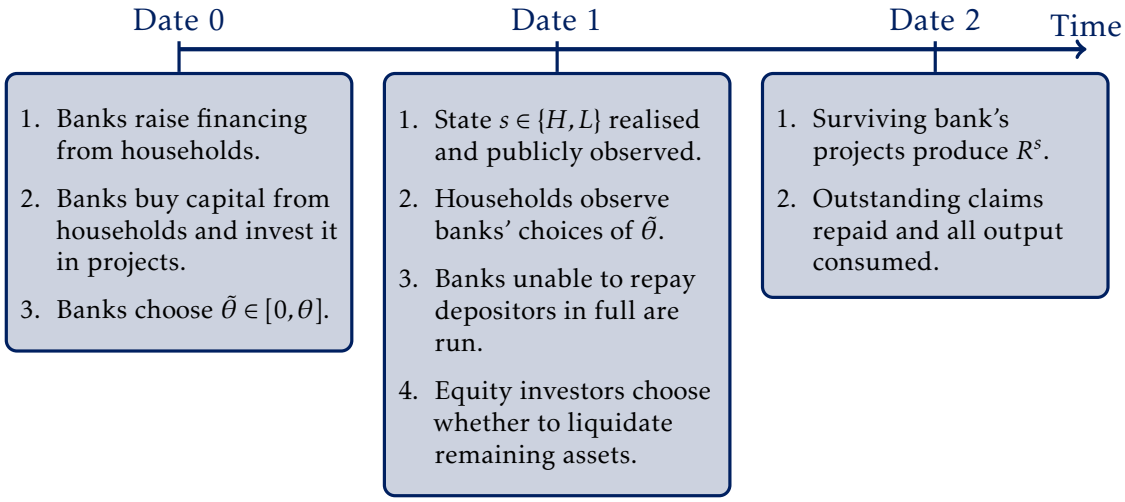


Figure 1.2: Timeline

1.5 Banks' choice of financing, equilibrium, and financial crises

In this section we first consider an individual profit maximizing bank that takes market price of capital as given. I show that when the probability of the high productivity state is relatively low, the bank is safe. It can raise the maximum feasible amount of external financing and maximize its payoff by relying on a safe mix of outside equity and deposits that it can always repay. However, when the high productivity state becomes more likely, the ability of a bank with safe level of deposits to issue outside equity declines due to moral hazard. At some point the bank finds it optimal to instead leverage itself up with deposits beyond what it can repay in the low state, i.e. become risky. This improves discipline and allows the bank to pledge more high-state returns to external investors, but comes at the cost of making a run inevitable in the low state. Second, I show that the model has a unique equilibrium. Similarly to the case of individual bank, when probability of the high state is sufficiently high, and the aggregate net worth of the banking sector is scarce, all banks choose risky strategy in equilibrium. This facilitates boom in asset prices and aggregate investment, but results in a systemic financial

crisis if the low state materializes.

1.5.1 Banker's problem

Throughout this subsection we suppose that $Q < \mathbb{E}[R^s] \equiv \pi R^H + (1-\pi)R^L$, so banks are financially constrained, earn positive expected profits, and want to expand. This will indeed be the case in equilibrium when the aggregate net worth of the banking sector is sufficiently scarce and the probability of the high state π is sufficiently high. The less interesting case $Q = \mathbb{E}[R^s]$ arises only when banks' financing constraints do not bind in equilibrium and the first best outcome is achieved.

Let b be the total amount of financing that a bank with net worth n raises, and so $m \equiv \frac{n+b}{Q}$ be the number of projects the bank starts. Let d and e^s be the face values of the bank's deposits and equity in state s, respectively. Without loss of generality, we can assume that e^s is an incentive-compatible amount that the bank actually pays out to equity investors in state s. In period 1 and state s, investors would liquidate the bank if $m\tilde{\theta}R^s < \max\{d, m\Lambda\}$, i.e. if either the pledgeable value of the bank's assets is not enough to repay its depositors in full, or is below the liquidation value of the assets.[21]

Depending on its capital structure, the bank can be either *safe* or *risky*. A bank is safe if it is solvent and avoids liquidation of its assets in both high and low aggregate productivity states. A risky bank, on the other hand, gets liquidated by its investors if the low state materialises. Intuitively, the bank's riskiness depends on how much it relies on deposits as the disciplining device. The bank will choose to be safe or risky depending on which strategy yields higher expected profit. We can then look at these two strategies separately, and compare the resulting payoffs.

[21] Of course, it always makes sense for a financially constrained bank to give investors all control rights in period 1 in this model.

Safe strategy

Taking as given that it is optimal for the bank to be safe, the bank's maximization problem can be written as:

$$\max \quad v = \pi\Big[mR^H - d - e^H\Big] + (1-\pi)\Big[mR^L - d - e^L\Big] \tag{1.1}$$

$$\text{s.t.} \quad m = \frac{n+b}{Q} \tag{1.2}$$

$$b = d + \pi e^H + (1-\pi)e^L \tag{1.3}$$

$$d + e^s \leq m\theta R^s,\ s \in \{L, H\} \tag{1.4}$$

$$d + e^L \leq \max\{d, m\Lambda\} \tag{1.5}$$

$$d + e^H \leq \frac{R^H}{R^L}\max\{d, m\Lambda\} \tag{1.6}$$

$$v \geq \pi\Big(mR^H - \max\{d, m\Lambda\}\Big) \tag{1.7}$$

Constraint (1.3) is the break-even condition for the bank's investors; (1.4) restricts liabilities of the safe bank to what it can feasibly pay out in each state; (1.5) and (1.6) place the limits on the maximum repayments in the low and high states respectively that the bank cannot reduce simply by lowering $\tilde{\theta}$ to $\frac{\max\{d,m\Lambda\}}{mR^L}$, and still avoiding liquidation in both states; lastly (1.7) is the incentive compatibility constraint that ensures that the safe bank does not benefit in expectation from squeezing equity investors out completely in the high state by setting $\tilde{\theta} = \frac{\max\{d,m\Lambda\}}{mR^H}$, and accepting the risk of failure in the low state.

Since $Q < \pi R^H + (1-\pi)R^L$, the bank will strive to raise as much external financing as possible, i.e. V is maximized when b is maximized. A quick analysis of the constraints reveals that this is achieved when $e^L = 0$ and $d = m\theta R^L$, i.e. bank operates with the highest possible safe level of deposits, and issues equity only against the high state. This is intuitive, because deposits provide discipline, and therefore relax constraints (1.5)-(1.7). The solution to the safe bank's profit maximization problem then depends on which one of the constraints (1.6) and

(1.7) binds, which in turn depends on the probability of the high state π:

Lemma 2. *A safe bank always operates with $e^L = 0$ and $d = m\theta R^L$, i.e. the largest safe level of deposit liabilities, in order to maximise discipline. Let*

$$\dot{\pi} \equiv \frac{(1-\theta)R^L}{\theta(R^H - R^L) + (1-\theta)R^L} \in (0,1). \tag{1.8}$$

- *When $\pi \leq \dot{\pi}$, constraint* (1.6) *binds, constraint* (1.7) *is slack, and the bank is able to pledge the maximum fraction θ of expected future returns to external investors. It raises $b_s/m_s = \theta\mathbb{E}[R^s]$ of external funds per project, supervises $m_s = \frac{n}{Q-\theta\mathbb{E}[R^s]}$ projects, and its expected payoff is $v_s = m_s(1-\theta)\left(\pi R^H + (1-\pi)R^L\right)$.*

- *When $\pi > \dot{\pi}$, constraint* (1.7) *binds, and the bank can only raise $b_s/m_s = \pi\theta R^L + (1-\pi)R^L$ of external financing per project, while remaining safe. The bank runs*

$$m_s = \frac{n}{Q - \pi\theta R^L - (1-\pi)R^L} \tag{1.9}$$

projects, and earns the expected payoff

$$v_s = m_s\pi\left(R^H - \theta R^L\right). \tag{1.10}$$

Corollary 1. *The ability of a safe bank to finance its projects externally falls as the high state becomes more likely.*

The intuition for the result is simple. As the high state becomes more likely, the banker's moral hazard becomes worse. He has a stronger incentive to squeeze equity investors and a get higher share of the pie in the high state, so the amount of outside equity the bank can issue falls. The ability of a safe bank to use deposits to reduce this moral hazard is limited by the fact that it must be able to repay depositors in full in all states. The solid line in Figure 1.3 illustrates the fraction of expected returns pledgeable by a safe bank as a function of π.

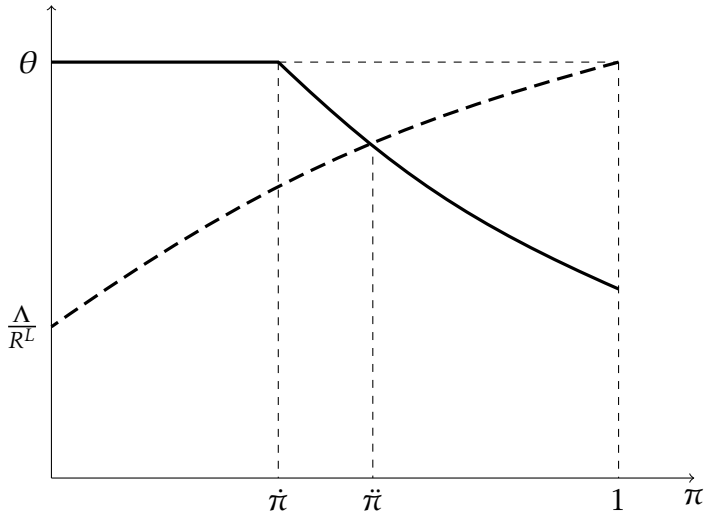


Figure 1.3: Fraction of expected projects' value that safe (solid line) and risky (dashed line) banks can pledge to external investors

Risky strategy

This brings us to the second, risky, strategy that the bank may pursue when the high state is sufficiently likely. Namely, the bank may accept that it will fail in the low state, but use deposits to raise more financing against the high state, and start more projects. Taking as given that the bank has $d > m\theta R^L$ and fails in the low state, its problem becomes:

$$\max \quad v = \pi\left(mR^H - d\right) \tag{1.11}$$

$$\text{s.t.} \quad m = \frac{n+b}{Q}$$

$$b = \pi d + (1-\pi)m\Lambda \tag{1.12}$$

$$d \leq m\theta R^H \tag{1.13}$$

Constraint (1.12) is the break-even condition of the bank's depositors, who run and liquidate the bank in the low state; (1.13) limits deposit liabilities of a risky bank to what it can feasibly repay in the high state. Risky bank issues no outside equity and $e^H = 0$, since otherwise it could squeeze equity investors out by setting $\tilde{\theta} = \frac{d}{mR^H}$, while still avoiding the run and liquidation in the high state.

Combining (1.11) and (1.12), it is easy to see that for the risky strategy to earn a positive expected profit, and thus for the bank to even consider it over the safe one, we must have $Q < \pi R^H + (1-\pi)\Lambda$, i.e. the capital price should not exceed the value of the projects, accounting for the liquidation in the L state. When that is the case, the bank's payoff is again maximized when external financing b is maximized. This is achieved when the bank issues the maximum level of deposit liabilities, $d = m\theta R^H$. This way the bank commits to repay the maximum pledgeable fraction θ of the returns in the high state.

Lemma 3. *The solution to the risky bank's problem is achieved when constraint* (1.13) *binds and the bank promises maximum pledgeable fraction of returns in the high state to depositors. This allows the bank to raise external financing* $b_r/m_r = \pi\theta R^H + (1-\pi)\Lambda$ *per project, supervise*

$$m_r = \frac{n}{Q - \pi\theta R^H - (1-\pi)\Lambda} \tag{1.14}$$

projects, and earn the expected payoff

$$v_r = m_r \pi(1-\theta)R^H. \tag{1.15}$$

The risky strategy allows the bank to raise more external financing against the high state, so as the high state becomes more likely, the ability of a risky bank to finance itself externally improves. The dashed line on Figure 1.3 shows the amount of external financing that a risky bank can raise as a fraction of the first-best expected value of its projects, $\mathbb{E}[R^s]$. As π approaches 1, this fraction approaches the maximum feasible value θ.

Safe or risky?

So when will the bank adopt a high, risky level of deposits?

Lemma 4. *The necessary conditions for a bank to consider risky strategy are* $Q < \pi R^H + (1-\pi)\Lambda$ *and* $\pi > \ddot{\pi}$*, where*

$$\ddot{\pi} \equiv \frac{R^L - \Lambda}{\theta(R^H - R^L) + R^L - \Lambda} \in (\dot{\pi}, 1). \tag{1.16}$$

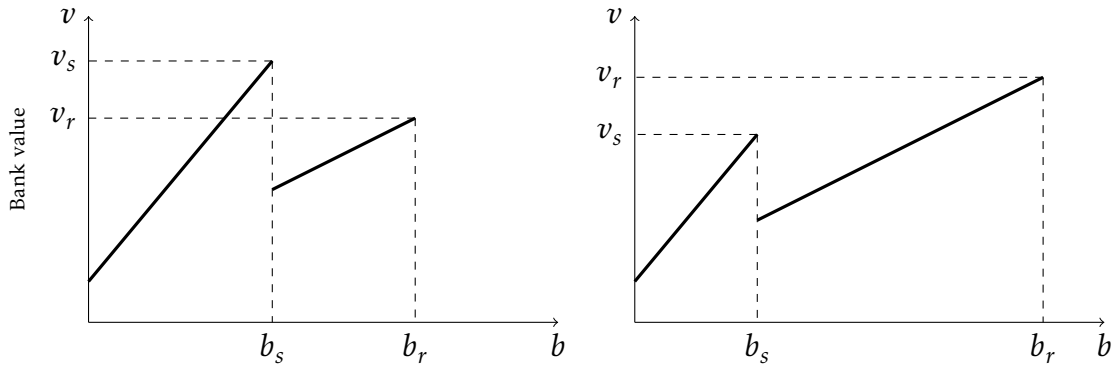
Then, the bank strictly prefers the risky strategy to the safe strategy iff $Q < \bar{Q}$*, where:*

$$\bar{Q}(\pi) \equiv \pi R^H + (1-\pi)\Lambda - \frac{(1-\pi)(1-\theta)(R^L - \Lambda)R^H}{\theta(R^H - R^L)}, \tag{1.17}$$

i.e., ceteris paribus, (a) π *is sufficiently high; (b)* Q *is sufficiently low.*

When $\pi < \ddot{\pi}$, the bank can obtain more external financing following the safe strategy, hence there is no benefit at all of increasing the leverage and becoming risky. When $\pi > \ddot{\pi}$, the bank trades off the ability to rise more financing and run more projects offered by the risky strategy vs. losses due the bank run and liquidation in the low state. Figure 1.4 depicts the banker's expected value as a function of the external financing the bank raises, holding the market price of capital fixed. The bank's value increases linearly up until b_s, at which point it drops discontinuously: to raise more external financing than b_s, the bank must lever itself up with deposits beyond what it can repay in the low state, implying a bank run and liquidation in the L state. The value then continues to increase linearly in b up until b_r, although now with a lower slope, reflecting the lower value of each project. The bank's problem reduces to simply choosing between the two peaks: v_s (playing it conservatively) and v_r (betting on the high growth, but suffering a run if it slows down), whichever yields a higher value. Comparing the value in (1.15) and (1.10), the second claim in the lemma follows.[22] The risky strategy is more appealing when the high productivity state becomes more likely and/or the capital becomes cheaper. The additional financing that the bank is

[22]Here I make use of the fact that in equilibrium we will necessarily have $Q > \pi\theta R^H + (1-\pi)\Lambda$ and $Q > \pi\theta R^L + (1-\pi)R^L$, otherwise banks could make infinite profits and would demand infinite amount of capital.



(a) Bank chooses safe strategy (b) Bank chooses risky strategy

Figure 1.4: Bank's value as a function of external financing

able to obtain translates into more additional projects when Q is low.

1.5.2 Equilibrium and financial crises

Equilibrium requires that financial and capital markets clear in period 0. The key equilibrating variable is the price of capital Q: the demand for capital from banks that take Q as given must equal the supply of capital by households. At the same time, the banks' demand for capital depends on the amount of external financing they obtain, which, as we have seen, also depends on Q.

We say that there is a risk of a financial crisis if all or a substantial fraction of banks choose the risky strategy in period 0, and are run and liquidated in the L state. The central result of this paper, formally established in the proposition below, is that such risk indeed emerges in equilibrium during booms, when the probability of the high productivity state is large.

Proposition 1. *There is an essentially unique equilibrium in this economy (with respect to aggregate outcomes),*[23] *which depends on the probability of the high state π. Suppose $N < (1-\theta)R^H$, i.e. the aggregate net worth of the banking sector is sufficiently scarce.*

[23] While there may be multiple equilibria, notably in cases (b) and (c), the essential uniqueness here means that these equilibria are always equivalent in terms of the aggregate outcomes, e.g. the amount of credit, the price of capital, and the fraction of banks that end up being risky.

Then there are thresholds

$$\underline{\pi} = \frac{N\frac{\theta(R^H-R^L)}{R^H-\theta R^L} + R^L - \Lambda}{\theta(R^H-R^L)+R^L-\Lambda} \quad \textit{and} \quad \overline{\pi} = \frac{N\frac{\theta(R^H-R^L)}{(1-\theta)R^H} + R^L - \Lambda}{\theta(R^H-R^L)+R^L-\Lambda},$$

with $\ddot{\pi} < \underline{\pi} < \overline{\pi} < 1$, *such that:*

a. *If* $\pi \geq \overline{\pi}$, *there is a unique equilibrium in which all banks are financed with the high, risky level of deposits. This facilitates a credit and asset price boom in period 0: banks borrow* $\pi\theta R^H + (1-\pi)\Lambda$ *per project, and drive the equilibrium price of capital to*

$$Q^* = N + \pi\theta R^H + (1-\pi)\Lambda. \tag{1.18}$$

There is a systemic financial crisis in which all banks fail if the low state materialises.

b. *If* $\pi \in (\underline{\pi}, \overline{\pi})$, *a fraction* ζ^* *of banks (measured by net worth)*[24] *end up choosing risky level of deposits in equilibrium, where*

$$\zeta^* = \frac{(1-\theta)R^H}{N\theta(R^H-R^L)}[Q^* - \pi\theta R^L - (1-\pi)R^L - N],$$

while the remaining banks are safe. The equilibrium price of capital is $\bar{Q}$ *given in* (1.17). *The fraction* ζ^* *of banks that fail if the low state materialises increases in the ex-ante probability of the high state:* $\zeta^{*\prime}_{\pi} > 0$.

c. *If* $\pi \leq \underline{\pi}$, *all banks are safe, and never fail in equilibrium.*

The condition $N < (1-\theta)R^H$ is realistic. Assuming otherwise would imply that the banking sector has enough own funds to bid up the capital price to the full value of the investment projects even when the high state is certain.[25] But it is

[24] Here we require a technical assumption that banks are atomistic, such that such split is always possible for any $\zeta \in (0,1)$.

[25] And thus banks can never earn positive profits.

unlikely that banks would have accumulated that much net worth during normal times when the probability of the high state is low.

While so far we have assumed that the aggregate capital is fixed, it is not difficult to recognize that booms in the asset price and credit are likely to coincide with a boom in the aggregate investment. In Appendix 1.B, I formally extend the above results to the case in which the aggregate capital investment can vary.

To summarise the findings, during booms in productivity, the worsening moral hazard induces banks to rely exclusively on deposits, facilitating a boom in asset prices, credit, and aggregate investment. If the anticipated growth fails to materialise, however, the excessive deposit financing leads to a systemic financial crisis.

1.6 Welfare and policy implications

In this section I discuss policy implications of the theory presented above. We'll see that ex-ante leverage restrictions and ex-post bailouts may hurt welfare by stifling investment and destroying discipline, respectively. However, the regulatory supervision of banks that enhances their transparency and governance can sometimes eliminate crises and improve welfare.

1.6.1 Ex-ante leverage restrictions

While financial crises in the model are destructive and undesirable ex post, equilibria that feature crises are ex-ante constrained efficient. Unlike many models with financial accelerator effects, there are no pecuniary externalities that agents do not take into account here. Even though a growing reliance of banks on deposit-like short-term financing introduces the risk of a financial disaster, it also allows to increase capital investment when it is the most valuable: during booms with strong prospects of high future returns.

It is easy to see that sufficiently hard restrictions on ability of banks to leverage with deposits can always preclude financial crises in the present setting. However, this will not be Pareto improving: restricted leverage also means lower capital prices, and necessarily lower welfare of households, who supply capital. Interestingly, in some cases leverage restrictions may lead to higher profits of the banking sector, since they effectively limit the competition between banks on the capital market, and allow them to enjoy higher returns per project. Nevertheless, there is a net loss to the society in the form of missed investment opportunities.

This point is quite general, and would remain an important consideration in more realistic settings that incorporate additional social costs of bank failures which may not be fully internalised by financial market participants, e.g. unemployment. Even if these additional costs tilt the balance in favour of using leverage restrictions to preclude financial crises, policy makers must understand that such policies have significant economic costs.

1.6.2 Cleaning-up ex post: government bailouts

Unlike the models in which financial crises stem from panics and illiquidity, notably Diamond and Dybvig (1983), in the present model banking crisis is not a result of agents' failure to coordinate on a 'good' equilibrium. If growth opportunities deteriorate during a boom, banks are run by depositors because they become fundamentally insolvent, in the sense that their liabilities on deposits exceed what they can possibly repay.

Despite this, saving banks in a crisis is, in principle, desirable *ex post*, since liquidation of banking assets destroys wealth for both households and bankers. In practice, government rescue could take many forms: costly government guarantees of problematic banks' debt; lending to troubled banks at below-market interest rates; or simply injecting equity into banks. Arguably, most of these tools were

used in the US during the 2008 Financial Crisis.[26]

However, an unrestrained government support to failing banking sector would be ill advised. The theory suggests that fragile capital structure may be a market solution to bank discipline problem. Therefore, similarly to a large existing literature on moral hazard in the banking sector,[27] expectations of government rescue could completely destroy ex-ante incentives. If creditors expect that the government will use taxpayers' money to bail out troubled banks, they will have no reasons to care about bankers' conduct and incentives, leading to an overheated credit market, inflated asset prices, and high rent extraction by bankers. Incentives of banks to maintain high pledgeability of returns would be lost, and aggregate investment could be severely distorted.

Government bailouts would thus only be desirable ex ante if they are carried out in a way that preserves the market discipline. This requires that the government must be able to commit to bail out banks only in a genuine crisis driven by a significant decline in expected economic growth, but let fail any bank that tries to extract rents from its creditors or taxpayers.[28] How to generate such a commitment remains an important question actively discussed in academic literature and policy circles. In essence, this is an institutional design problem of giving proper incentives to the regulator responsible for bailouts. I conjecture that the optimal design is likely to require that the regulator (i) is well informed about the aggregate state of the economy, and (ii) is more reluctant to assist banks than a benevolent social planner would be. The latter would ensure that the regulator does not intervene unless there are significant benefits to the society from doing

[26] Despite the fact that Lehman Brothers was allowed to fail because the Fed deemed it fundamentally insolvent at the time, one can argue that other insolvent banks were effectively bailed out. Notwithstanding popular claims that the US government has made money on assisting troubled banks, this is not true if one takes into account the fair economic value of government's assistance: Lucas (2019) estimates the direct cost of bailouts to the US government to be in excess of £500bn, or 3.5% of GDP.

[27] Some of the prominent papers on the topic, among others, are Chari and Kehoe (2016); Farhi and Tirole (2012).

[28] Of course, if such commitment is in place, no bank would attempt to do so in equilibrium.

so, and could be achieved by making bailouts legally difficult and politically costly to the regulator.

1.6.3 Addressing financial frictions directly

The fundamental problem described in the paper is the inability of banks to credibly pledge future returns to external investors. The reliance of banks on rigid, demandable deposit-like instruments during booms, and the occasionally resulting financial crises are thus just the symptoms of the underlying issue.

One implication of the theory is the prediction that banking crises are more likely to erupt in developing countries that are characterised by weak financial institutions and governance problems, while, at the same time, having less stable political regimes and more volatile economic trends. Empirically, we know this to be true. The second implication is that the best remedy against financial crises is a developed and transparent financial sector with sufficient intermediation capacity to allow banks to refrain from fragile short-term financing – a point also forcefully stressed in Diamond and Rajan (2001a).

We can also ask a more immediate question: should the government monitor banks during booms to mitigate agency problems and enforce repayments to outside investors? The answer depends on the feasibility and costs of monitoring. Formally, suppose that by supervising banks, the regulator can eliminate moral hazard completely and ensure that a fraction θ of banks' returns gets repaid. However, an active regulator's interference in banks' business also reduces future returns by a fraction $1-\chi$, such that banks' projects only generate χR^s in state s. Let us assume that the regulator is household-friendly, i.e. its sole objective is to maximize the welfare of household, and places zero weight on bankers' profits. Since households supply capital, and always break-even in expectation on the financial market, household welfare maximization is tantamount to maximizing the capital price, which in turn depends on the amount of financing that banks

can sustain. Then one can show that:

Proposition 2. *If the cost of supervision* $1-\chi$ *is sufficiently low, specifically*

$$\chi > \frac{\max\left\{\underline{\pi}R^L + \frac{1-\underline{\pi}}{\theta}R^L, N + \underline{\pi}\theta R^L + (1-\underline{\pi})R^L\right\}}{\underline{\pi}R^H + (1-\underline{\pi})R^L}, \tag{1.19}$$

there exist π_1, π_2*, with* $\pi_1 < \underline{\pi} < \pi_2 < 1$*, such that banks are supervised iff* $\pi \in (\pi_1, \pi_2)$*. However, there is no supervision otherwise, and the equilibrium features the risk of a financial crisis in the low state when* $\pi > \pi_2$*.*

The main message is that the regulator will indeed actively supervise banks and eliminate the risk of a financial crisis when the probability of the high state is moderately high. The regulator, however, will choose to leave the market to its own devices once the high productivity state becomes very likely. Intuitively, when the probability of the low state becomes small, expected benefit of preventing a crisis falls; regulatory supervision thus becomes more costly than deposits in forcing repayments from banks in the high state. This potentially sheds some light on why, for example, regulation and supervision of financial sector in the US appears particularly lax in the years preceding the 2008 Financial Crisis: on the back of strong confidence in the high productivity growth, the regulators might had been afraid to do more harm than good by interfering.

We can graphically illustrate the proposition in the special case when $N \to 0$ (i.e. when the banking sector has negligibly low aggregate net worth), which is particularly easy to depict. When banks have very low net worth, the ability to raise external financing becomes their crucial concern in equilibrium, and banks are willing to become risky the moment it allows them to better finance themselves, i.e. when $\pi > \ddot{\pi}$. Figure 1.5 plots the price of capital Q as a fraction of its first-best level, $\mathbb{E}R = \pi R^H + (1-\pi)R^L$, assuming that either (i) the regulator actively supervises banks (dotted line), (ii) all banks rely on the market discipline and choose the

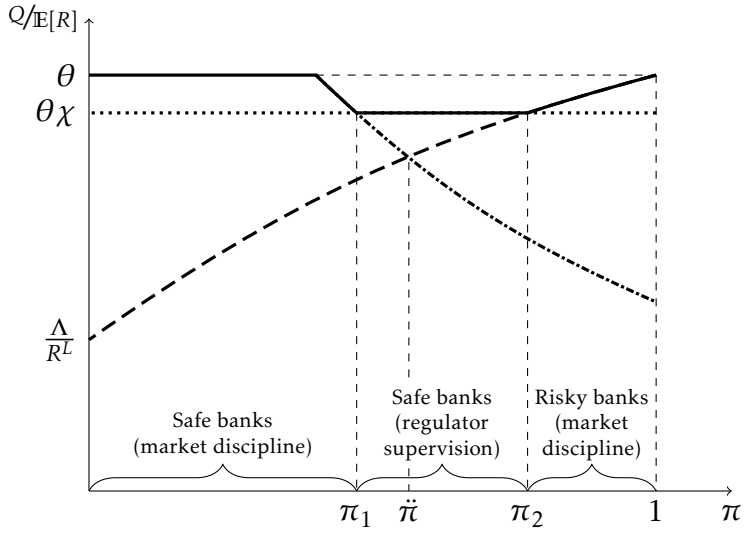


Figure 1.5: Capital price as a fraction of its first best level: market discipline vs regulatory supervision

safe strategy (dash-dotted line), or (iii) all banks rely on the market discipline and choose the risky strategy (dashed line). The latter two lines are effectively the same as in Figure 1.3, because the equilibrium capital price equals financing that banks can raise per project when $N \to 0$. In equilibrium under optimal supervision, the price of capital is the highest of these three lines, highlighted by the solid line on the plot. Supervision can achieve greater repayments from banks than the market discipline when $\pi \in [\pi_1, \pi_2]$, and it eliminates financial crises when $\pi \in (\ddot{\pi}, \pi_2]$. However, the regulator prefers to let banks finance themselves with risky deposits when $\pi > \pi_2$.

1.7 Conclusion

This paper presents a model in which financial crises result from build-ups of bank leverage and financial fragility during periods of rational optimism, when banks expect high returns on their long-term investments that ultimately do not materialise. The key mechanism is an increasing reliance of banks on deposit financing, which helps them overcome agency problems and increase funding, but

also exposes them to runs. The model is consistent with empirical evidence that credit booms and financial crises appear to be preceded by swings in the long-run productivity trends.

Ex-ante leverage restrictions can eliminate crises, but may hurt welfare, because they stifle investment at the time when it is most valuable. Unrestrained ex-post bailouts destroy ex-ante incentives and lead to inefficient levels of investment. They are therefore desirable only if the government can commit to interfere only when banks require help because of deteriorating fundamentals, but otherwise allow them to fail. How to engender such a commitment is a fascinating direction for future research.

Regulatory supervision of banks that enhances their accountability to investors can sometimes eliminate crises and improve welfare. However, it may still be optimal for the regulator to back away and let banks rely solely on market discipline during times when the confidence in high productivity growth is very strong. More generally, the theory presented here emphasizes that a transparent financial sector with strong corporate governance may not only enable superior allocation of resources, but also improve financial stability.

Bibliography

Aguiar, M. and Gopinath, G. (2007). Emerging Market Business Cycles: The Cycle Is the Trend. *Journal of Political Economy*, 115(1):69–102.

Akıncı, Ö. and Chahrour, R. (2018). Good news is bad news: Leverage cycles and sudden stops. *Journal of International Economics*, 114:362–375.

Allen, F. and Gale, D. (1998). Optimal financial crises. *Journal of Finance*, 53(4):1245–1284.

Anzoategui, D., Comin, D., Gertler, M., and Martinez, J. (2019). Endogenous Tech-

nology Adoption and R&D as Sources of Business Cycle Persistence. *American Economic Journal: Macroeconomics*, 11(3):67–110.

Benati, L. (2007). Drift and breaks in labor productivity. *Journal of Economic Dynamics and Control*, 31(8):2847–2877.

Bianchi, F., Kung, H., and Morales, G. (2019). Growth, slowdowns and recoveries. *Journal of Monetary Economics*, 101:47–63.

Boissay, F., Collard, F., and Smets, F. (2016). Booms and banking crises. *Journal of Political Economy*, 124(2):489–538.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2018). Diagnostic Expectations and Credit Cycles. *Journal of Finance*, LXXIII(1).

Brunnermeier, M. K. and Sannikov, Y. (2014). A Macroeconomic Model with a Financial Sector. *American Economic Review*, 104(2):379–421.

Calomiris, C. and Kahn, C. (1991). The Role of Demandable Debt in Structuring Optimal Banking Arrangements. *The American Economic Review*, 81(3):497–513.

Calomiris, C. W. and Gorton, G. (1991). The origins of banking panics: Models, facts, and bank regulation. In *Financial Markets and Financial Crises*, number January, pages 93–163.

Calomiris, C. W. and Mason, J. R. (2003). Fundamentals, panics, and bank distress during the depression. *American Economic Review*, 93(5):1615–1647.

Cao, D. and L'Huillier, J. P. (2018). Technological revolutions and the Three Great Slumps: A medium-run analysis. *Journal of Monetary Economics*, 96:93–108.

Cerra, V. and Saxena, S. C. (2008). Growth dynamics: The myth of economic recovery. *American Economic Review*, 98(1):439–457.

Cette, G., Fernald, J., and Mojon, B. (2016). The pre-Great Recession slowdown in productivity. *European Economic Review*, 88:3–20.

Chari, V. V. and Kehoe, P. J. (2016). Bailouts , Time Inconsistency, and Optimal Regulation: A Macroeconomic View. *American Economic Review*, 106(9):2458–2493.

Cummins, J. G. and Violante, G. L. (2002). Investment-Specific Technical Change in the United States (1947-2000): Measurement and Macroeconomic Consequences. *Review of Economic Dynamics*, 284:243–284.

De Long, J. B. (2003). Productivity Growth in the 2000s. In *NBER Macroeconomics Annual*.

Dell'Ariccia, G., Igan, D., Laeven, L., and Tong, H. (2012). Policies for Macrofinancial Stability: How to Deal with Credit Booms. Technical report, IMF Staff Discussion Note.

Diamond, D. W. and Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, 91(3):401–419.

Diamond, D. W. and Rajan, R. G. (2000). A theory of bank capital. *Journal of Finance*, 55(6):2431–2465.

Diamond, D. W. and Rajan, R. G. (2001a). Banks, Short Term Debt and Financial Crises: Theory, Policy Implications and Applications. In *Carnegie-Rochester Conference Series on Public Policy*, volume 54, pages 37–71.

Diamond, D. W. and Rajan, R. G. (2001b). Liquidity risk, liquidity creation, and financial fragility: A theory of banking. *Journal of Political Economy*, 109(2):287–327.

Edge, R. M., Laubach, T., and Williams, J. C. (2007). Learning and shifts in long-run productivity growth. *Journal of Monetary Economics*, 54(8):2421–2438.

Ennis, H. M. and Keister, T. (2003). Economic growth, liquidity, and bank runs. *Journal of Economic Theory*, 109(2):220–245.

Farhi, E. and Tirole, J. (2012). Collective Moral Hazard, Maturity Mismatch, and Systemic Bailouts. *American Economic Review*, 102(1):60–93.

Fernald, J. G. (2014). *Productivity and potential output before, during, and after the great recession*, volume 29.

Fisher, I. (1933). The Debt-Deflation Theory of Great Depressions. *Econometrica*, 1(4):337.

Flemming, J., L'Huillier, J. P., and Piguillem, F. (2019). Macro-prudential taxation in good times. *Journal of International Economics*, 121:103251.

Garcia-Macia, D. (2017). The Financing of Ideas and the Great Deviation.

Gertler, M. and Karadi, P. (2011). A model of unconventional monetary policy. *Journal of Monetary Economics*, 58(1):17–34.

Gertler, M. and Kiyotaki, N. (2015). Banking, liquidity, and bank runs in an infinite horizon economy. *American Economic Review*, 105(7):2011–2043.

Gertler, M., Kiyotaki, N., and Prestipino, A. (2019a). A Macroeconomic Model with Financial Panics. *The Review of Economic Studies*, (May 2019):240–288.

Gertler, M., Kiyotaki, N., and Prestipino, A. (2019b). Credit Booms, Financial Crises and Macroprudential Policy.

Gordon, R. J. (2003). Exploding Productivity Growth : Context , Causes , and Implications. *Brookings Papers on Economic Activity*, 2:207–298.

Gorton, G. and Metrick, A. (2012). Securitized banking and the run on repo. *Journal of Financial Economics*, 104(3):425–451. Market Institutions, Financial Market Risks and Financial Crisis.

Gorton, G. and Ordoñez, G. (2019). Good Booms, Bad Booms. *Journal of the European Economic Association*, 00(0):1–48.

Guerron-quintana, P. A. and Jinnai, R. (2014). Liquidity, Trends, and the Great Recession.

He, Z. and Krishnamurthy, A. (2019). A macroeconomic framework for quantifying systemic risk. *American Economic Journal: Macroeconomics*, 11(4):1–37.

Jorgenson, D. W., Ho, M. S., and Stiroh, K. J. (2004). Will the U.S. Productivity Resurgence Continue? *Current Issues in Economics and Finance*, 10(13):1–7.

Kahn, J. A. and Rich, R. W. (2007). Tracking the new economy: Using growth theory to detect changes in trend productivity. *Journal of Monetary Economics*, 54(6):1670–1701.

Krishnamurthy, A. and Muir, T. (2017). How Credit Cycles across a Financial Crisis.

L'Huillier, J. P., Phelan, G., and Wieman, H. (2020). Minsky Cycles: A Technology-Based Theory of Financial Crises and Macroeconomic Slumps.

Lorenzoni, G. (2008). Inefficient Credit Boom. *Review of Economic Studies*, pages 809–833.

Lucas, D. (2019). Measuring the Cost of Bailouts. *Annual Review of Financial Economics*, pages 85–108.

Martin, A., Skeie, D., and Thadden, E. L. V. (2014). Repo runs. *Review of Financial Studies*, 27(4):957–989.

Mendoza, E. G. (2010). Sudden Stops, financial crises, and leverage. *American Economic Review*, 100(5):1941–1966.

Mendoza, E. G. and Terrones, M. E. (2008). An Anatomy of Credit Booms: Evidence from Macro Aggregates and Micro Data.

Mendoza, E. G. and Terrones, M. E. (2014). An Anatomy of Credit Booms and their Demise. In Fuentes, M. D., Raddatz, C. E., and Reinhart, C. M., editors, *Capital Mobility and Monetary Policy*, pages 165–204. Central Bank of Chile, Santiago, Chile.

Paul, P. (2020). Historical Patterns of Inequality and Productivity around Financial Crises.

Queralto, A. (2019). A model of slow recoveries from financial crises. *Journal of Monetary Economics*, (xxxx).

Reinhart, C. M. and Reinhart, V. R. (2010). After the Fall. In *Macroeconomic Challenges: The Decade Ahead*, pages 17–60. Federal Reserve Bank of Kansas City Economic Policy Symposium.

Reinhart, C. M. and Rogoff, K. S. (2009). The Aftermath of Financial Crises. *American Economic Review: Papers & Proceedings*, 99(2):466–472.

Reinhart, C. M. and Rogoff, K. S. (2014). Recovery from financial crises: Evidence from 100 episodes. *American Economic Review: Papers & Proceedings*, 104(5):50–55.

Schularick, M. and Taylor, A. M. (2012). Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870-2008. *American Economic Review*, 102(2):1029–1061.

Seoane, H. D. and Yurdagul, E. (2019). Trend shocks and sudden stops. *Journal of International Economics*, 121.

Stein, J. C. (2012). Monetary policy as financial stability regulation. *Quarterly Journal of Economics*, 127(1):57–95.

Stiroh, K. and Botsch, M. (2007). Information Technology and Productivity Growth in the 2000s. *German Economic Review*, 8(2):255–280.

Thakor, A. V. (2018). Post-crisis regulatory reform in banking: Address insolvency risk, not illiquidity! *Journal of Financial Stability*, 37:107–111.

Appendices

1.A Productivity, booms, and the 2008 Crisis

There is ample empirical evidence of low-frequency variation in the long-run growth rate of productivity. For example, Kahn and Rich (2007) show that a Markov switching model in which labour productivity growth can be either high or low, with a small probability of transition between the states, fits the postwar US data surprisingly well.[29] Moreover, accelerations in average productivity growth appear to be associated with starts of credit booms; and, while financial crises often follow credit booms, many credit booms do not, in fact, end up in crises (Gorton and Ordoñez, 2019; Dell'Ariccia et al., 2012). Crises, however, typically follow after several years of slow productivity growth (Gorton and Ordoñez, 2019; Paul, 2020). Evidence thus appears to suggest that crises are productivity booms gone wrong: they follow booms in which accelerated productivity growth ultimately disappoints.

The Great Recession is a case in point. It is well documented that there has been a marked acceleration in the US productivity growth in the mid 1990s after more than 25 years of sluggish growth, widely attributed to information and communication technologies.[30] Figure 1.A.1, taken from the up-to-date version

[29] See also Edge et al. (2007); Benati (2007), and references therein.

[30] See Benati (2007); Kahn and Rich (2007); Fernald (2014). In addition, Fernald (2014) demonstrates that although capital deepening played a role, high growth of labour productivity in the late 1990s- early 2000s was driven primarily by TFP.

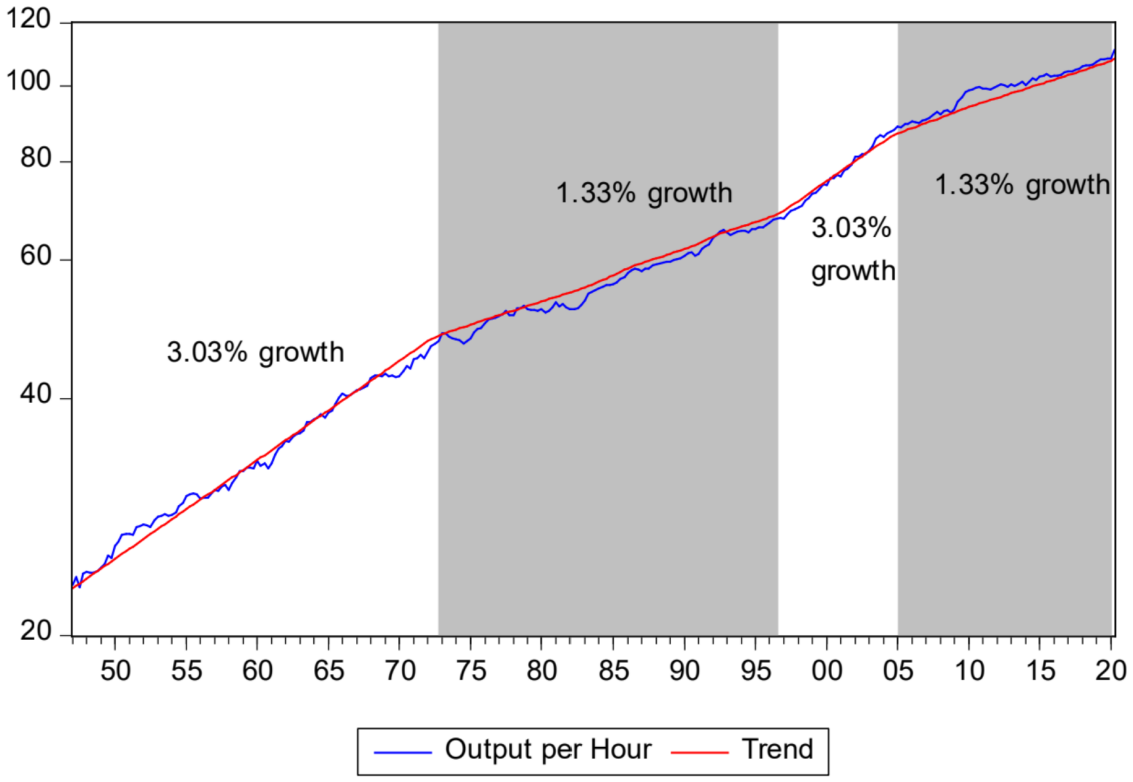


Figure 1.A.1: US labour productivity growth trends (log scale)
Source: James Kahn's website.

of Kahn and Rich (2007), demonstrates a shift to the high-growth regime around 1997. The gradual recognition of the New Economy prompted many people, including professional economists, to express confidence in high rate of growth for the decades to come.[31]. Figure 1.A.2 shows that medium professional forecasts of 10-year average productivity growth nearly doubled in the 3 years 1997-2000, and were hardly affected by the dot-com bust. At the time, they were not necessarily unreasonable: since shifts in trend are infrequent and difficult to predict, the best forecast following an acceleration of productivity may well be to expect that the economy will experience a prolonged period of high growth. The figure also shows that this period was marked by a large credit boom in the US. As is evident from the figures, however, the trend productivity growth has slowed down again around the beginning of 2005. The evidence that productivity has slowed *well*

[31] Among many others were: De Long (2003); Cummins and Violante (2002); Jorgenson et al. (2004); Stiroh and Botsch (2007); Gordon (2003).

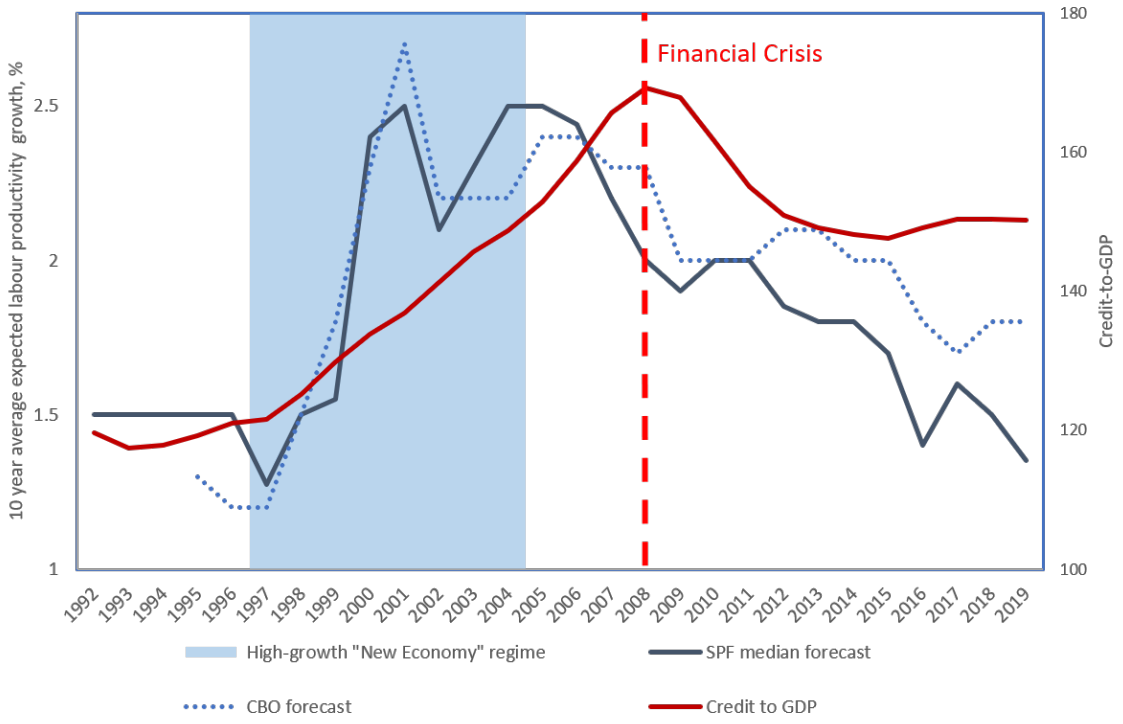


Figure 1.A.2: The New Economy, long-run productivity growth expectations, and credit in the US
Source: Kahn and Rich (2007), BIS, FRB Philadelphia

before rather than as a result of the Great Depression is also presented in Fernald (2014); Cette et al. (2016); Cao and L'Huillier (2018). Long-term forecasts began to decline, and the Financial Crisis broke out three years later.

1.B Capital investment

In this appendix I relax the assumption that the stock of capital is fixed. This introduces realistic deadweight losses from restricting banks' leverage, since this affects the aggregate investment.

Suppose that capital is produced by households, who only use their labour as an input. Each household runs a competitive capital producing firm and chooses the level of output to maximize its profit:

$$\max_{k_i} Qk_i - \gamma_i \frac{k_i^2}{2}, \tag{1.B.1}$$

where k_i is the amount of capital produced, and the quadratic term represents the disutility of labour required to produce k_i units of capital. First-order condition is

$$k_i(Q) = \frac{Q}{\gamma_i}, \tag{1.B.2}$$

which, aggregated over all households, yields the aggregate supply of capital:

$$K^S(Q) = \int_{i \in \mathcal{H}} \frac{Q}{\gamma_i} \mathrm{d}F_h(i) = \frac{Q}{\gamma}. \tag{1.B.3}$$

The following proposition, similarly to Proposition 1 in the paper, establishes that the risk of a financial crisis emerges in equilibrium when aggregate banks' net worth is relatively scarce, and the probability of a high productivity state is sufficiently high.

Proposition 1.B.1. *There is an essentially unique equilibrium,*[32] *which is parameter-dependent. Suppose the aggregate net worth of the banking sector is sufficiently scarce, namely* $N < \frac{1-\theta}{\gamma} R^{H\,2}$. *Then there exist thresholds* $\underline{\pi}, \overline{\pi}$, *with* $\ddot{\pi} < \underline{\pi} < \overline{\pi} < 1$, *such that:*

- *If* $\pi \geq \overline{\pi}$, *all banks choose risky strategy in equilibrium, and there is a financial crisis if the low state materialises.*
- *If* $\pi \in (\underline{\pi}, \overline{\pi})$, *some banks are safe, and some are risky in equilibrium.*
- *If* $\pi \leq \underline{\pi}$, *all banks choose safe strategy, and never fail in equilibrium.*

Again, $N < \frac{1-\theta}{\gamma} R^{H\,2}$ is a plausible assumption. It implies that the banking sector does not have enough own resources to be able to bid up the price of capital to the first-best level even when the high state becomes certain.

The central results therefore remain similar when we allow for the aggregate capital to vary, although it is more challenging to obtain closed-form solutions

[32] With respect to aggregate outcomes.

for probability thresholds, capital prices, and investment. However, it is straightforward to characterize the equilibrium in the limiting case where $N \to 0$, i.e. the banking sector is close to penniless in the aggregate, and raises most of its financing externally. In that case, the price of a unit of capital approaches the amount of external financing that banks can raise per project, and we have the following result:

Corollary 1.B.1. *As $N \to 0$, both thresholds $\underline{\pi}, \overline{\pi} \to \ddot{\pi}$, and the equilibrium capital price is*

$$Q^* \to \begin{cases} \theta\left(\pi R^H + (1-\pi)R^L\right) & \text{if } \pi \leq \dot{\pi} \\ \pi\theta R^L + (1-\pi)R^L & \text{if } \pi \in (\dot{\pi}, \underline{\pi}] \\ \pi\theta R^H + (1-\pi)\Lambda & \text{if } \pi \geq \overline{\pi}. \end{cases} \tag{1.B.4}$$

The equilibrium capital investment is $K^* = Q^*/\gamma$.

When banks have a very low net worth, the ability to raise external financing becomes a crucial concern in equilibrium, and banks are willing to become risky the moment it allows them to better finance themselves, as reflected in $\overline{\pi}$ falling to $\ddot{\pi}$.

Because risky banks are able to raise more financing, increasing the demand for capital, we also have:

Corollary 1.B.2. *When π is large and the equilibrium features the risk of a financial crisis, the capital prices and aggregate investment are higher compared to a case in which all banks restrict their leverage to avoid a financial crisis in the L state.*

In other words, the deposit financing indeed enables a boom in asset prices and investment when there is a productivity boom.

1.C Proofs

Proof of Proposition 1. It is straightforward to verify that $\ddot{\pi} < \underline{\pi} < \overline{\pi} < 1$ when $N \in (0, (1-\theta)R^H)$. Let us prove each part of the proposition consecutively.

a. First, we show this is indeed an equilibrium. If all banks choose risky strategy and maximum level of deposits they can repay only in the high state, the aggregate demand for capital is:

$$K^D(Q) = \frac{N}{Q - \pi\theta R^H - (1-\pi)\Lambda}.$$

which follows from aggregating (1.14). Together with fixed capital supply $K^S = 1$, this immediately implies the equilibrium price of capital in (1.18). Banks are indeed behaving optimally iff $Q^* \leq \bar{Q}$ by Lemma 4, which, with a little bit of algebra, can be shown to hold iff $\pi \geq \overline{\pi}$, as required.

To prove uniqueness, suppose there is a different equilibrium, in which fraction $\zeta < 1$ of banks, measured by net worth, are risky, and the rest $1-\zeta$ are safe. Aggregate demand for capital is given by

$$K^D = \frac{\zeta N}{Q - \pi\theta R^H - (1-\pi)\Lambda} + \frac{(1-\zeta)N}{Q - \pi\theta R^L - (1-\pi)R^L}. \tag{1.C.1}$$

Combined with equilibrium condition $K^D = K^S \equiv 1$, it is straightforward to show that equilibrating price Q is an increasing function of ζ. So in this other equilibrium price of capital is below Q^* found in (1.18), and thus also below $\bar{Q}$. But then all banks would strictly prefer risky strategy by Lemma 4, a contradiction.

b. To show that it is indeed an equilibrium outcome, first note that because equilibrium price is $\bar{Q}$, individual banks are indifferent between safe and risky strategy. Combining (1.17) with aggregate demand in (1.C.1) and

market clearing condition, it is straightforward to show that $\bar{Q}$ is indeed the equilibrating price when exactly fraction ζ^* of banks (measured by net worth) choose risky level of deposits.

There are multiple equilibria, since there are numerous strategy profiles, pure and mixed, that result in fraction ζ^* of banks choosing risky level of deposits. The equilibrium aggregate outcome described in the proposition, however, is unique. There is no equilibrium in which more than fraction ζ^* of banks chose risky level of deposits, since that would lead to more financing and higher capital price in period 0, but then all banks would strictly prefer to be safe. Similarly, there is no equilibrium in which less than fraction ζ^* of banks end up choosing risky strategy, because that would imply lower capital price and all banks preferring risky strategy.

c. When $\pi \leq \underline{\pi}$, there is no equilibrium in which any of the banks end up being risky. Firstly, when $\pi \leq \ddot{\pi}$, all banks are safe by Lemma 4. Secondly, if such an equilibrium where to exist when $\pi \in (\ddot{\pi}, \underline{\pi}]$, the demand for capital would be given, again, by (1.C.1). But then it is straightforward to verify that equilibrating capital price would always exceed $\bar{Q}$ whenever $\zeta > 0$. But then all banks would strictly prefer being safe, a contradiction.

Q.E.D.

Proof of Proposition 2. We are going to proceed with the proof as follows. First, we derive the condition (1.19), which the condition under which the regulator implements supervision policy when the probability of high state is $\underline{\pi}$. Second, we demonstrate that if the condition (1.19) is satisfied, there exist π_1, π_2, with $\pi_1 < \underline{\pi} < \pi_2 < 1$, such that banks are supervised iff $\pi \in (\pi_1, \pi_2)$. Lastly, we show that it is never optimal for the regulator to supervise banks if it is not optimal to do so when $\pi = \underline{\pi}$, i.e. (1.19) does not hold.

Without bank supervision, the equilibrium price of capital is

$$Q^*(\pi) = \begin{cases} \min\left\{N + \theta[\pi R^H + (1-\pi)R^L], \pi R^H + (1-\pi)R^L\right\} & \text{if } \pi \leq \dot{\pi} \\ \min\left\{N + \pi\theta R^L + (1-\pi)R^L, \pi R^H + (1-\pi)R^L\right\} & \text{if } \pi \in (\dot{\pi}, \underline{\pi}] \\ \bar{Q}(\pi) & \text{if } \pi \in (\underline{\pi}, \overline{\pi}) \\ N + \pi\theta R^H + (1-\pi)\Lambda & \text{if } \pi > \overline{\pi}, \end{cases} \tag{1.C.2}$$

where the last two cases are already given in Proposition 1. The first two follow from either: (1) aggregating capital demand in Lemma 2 and combining it with market clearing condition, under the assumption that banks are constrained and exhaust their financing capacity, or (2) the maximum price that unconstrained banks would pay for capital, $\mathbb{E}[R^S] = \pi R^H + (1-\pi)R^L$, whichever is *smaller*. It is straightforward to verify that $Q^*(\pi)$ is continuous on $[0,1]$.

When policy is implemented, the price of capital is simply

$$Q_p(\pi) = \min\left\{N + \theta\chi[\pi R^H + (1-\pi)R^L], \chi[\pi R^H + (1-\pi)R^L]\right\}. \tag{1.C.3}$$

The first term in the min operator again follows from the capital market clearing condition assuming banks are constrained and exhaust their financing capacity. The second term caps the equilibrating price at the maximum level that unconstrained banks would pay for capital when the policy is implemented, $\chi\mathbb{E}[R^S]$. $Q_p(\pi)$ is an increasing and continuous function on $[0,1]$.

At $\pi = \underline{\pi}$, banks are constrained and so $Q^*(\underline{\pi}) = N + \underline{\pi}\theta R^L + (1-\underline{\pi})R^L$. The household-friendly regulator will find it optimal to supervise banks if and only if $Q_p(\underline{\pi}) > Q^*(\underline{\pi})$, from which the condition (1.19) follows.

Suppose the condition (1.19) is satisfied, and $Q_p(\underline{\pi}) > Q^*(\underline{\pi})$. Then there is $\pi_1 \in (\dot{\pi}, \underline{\pi})$ s.t. $Q_p(\underline{\pi}) > Q^*(\pi)$ when $\pi \in (\pi_1, \underline{\pi}]$, but $Q_p(\underline{\pi}) \leq Q^*(\pi)$ when $\pi \leq \pi_1$. This follows by recognizing that $Q^*(\pi) > Q_p(\pi)$ when $\pi \leq \dot{\pi}$ or when banks are

unconstrained and $Q^* = \mathbb{E}[R^s]$. Therefore, we can only have $Q_p(\pi) > Q^*(\pi)$ in the region where $Q^*(\pi) = N + \pi\theta R^L + (1-\pi)R^L$, which is a decreasing function of π, and so must intersect $Q_p(\pi)$ at a single point, π_1.

There is also $\pi_2 \in (\underline{\pi}, 1)$ s.t. $Q_p(\underline{\pi}) > Q^*(\pi)$ when $\pi \in [\underline{\pi}, \pi_2)$, but $Q_p(\underline{\pi}) \leq Q^*(\pi)$ when $\pi \geq \pi_2$. This follows because, firstly, $Q_p(\pi)$ must cross $Q^*(\pi)$ at least once, since it is easy to verify that $Q^*(1) > Q_p(1)$. Secondly, once $Q_p \leq Q^*$ at some $\hat{\pi} > \underline{\pi}$, then $Q_p(\pi) \leq Q^*(\pi)\ \forall \pi \in [\hat{\pi}, 1]$. Let us now show the latter.

First, it is straightforward to see that if $Q_p \leq Q^*$ at some $\hat{\pi} \in [\underline{\pi}, \overline{\pi}]$, then $Q_p(\pi) \leq Q^*(\pi)\ \forall \pi \in [\hat{\pi}, \overline{\pi}]$. This is because on the interval $[\underline{\pi}, \overline{\pi}]$, $\bar{Q}(\pi)$ is a linear function with a slope greater than $R^H - \Lambda$, and so always greater than the slope of $Q_p(\pi)$, which is at most $\chi(R^H - R^L)$. Next, we also have that if $Q_p \leq Q^*$ at $\hat{\pi} \in [\overline{\pi}, 1]$, then $Q_p(\pi) \leq Q^*(\pi)\ \forall \pi \in [\overline{\pi}, 1]$. This is straightforward to see if, when supervised, banks are constrained in equilibrium at $\hat{\pi}$ and so $Q_p(\pi) = N + \theta\chi[\pi R^H + (1-\pi)R^L]$ on $[\hat{\pi}, 1]$, which again has lower slope than that of Q^*. If, when supervised, banks are unconstrained at $\hat{\pi}$, then $Q_p(\hat{\pi}) = \chi[\hat{\pi}R^H + (1-\hat{\pi})R^L]$. With a bit of algebra, we can verify, however, that $Q_p(\hat{\pi}) \leq Q^*(\hat{\pi})$ at this point also implies that $N + \theta\chi[\hat{\pi}R^H + (1-\hat{\pi})R^L] < Q^*(\hat{\pi})$.[33] This, in turn, implies that Q_p cannot possibly exceed Q^* as π increases.

Taken together, the above implies that the regulator supervises banks iff $\pi \in (\pi_1, \pi_2)$ when (1.19) holds.

The above considerations about the behaviour of the functions $Q^*(\pi)$ and $Q_p(\pi)$ also lead us to conclude that there is no π at which the regulator would find it optimal to supervise banks when the condition (1.19) is not satisfied. As $Q_p \leq Q^*$ at $\underline{\pi}$, it must be the case that $Q_p \leq Q^*$ when $\pi < \underline{\pi}$, since even in the region where banks are constrained and $Q^*(\pi) = N + \pi\theta R^L + (1-\pi)R^L$, $Q^*(\pi)$ increases

[33] Using $Q_p(\hat{\pi}) \leq Q^*(\hat{\pi})$, it follows that $N + \theta\chi[\hat{\pi}R^H + (1-\hat{\pi})R^L] < Q^*(\hat{\pi})$ whenever $\theta N < (1-\theta)[\hat{\pi}\theta R^H + (1-\hat{\pi})\Lambda]$. The right-hand-side of the last inequality increases in $\hat{\pi}$. This inequality holds, however, even when $\hat{\pi} = \overline{\pi}$, which can be verified by substituting in $\overline{\pi}$ from Proposition 1 and rearranging.

and $Q_p(\pi)$ falls when π decreases below $\underline{\pi}$. Similarly, $Q_p \leq Q^*$ at $\underline{\pi}$ implies that $Q_p(\pi) \leq Q^*(\pi)\ \forall \pi > \underline{\pi}$. *Q.E.D.*

Chapter 2

Sinking ships: asset prices and the predictability of returns on shipping in recessions

2.1 Introduction

> *In recessions ... the financial strength of shipping companies is tested and only the strongest survive. If the low rates persist the weak companies may end up selling their ships to the financially strong at distress prices, a game of pure Darwinian economics.*
>
> — Martin Stopford

Following the Financial Crisis of 2008, there has been an increased interest in the economics literature in boom-bust cycles in asset prices and real economic activity. One particularly familiar narrative is that economic crashes are inevitable consequences of unsustainable expansions, characterised by overly optimistic expectations, excessive credit growth, speculation, and overin-

vestment. Large booms in asset prices and investment, then, predictably lead to economic busts and reversals of fortunes for investors.

In this paper I study the link between movements in asset prices, investment, and future returns on real assets in a capital intensive industry. I first document empirical evidence that challenges the conventional wisdom in the context of dry bulk shipping. There appears to be no association between the current level of industry activity and future returns on investments during expansions. In other words, good times do not invariably lead to bad times. However, I find that during recessions, the returns on capital are strongly and negatively related to the industry earnings, asset prices, and investment. That is, the worse is the plunge, the handsomer is the expected reward to those who buy ships. Consistent with the anecdotal evidence from industry experts, I suggest an explanation of my empirical findings based on binding liquidity constraints faced by industry participants during downturns.

Dry bulk shipping provides the perfect setting for my analysis. With a large number of small shipowners and little scope for differentiation, it resembles perfect competition. It is, nevertheless, a highly capital-intensive industry. Because it takes several years to build a ship, the dry bulk fleet responds only slowly to developments in the demand for shipping services, which in turn is closely linked to global macroeconomic conditions. The combination of fairly inelastic demand and supply that is practically fixed in the short run results in enormous volatility of freight rates, and hence also industry earnings and returns.

Recurring plunges in dry bulk earnings are often blamed on the excessive optimism of shipping firms during booms. According to this view, it is overordering of ships in good times that ultimately causes a substantial supply glut and collapse in freight rates (Stopford, 2009). Indeed, Greenwood and Hanson (2015) find that second hand ship prices, earnings, and industry investment all negatively forecast future excess returns on owning dry bulk ships during 1976-2010. Their evidence

is consistent with the overinvestment story, and not simple bad luck. They suggest a behavioural theory of waves in ship prices and investment, according to which firms tend to both overextrapolate current demand conditions into the future, and also neglect the investment responses of their competitors. This results in classic cobweb-like industry dynamics, in which shipowners overreact during good times, and are predictably disappointed by the subsequent low returns.

I document new empirical evidence that casts doubt on this interpretation of the industry cycle, however. Given that international trade is closely related to the state of the world economy, the demand for shipping services is strongly affected by global economic booms and recessions. It is therefore natural to investigate whether the predictability of shipping excess returns varies systematically over the global business cycle. To evaluate the presence and magnitude of return predictability in shipping separately during macroeconomic expansions and recessions, I estimate regime-dependent predictive regressions. Specifically, I regress subsequent 12-, 24-, and 36- months cumulative excess returns on owning a five year old ship on each of current second hand ship prices, earnings, and industry investment, allowing the coefficients to differ between the two regimes. My monthly sample runs from 1976 through 2020, and I use the OECD based Recession Indicator for OECD and Non-member Economies as the regime dummy indicator of global recessions – one of the most comprehensive indicators of global recessions available.[1] To account for heteroscedasticity and autocorrelation of residuals, especially in the presence of overlapping observations, I employ the state-of-the-art robust inference method developed by Ye and Sun (2018).[2]

I find that return predictability in shipping is strongly countercyclical. During global recessions, the estimates of predictive slope coefficients on most variables and forecasting horizons are statistically significant, and appear economically

[1] The results are similar when alternative indicators of recessions are used, e.g. the NBER indicator of US recessions.

[2] My results are only stronger when Newey-West standard errors are used, tests based on which are known to over-reject the null.

large. For example, a one standard deviation decrease in real ship prices during recessions predicts an 11.7 percentage point (pp) increase in the subsequent 12-months excess returns, and a 25.1 pp increase in excess returns over the following 24 months. Meanwhile, I cannot reject the null of no predictability during expansions, and the estimates of slope coefficients are small in absolute value. Furthermore, in the majority of cases I can reject the null that slope coefficients in predictive regressions are equal across the two regimes. These results are robust to various checks designed to control for potential biases, outliers, inference issues, and measurement errors.

How can we explain such a significant return predictability arising specifically during downturns? It does not fit well with the idea that firms overreact to industry conditions, because the latter implies that predictability would always be present. Yet, I find little evidence of predictability in the whole estimation sample when not allowing for regimes.[3] Many financial economists instead suggest that the return predictability reflects systematic variation over time in the rate of return required by investors (Cochrane, 2011). Indeed, the finding that return predictability emerges strongly during recessions squares well with the intuition that required risk premia increase during downturns. However, I do not find compelling empirical evidence to support this hypothesis either. During recessions, ships appear to have particularly low hedge value (and hence are potentially required to generate higher returns) when ship prices, earning, and investments are high, rather than low. The predictive regressors are *negatively* associated with future shipping risk, and their ability to forecast future returns also does not weaken when controlling for variation over time in measures of the economy-wide risk premium.

Instead, I argue, it is binding financial constraints that cause slumps in ship

[3]This is in contrast with the main finding of Greenwood and Hanson (2015). The difference arises due to the combination of a longer data set, refined variable construction, and the more robust testing methodology.

prices and investment in recessions. The view that financial woes put huge downward pressure on second hand prices and investment during recessions is supported by numerous accounts from shipping industry experts and practitioners. Substantial return predictability is then simply a manifestation of ship prices falling below their fundamental value due to the lack of liquidity during these episodes.

It is only natural to expect financial constraints to play a critical role in industries like dry bulk shipping. Shipping is dominated by many small, often family-owned firms with specialist knowledge and informal corporate structures. The industry traditionally relies on bank financing, and, in addition, is characterised by high volatility of cash flows and asset prices. For these reasons, traditional lenders to shipping companies typically possess considerable industry expertise, while outsiders may find it difficult to make a profitable jump into shipping.

When earnings fall and cash flows dry up in recessions, many firms are forced to sell their assets in order to repay outstanding debts. At the same time, the potential buyers, – mostly other firms in the industry, – are also likely to be tight on cash, and may find it difficult to borrow funds. Their ability to absorb second hand ships from the distressed firms is therefore limited: prices of used ships have to fall below their fundamental value to equilibrate the distressed supply with the constrained demand. The worse is the recession, the more constrained the buyers are, and so the deeper prices fall, resulting in higher excess returns to the firms that do manage to buy ships. At the same time, the lack of financing and falling prices for second hand tonnage lead to a collapse in industry investment as well, implying also a negative association between investment and future returns.[4] Observe that while sales of second hand ships at depressed prices simply represent

[4]First, a fall in industry investment is driven by, and therefore correlated with the fall in earnings and asset prices. Second, constrained ordering of new ships due to financial frictions also implies higher expected earnings per ship in a few years time.

a transfer from the sellers to the buyers, plunges in industry investment likely reflect industry inefficiencies due to financial constraints.

I formalize this argument in a model of a competitive, financially constrained industry, in which firms are exposed to both aggregate and idiosyncratic shocks. Because of limited commitment problems on the firms' side, the optimal financial contracts in the model take the form of simple non-recourse debt secured by the resale value of capital.[5] In an expansion, most firms generate enough cash flows to repay debts and also have spare liquidity to purchase assets at their full fundamental value from the distressed firms hit by bad idiosyncratic shocks. In a recession, however, even firms with good idiosyncratic shocks have limited spare liquidity, and are unable to borrow, while the amount of capital that firms with bad shocks have to sell increases. The worse is the aggregate shock, the deeper the asset price falls below the fundamental value in order to equilibrate constrained demand and the forced sales. The buyers end up earning high excess returns, which are larger the deeper the recession is.

This paper contributes to the extensive empirical asset pricing literature on return predictability. For a comprehensive survey of this literature, see Cochrane (2011). Notably, Rapach et al. (2009), Henkel et al. (2011), and Dangl and Halling (2012) document that predictability of stock market returns is significantly stronger during recessions. However, in contrast with most of the asset pricing literature which focuses on returns on financial claims, this paper directly analyses the returns on real capital investments. To the best of my knowledge, Greenwood and Hanson (2015) is the only prior paper that studies return predictability on real assets in a capital-intensive industry. I significantly extend prior findings by documenting that predictability of returns on owning ships is present only during global recessions. To my knowledge, the present paper is also the first in the predictability literature to suggest that this phenomenon arises due

[5]In fact, Franks et al. (2021b) document that this is the commonplace debt contract in the shipping industry.

to liquidity constraints faced by firms in the industry.

My paper is also closely related to the literature on financial frictions and fire sales. Shleifer and Vishny (1992) were among the first to argue that when a financially distressed firm liquidates assets, other firms in the same industry, who are the natural buyers, are also likely to be experiencing hard times. This leads to specialized assets selling for prices below their value in best use, and also creates the possibility of misallocation if less efficient users end up buying them.[6] There has been a substantial empirical literature testing this hypothesis, notably Pulvino (1998) in the context of the airline industry, Campbell et al. (2011) in the context of the housing market, and Rajan and Ramcharan (2016) in the context of the US banking industry during the Great Depression.[7] Franks et al. (2021a) document that fire sale discounts in the airline industry predominantly reflect the liquidity channel and quality impairments due to under-maintenance of assets, while they find no evidence of misallocation. Because in this paper I focus on relatively new, five year old ships, under-maintenance is unlikely to be a major driver of my results. The sharp increases in expected returns that I document when asset prices fall in recessions, coupled with the ample anecdotal evidence of financial woes faced by the shipping firms, provides further evidence in support of liquidity constraints and fire sale discounts in the context of the shipping industry. The fact that plunges in ship cash flows and prices also appear to be accompanied by slumps in industry investment highlights the potential efficiency costs of liquidity constraints.

[6]The dynamic version of this idea has been at the centre of much of the literature on financial frictions and financial accelerator in macroeconomics, notably Kiyotaki and Moore (1997), Bernanke et al. (1999), and more recently Gertler and Karadi (2011) and Brunnermeier and Sannikov (2014), among others.

[7]For a comprehensive review of literature on fire sales in macroeconomics see Shleifer and Vishny (2011).

2.2 Dry bulk shipping: background and data

2.2.1 Basic facts about dry bulk shipping

Dry bulk shipping is an attractive industry for economic analysis, with a growing number of recent applications in the literature.[8] With a large number of small shipowners and little scope for differentiation, dry bulk shipping closely resembles the perfect competition setting. Figure 2.1 illustrates the point by showing the distribution of ship ownership across firms. The largest firm owns merely 3% of the fleet, and there is a long tail of small shipowners with just one ship each.

Dry bulk vessels mainly transport raw materials such as iron ore, grain, coal, phosphates and bauxite (five "major bulks"), as well as minor bulks. There are four major dry bulk ship categories: Handysize (below 40,000 DWT[9]), Handymax (40,000 - 60,000 DWT), Panamax (60,000 - 100,000 DWT), and Capesize (above 100,000 DWT). As Kalouptsidi (2014) points out, since vessels in different size categories can carry different products, take different routes, and approach different ports, practitioners sometimes treat them as different markets. However, the separation is not rigid, since in practice ships can move between adjacent segments in response to freight rates.

Demand for seaborne shipping is closely related to developments in the world economy and international trade. At least in the short run, it is fairly price inelastic: changes in transport costs have little impact on the volume of cargo transported. This is due to the fact that shipping bulk goods by sea is one of the cheapest options with few alternatives, and, moreover, freight rates constitute a relatively small fraction of the overall costs that shippers face (Stopford, 2009).

The elasticity of supply of shipping services in the short run depends on the

[8] Notably, Kalouptsidi (2014), Greenwood and Hanson (2015) and Franks et al. (2021b).

[9] Deadweight tonnes - a standard measure of ship size according to how much weight it can carry.

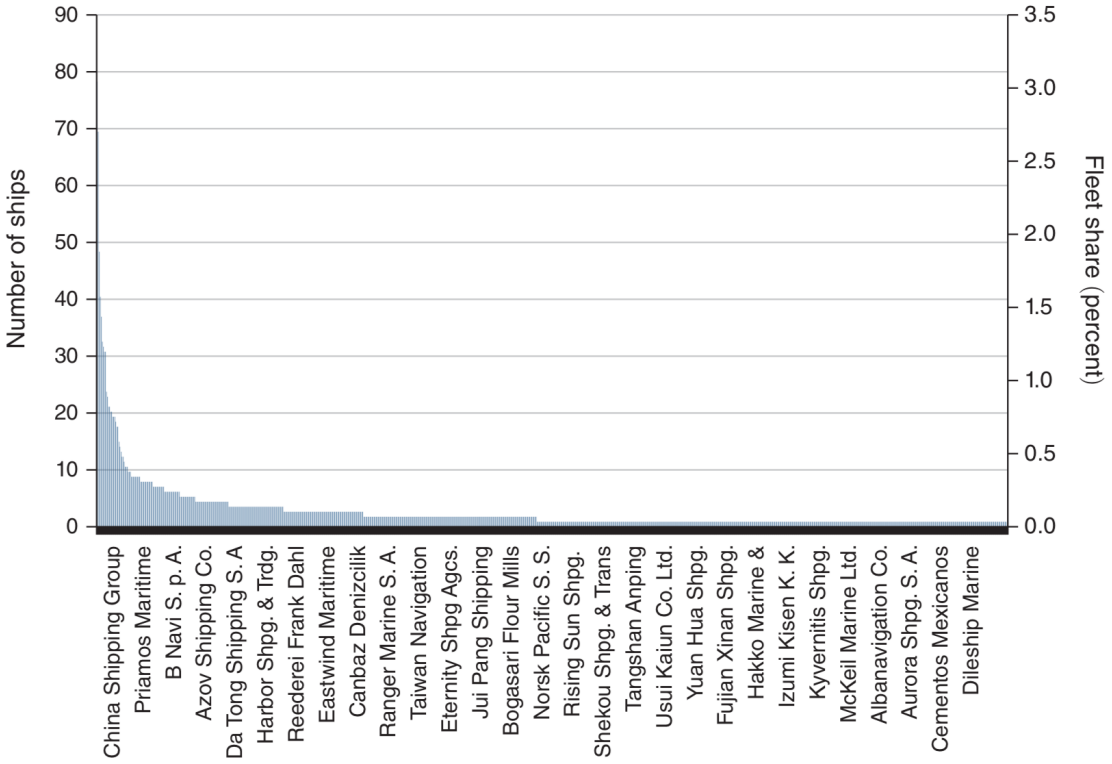


Figure 2.1: Fleet Shares of Dry Bulk Shipping Firms (Handysize)
Source: Kalouptsidi (2014), p. 568

aggregate conditions. Roughly, it can be in one of the two regimes: (i) very inelastic during normal times, when most of the vessels are in active operation, and (ii) very elastic during industry recessions, when freight rates are low and a substantial fraction of ships may be laid up (Evans, 1988). In the long run, on the other hand, the world fleet is determined by deliveries and scrappages. Since it takes one to four years to build a ship, and the average economic life of a ship is around 25 years, the fleet changes only slowly in response to developments in demand (Stopford, 2009).

The combination of often inelastic supply and inelastic demand for shipping services should make it hardly surprising that freight rates and earnings are highly volatile: even relatively modest changes in demand for shipping cargo given practically fixed fleet can lead to large changes in rates.

2.2.2 Data

The data on ship prices, earnings, and investment comes from Clarksons Research, one of the leading providers of intelligence on global shipping. The sample runs from January 1976 until December 2020. For the main results, I use data on 5 year old 76k DWT Panamax-class vessels[10] – mid-size bulk carriers typically considered representative of the broader fleet.[11] The data includes:

- Price P_t (real) – second hand prices of 5 year old ships, expressed in constant 2011 dollars using the CPI index,[12] available monthly.

- Net Earnings Π_t (real) – earnings per ship net of depreciation, in constant 2011 dollars. This is a monthly series of annual real earnings accruing to an owner of a Panamax class ship who leases their vessel for the next 12 months via a time charter agreement. The owner typically provides crew and maintenance, while the charterer bears all the remaining costs, including fuel. Accounting for the fact that a ship spends an average of 8 days per year on maintenance, I follow Greenwood and Hanson (2015) and compute annual earnings as:

$$\Pi_t = 357 \cdot DailyLeaseRate_t - 365 \cdot DailyCrewCost_t - Depreciation_t. \quad (2.1)$$

  Here $DailyLeaseRate_t$ is the 12-months time charter rate in month t, expressed in 2011 dollars using the CPI, $DailyCrewCost_t$ is estimated to be around $\$6,000$ in 2011 dollars, and depreciation rate is set at 4% of the value of the ship, in line with the linear depreciation method typically employed

[10] Details on the provided time series reveal that throughout history the series for Panamax class vessels were based on ships of slightly different sizes, ranging from 60 to 76 k DWT. To be consistent, and following Greenwood and Hanson (2015), I adjust observations where size of the ship is $S_t < 76$ by $76/S_t$. This adjustment, however, does not affect any of the main conclusions.

[11] As I verify in the Appendix, the results equally apply to smaller Handysize and larger Capesize ships.

[12] Retrieved from FRED and provided by the US Bureau of Labour and Statistics.

in shipping.

- Investment I_t measures aggregate investment decisions of the shipping industry participants over the past year, relative to the current size of the fleet.[13] Specifically:

$$I_t = \frac{\sum_{\tau=t-11}^{t} \left(NetContracting_\tau - Demolitions_\tau + Additions_\tau - Removals_\tau \right)}{Fleet_t}, \tag{2.2}$$

where $NetContracting_\tau$ is the new orders placed for newbuilt ships in month τ net of cancellations of previously made orders, while $Demolitions_\tau$ represents scrapping of the old fleet in month τ. Demolitions are typically timed according to market conditions, e.g. can be postponed during booms or brought forward during downturns, and hence constitute part of active investment decisions. $Additions_\tau$ and $Removals_\tau$ represent rare conversions of ships into and out of the bulkcarrier fleet, while $Fleet_t$ is the total size of the global bulk cargo fleet in month t. In line with Greenwood and Hanson (2015), I focus on the smoothed 12-month investment measure due to high volatility of $NetContracting_\tau$ and $Demolitions_\tau$ on the monthly basis.

All variables in (2.2) are expressed in deadweight tonnes (DWT) and are directly available from Clarksons, with the exception of $NetContracting_\tau$. The latter is constructed from the available monthly ship orderbook data as:

$$\begin{aligned} NetContracting_\tau &= NewOrders_\tau - Cancellations_\tau \\ &= Orderbook_{\tau+1} - (Orderbook_\tau - Deliveries_\tau), \end{aligned} \tag{2.3}$$

where $Orderbook_\tau$ is the total outstanding orders for new ships in month τ, and $Deliveries_\tau$ is newbuilt ships delivered in month τ, both measured

[13] Note that the investment measure is for the entire bulkcarrier fleet, not just Panamax ships.

in DWT. Unfortunately, as the orderbook data is only available going back to January 1996, my investment series is significantly shorter compared to other variables.[14]

- rx_{t+K} – cumulative log excess return on ships over the following K months, $K \in \{12, 24, 36\}$. To be precise, this is the excess return that accrues to an owner who leases their ship for K months by signing sequential 12-months time charter contracts every year.[15]

  Following the literature, I first compute the nominal 12-months holding period return as

$$r_{t+12} = \frac{P^N_{t+12} + \Pi^N_t}{P^N_t} - 1, \tag{2.4}$$

  using nominal ship prices and earnings (nominal values denoted by the superscript *'N'*). I then compute the 12-month log excess return as

$$rx_{t+12} = \ln(1 + r_{t+12}) - \ln(1 + r^f_{t+12}), \tag{2.5}$$

  where r^f_{t+12} is the cumulative Treasury Bill rate from Ken French's database.[16] Given the assumption that a ship owner signs a new 12-months time charter contract each year, I can simply sum up 12-months excess returns to compute 2 and 3 years cumulative excess returns rx_{t+24} and rx_{t+36}.

Table 2.1 summarizes the shipping data. In addition, as an indicator of global

[14] While data on ex-post fleet delivery realisations is available going back to 1976, it is difficult to back out at what time the orders for these ships were made. This is due a substantial and uncertain time to build: it can take 1 to 4 years to build a ship, and there are often delays past contracted dates. Even average time to build is hardly constant, and, unsurprisingly, increases with the amount of outstanding orders. For these reasons, I chose to stick with a shorter investment series which best captures the firm's investment decisions at time t.

[15] Owners may also choose to operate ships themselves and earn spot freight rates instead. Stopford (2009) demonstrates that spot trading and time chartering remain two important competing strategies of employing ships, between which firms actively choose.

[16] By subtracting the nominal risk-free interest rate from the nominal returns on ships, I also account for (possibly predictable) effects of inflation.

Table 2.1: Summary statistics

Variable	T	Mean	Median	Std. dev.	Min.	Max.	ϕ_1	ϕ_{12}	ϕ_{24}	ADF test
Net Earnings Π_t	540	3.67	2.32	4.25	-1.03	26.91	0.97	0.44	0.17	0.02
Price P_t	540	33.92	31.68	16.71	12.36	103.23	0.98	0.59	0.31	0.07
Investment I_t	289	0.069	0.040	0.098	-0.040	0.423	0.99	0.52	0.20	0.02
rx_{t+12}	528	0.057	0.051	0.302	-0.686	0.862	0.96	0.08	-0.15	<0.01
rx_{t+24}	516	0.121	0.060	0.446	-1.154	1.201	0.98	0.46	-0.05	<0.01
rx_{t+36}	504	0.187	0.075	0.529	-0.891	1.442	0.99	0.71	0.33	0.31

Notes. T is the number of observations. ϕ_k is the lag k autocorrelation. ADF test contains p-values from the Augmented Dickey-Fuller Test (alternative hypothesis: stationary time series).

recessions in the following analysis I use

- OECD Global Recession $\mathbb{1}_t^R$ – a composite recession indicator for OECD and Non-member economies[17] provided by FRED and based on OECD business cycle turning points developed as part of the Composite Leading Indicators (CLI) project. The OECD CLI system is based on the "growth cycle" approach, where business cycles and turning points are measured and identified in the deviation-from-trend series. Gross Domestic Product (GDP) is used as the reference for identification of turning points in the growth cycle for all countries, except for China for which the OECD relies on the value added of industry at 1995 prices. The turning point detection algorithm is based on the Bry and Boschan routine (OECD, 2019). This is one of the most comprehensive global recession indicators available.

2.2.3 Time series properties of earnings, prices, and returns

Figure 2.2 shows that ship earnings and prices are highly volatile and closely track each other. Both series also appear to be strongly mean-reverting. For example, while the 1-month autocorrelation of the earning series is 0.97, at 12 and 24

[17]The full list of countries covered is: Australia, Austria, Belgium, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States, Brazil, China, India, Indonesia, Russia, South Africa.

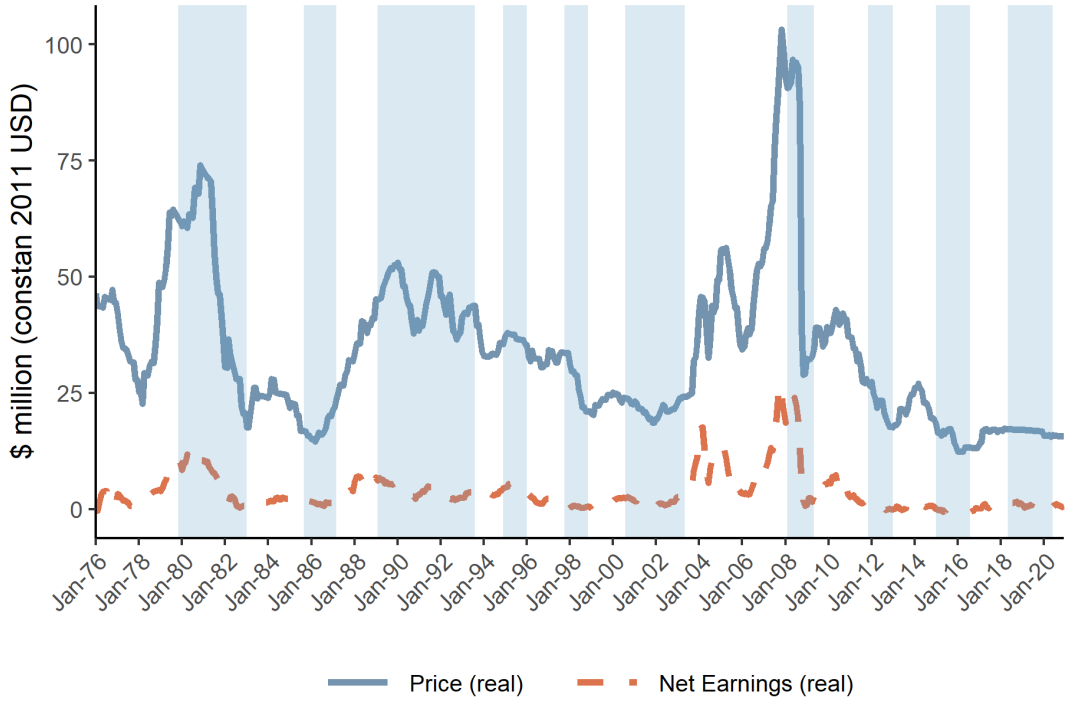


Figure 2.2: Real Net Earnings per Ship and Ship Prices for Panamax Vessels

Notes. In this and subsequent figures light blue bars indicate OECD global recessions.

month horizons it is only 0.44 and 0.17, respectively. Industry investment shown in Figure 2.3 follows a similar pattern. Lastly, Figure 2.4 shows 12 and 24 months cumulative log excess returns on owning ships. Again, cumulative returns are highly volatile series, and, while persistent from month to month, they strongly mean revert within 2 to 3 years.

It should not be surprising that ship earnings and prices are highly autocorrelated at the monthly frequency: there is simply not much that firms can do to respond to demand shocks, given the long time to build. However, the high degree of competition ensures that within 2 or 3 years shipping capacity adjusts sufficiently to revert earnings per ship back to the competitive level. Assuming a stable long-run equilibrium, we should thus expect real earnings per ship and ship prices to be stationary series. This logic remains valid even if the demand process for shipping services is a random walk.

In agreement with the above argument, the ADF test reported in Table 2.1

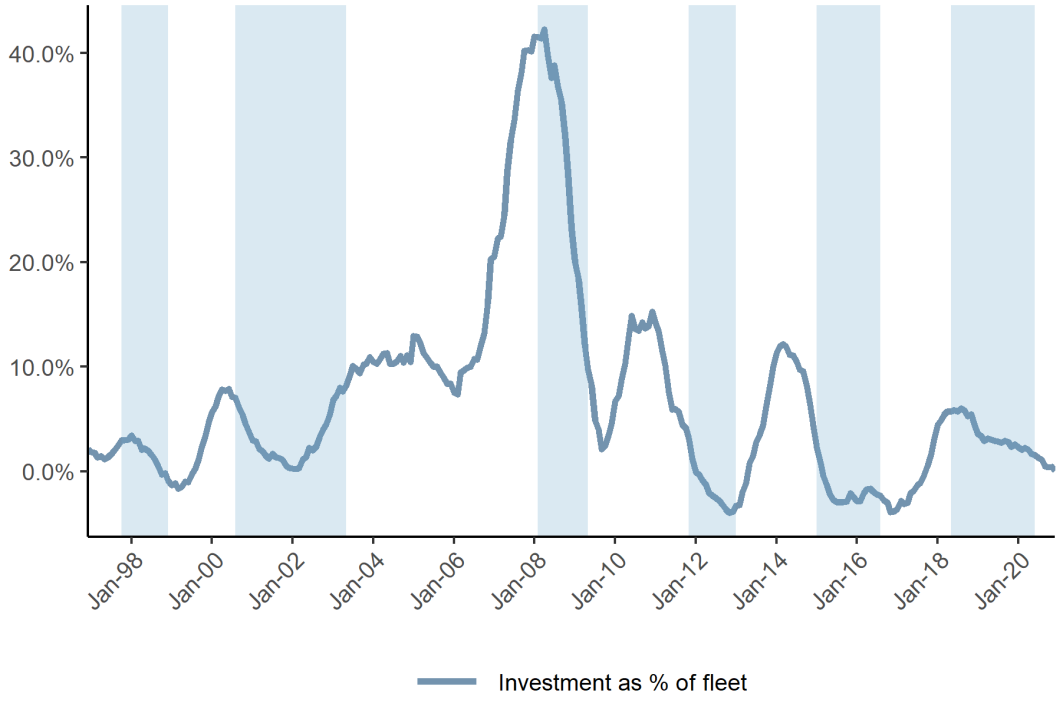


Figure 2.3: Investment in Bulkcarrier Fleet



Figure 2.4: Excess Nominal Returns on Shipping

rejects a unit root in real ship prices at the 10% level, and in net earnings per ship and investment at the 5% level. We can also reject a unit root in 12 and 24 months cumulative excess returns at less than the 1% significance level. Although the ADF test does not reject presence of a unit root in 36 months cumulative returns, this is likely due to the high level of autocorrelation induced by the overlapping nature of the observations.[18]

2.3 Countercyclical predictability of shipping returns

We now turn to investigating the relationship between current ship earnings, prices, and investment, and future excess returns. In contrast to previous literature, I find only weak evidence that the returns in shipping are predictable in the whole sample. However, I show that predictability appears strongly during global recessions, and is not present during expansions.

2.3.1 Return predictability in simple univariate regressions

Previous literature (Greenwood and Hanson (2015)) documents that variables like earnings, second hand ship prices, and industry investment all negatively forecast returns on owning ships over the following 1 to 3 years in univariate linear predictive regressions of the form

$$rx_{t+K} = a + bX_t + v_{t+K} \tag{2.6}$$

[18]For the ADF test I use the default number of lags in the R *tseries* package: $\left\lfloor (T-1)^{\frac{1}{3}} \right\rfloor$. It is known, however, that lag selection can affect the outcome of the test, and that it may suffer from potentially large size distortions. To check the robustness of the result, I also employ state-of-the-art bootstrap model averaging unit root inference by Hansen and Racine (2018). They demonstrate that an automatic model averaging procedure has the lowest size distortions among its peers, while maintaining superior power. This HR test rejects the unit root hypothesis for all series except investment at least at the 10% level.

where rx_{t+K} is K-months cumulative log excess return between t and $t+K$, X_t is a predictive regressor, and v_{t+K} is the error term. Previously reported predictability is both statistically significant and economically large.

The overlapping nature of the monthly observations on returns means that in univariate predictive regressions of the form (2.6) one would expect residuals to be highly autocorrelated. Even with non-overlapping data, one would expect residuals to be autocorrelated if returns are indeed predictable. This creates problems when testing $H_0 : b = 0$, since the recent research shows that even tests based on Newey-West and similar HAC covariance matrix estimators are usually severely oversized and tend to over-reject the null (see, for example, Harri and Brorsen (2009)). While there has been great theoretical progress in heteroscedasticity- and autocorrelation- robust inference starting with seminal papers by Kiefer et al. (2000) and Kiefer and Vogelsang (2002), and with recent important contributions by Ye and Sun (2018) and Lazarus et al. (2018), the applied work has not generally kept up with these new improved methods.

Accounting for these concerns, I estimate linear regressions (2.6) using OLS with state-of-the-art Heteroscedasticity and Autocorrelation Robust (HAR) inference developed by Ye and Sun (2018). I report tests using the orthonormal series (OS) HAR variance estimator that has been shown to have competitive power compared to the alternatives, while maintaining the advantage of asymptotically having a standard F distribution (Lazarus et al., 2018).

Table 2.2 presents the estimates of slope coefficients when using, in turn, ship earnings, prices and investments as predictive regressor to forecast cumulative excess returns at 12, 24, and 36 months horizons. The first observation is that point estimates for earnings and prices are substantially smaller in absolute value compared to those found by Greenwood and Hanson (2015). For example, the estimated coefficients for net earnings are less than half in absolute value. Secondly, the strength of statistical evidence for predictability in the whole sample also

Table 2.2: Return Predictability in Shipping: Simple Regressions

	X = net real earnings Π			X = used ship price P			X = investment I		
K	12m	24m	36m	12m	24m	36m	12m	24m	36m
b	-0.007	-0.019	-0.018	-0.004*	-0.009+	-0.011	-0.406	-0.762*	-0.557
	(0.007)	(0.012)	(0.021)	(0.001)	(0.004)	(0.007)	(0.313)	(0.266)	(0.479)
	[0.390]	[0.206]	[0.429]	[0.011]	[0.088]	[0.165]	[0.264]	[0.046]	[0.309]
T obs.	528	516	504	528	516	504	277	265	253
R^2	0.009	0.032	0.023	0.061	0.113	0.124	0.018	0.032	0.013
Adj. R^2	0.007	0.030	0.021	0.059	0.112	0.123	0.015	0.029	0.009

Notes. This table reports the estimates of slope coefficients in univariate time series regressions of the form $rx_{t+K} = a + bX_t + v_{t+K}$ where X_t is one of the forecasting variables, K is the forecasting horizon in months, and rx_{t+K} is cumulative log excess return between months t and $t+K$. Standard errors are reported in parentheses, and p-values are reported in brackets, all obtained using OS HAR inference. $^{+}$,*,**, and *** indicate significance at the 0.1, 0.05, 0.01, and 0.001 levels, respectively.

appears substantially weaker than previously reported. We reject the null of no predictability only in three out of nine cases: prices appear to negatively forecast future returns at 1 and 2 year horizons, at the 5% and 10% significance levels respectively, and investment is negatively associated with 2-year ahead returns at the 5% significance level. We do not reject the null of no predictability of returns by net earnings at any horizon at any conventional significance level.[19]

To sum up, while there is still some evidence of shipping return predictability in my sample, in particular by ship prices, it is considerably weaker than reported in previous literature. We now turn to investigate whether predictability varies systematically over the global business cycle.

2.3.2 Regime-dependent predictability

The demand for shipping services is driven by the volume of international trade, which in turn is closely related to the state of the global economy. Global economic

[19]Incidentally, the weak statistical significance does not appear to be primarily driven by the use of robust OS HAR inference. Tests using the conventional Newey-West standard errors still reject the null of no predictability only in 4 out of the 9 regressions. The difference in statistical conclusions compared to the previous literature therefore is mainly due to smaller point estimates obtained in the new dataset.

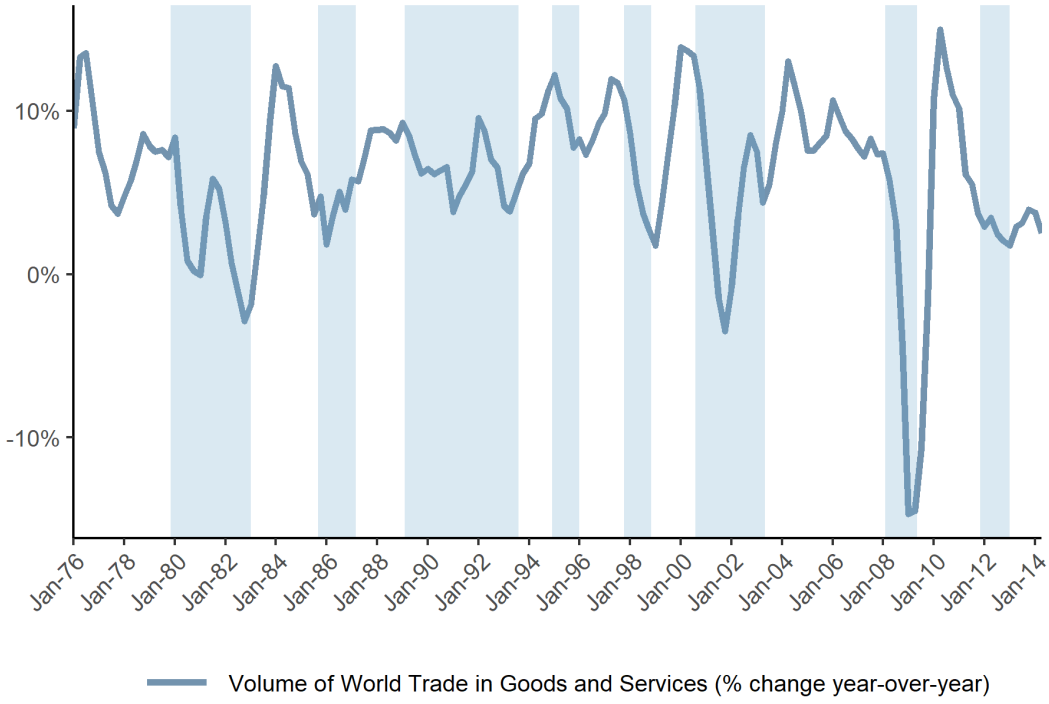


Figure 2.5: Global Recessions and International Trade

downturns are thus likely to be associated with weak shipping demand. Indeed, Figure 2.5 shows that year-over-year growth in the volume of world trade[20] usually significantly weakens during the global recessions.

Looking back at Figures 2.2, 2.3 and 2.4, it appears that ship prices, earnings and investment contract particularly sharply around global recessions, whereas forward returns tend to rise substantially. It is possible then that return predictability varies systematically as conditions faced by the shipping firms change over the business cycle, and is concentrated during recessions.

I test this by allowing the predictive relationship to change across recessions and expansions as follows:

$$rx_{t+K} = \mathbb{1}_t^R(a_1 + b_1 X_t) + (1 - \mathbb{1}_t^R)(a_2 + b_2 X_t) + v_{t+K}, \tag{2.7}$$

where $\mathbb{1}_t^R$ is the OECD based indicator of global recessions. This way, I let coeffi-

[20] Retrieved from FRED and provided by OECD (Main Economic Indicators). Series was discontinued after 2014.

Table 2.3: Regime-Dependent Return Predictability in Shipping

	X = net real earnings Π			X = used ship price P			X = investment I		
K	12m	24m	36m	12m	24m	36m	12m	24m	36m
b_1	-0.027***	-0.050	-0.062^{+}	-0.007**	-0.015^{+}	-0.020^{+}	-0.801*	-0.810^{+}	-1.268
	(0.006)	(0.026)	(0.032)	(0.002)	(0.006)	(0.008)	(0.311)	(0.408)	(0.907)
	[0.001]	[0.130]	[0.096]	[0.004]	[0.063]	[0.072]	[0.033]	[0.075]	[0.212]
b_2	0.008	0.007	0.016	-0.000	-0.001	0.000	-0.180	-0.548	0.373
	(0.010)	(0.016)	(0.019)	(0.002)	(0.004)	(0.006)	(0.393)	(0.462)	(0.656)
	[0.420]	[0.700]	[0.427]	[0.883]	[0.847]	[0.991]	[0.659]	[0.301]	[0.600]
F-test $b_1 = b_2$	0.004	0.141	0.108	0.014	0.086	0.043	0.232	0.601	0.083
T_1 obs.	237	225	217	237	225	217	115	103	95
T_2 obs.	291	291	287	291	291	287	162	162	158
R^2	0.108	0.109	0.135	0.135	0.186	0.225	0.067	0.053	0.060
Adj. R^2	0.103	0.103	0.129	0.130	0.181	0.221	0.057	0.043	0.049

Notes. This table reports the estimates of slope coefficients and F-test statistics for regime switching time series regressions of the form $rx_{t+K} = \mathbb{1}_t^R(a_1 + b_1 X_t) + (1 - \mathbb{1}_t^R)(a_2 + b_2 X_t) + v_{t+K}$, where $\mathbb{1}_t^R$ is the OECD global recession indicator, X_t is one of the forecasting variables, K is the forecasting horizon in months, and rx_{t+K} is cumulative log excess return between months t and $t+K$. Standard errors are reported in parentheses, and p-values are reported in brackets, while the F-test row reports p-value for testing $H_0 : b_1 = b_2$, all obtained using OS HAR inference. $^{+}$,*,**, and *** indicate significance at the 0.1, 0.05, 0.01, and 0.001 levels, respectively.

cients take different values depending on whether month t (the month in which we attempt to forecast returns over the following K months, using data on the predictor variable for month t) is classed as being in a recession or not; and not depending on whether the period spanned by the dependent variable is wholly or partly classed as a recession period. We are interested in whether coefficients b_1 and b_2 are systematically different across the two regimes. As before, I estimate (2.7) using OLS with OS HAR inference.

Table 2.3 presents the results. Firstly, all but two of the estimates of predictive slope coefficients in recessions (b_1) are statistically significant at least at the 10% level, while none is remotely close to being significant in expansions (b_2). There is particularly strong statistical evidence that in recession months future returns are predictable at the 12-months horizon, as coefficient estimates on net earnings and prices are significant at the 0.1% and 1% levels, respectively. The estimate

of the coefficient on investment, for which I have only about half the number of observations, is significant at the 5% level. Most coefficient estimates for predictability at 24 and 36 month horizons are significant at the 10% level.

Secondly, for the majority of these predictive regressions, we can reject the null that slope coefficients are equal across regimes $H_0 : b_1 = b_2$, notably at the 1% level for earnings at the 12 months horizon, and at the 5% level for prices at the 12 and 36 months horizons. Even when we cannot reject the null at the standard levels for earnings, the p-values are just above the 10% threshold. Although we fail to reject $H_0 : b_1 = b_2$ for investment at the 12 and 24 months horizons, it is likely because of limited power due to a smaller number of observations.

Lastly, the point estimates of slope coefficients during recessions are substantially larger in absolute value than the estimates during expansions (or, for that matter, than the estimates in the full sample univariate regressions reported in Table 2.2). The predictability during recessions is economically large: for example, a one standard deviation fall in net real earnings is associated with 11.5 and 21.3 percentage point increases in the expected cumulative excess returns over the following 1 and 2 years, respectively. Similarly, a one standard deviation decline in real ship prices during recessions predicts an 11.7 percentage point increase in the subsequent 12 months excess returns, and a 25.1 percentage point increase in excess returns over the following 24 months.

The presented empirical evidence thus indicates that return predictability in the shipping industry is both statistically significant and economically large during global recessions, but not at other times.

2.3.3 Robustness

The Appendix provides a detailed description of several important robustness checks and extensions. Here I provide a brief summary.

First, I verify that the results are similar for smaller Handysize and larger

Capesize vessels, and so are indeed representative of the broader shipping industry (see Table 2.A.1 in Appendix 2.A). Second, I show that using an alternative and more conservative measure of recessions, namely the dates set by NBER for the US, makes no material difference to the finding that the returns are strongly predictable in downturns, but not in expansions (see Table 2.B.1 in Appendix 2.B).

Third, we can see from Figure 2.2 that the shipping industry has experienced a super boom-bust cycle during 2006-2009. This raises a concern that my main results could be driven by the Great Recession. I show that the results are similar when I only include the data available until 2006 in the estimation sample (see Table 2.C.1 in Appendix 2.C).

Fourth, I address the potential concern that my results could be driven by an upward bias in the measure of earnings, and therefore returns, during recessions. The bias could arise because some ships are likely to end up idle during downturns, as their owners may be unable to employ them. In that case, my measure of returns, which assumes ships to always be chartered away, will overstate the true returns to owning ships in recessions. Moreover, the number of idle ships, and therefore the size of this bias, is likely to increase when the market conditions deteriorate, potentially generating an impression of regime-dependent predictability. To avoid the bias, I estimate the predictive regressions using excess capital gains (net of depreciation) as the dependent variable, i.e. effectively setting gross ship earnings to zero (see Table 2.D.1 in Appendix 2.D).[21] The results are very similar to those reported in Table 2.3, hence it is unlikely that they are driven by a potential bias in the measure of returns.

Fifth, to overcome the concern that my findings could be driven by the overlapping nature of returns and artificially high degree of autocorrelation, I estimate (2.7) using only annually sampled data, and again get results fully consistent with those in Table 2.3 (see Table 2.E.1 in Appendix 2.E).

[21] In other words, I set $DailyLeaseRate_t$ and $DailyCrewCost_t$ to zero in (the nominal version of) (2.1), and then use this amended measure of earnings to compute returns.

Sixth, because the regressors I use are highly persistent at the monthly frequency, the results could be driven by the Stambaugh bias (Stambaugh, 1999), or near-unit-root inference. I employ the IVX procedure developed by Magdalinos and Phillips (2011) for econometric inference in the vicinity of unity, and adapted by Kostakis et al. (2015) to the context of return predictability. I use this procedure to test for predictability of shipping net excess capital gains and excess returns in recessions and expansions, and again find strong evidence of countercyclical return predictability (see Table 2.F.1 in Appendix 2.F).

Seventh, some of the results could potentially be driven by correlated and persistent measurement errors in the second-hand ship prices and earnings series. I address this by further lagging the regressors used in the predictive regressions. While this weakens the power of the statistical evidence for both return predictability and predictability being stronger during recessions, the results remain qualitatively similar (see Table 2.G.1 in Appendix 2.G).

Lastly, I verify that the main results reported in Table 2.3 are not the artefact of the adjustments made to the raw time series (see Table 2.H.1 in Appendix 2.H).

2.3.4 Risk-based explanations of return predictability

The default explanation of the documented return predictability by most financial economists would likely be that the required rate of return changes over time. The central finding that return predictability emerges strongly during recessions squares particularly well with the intuition that risk premia increase during downturns due to increased uncertainty. However, as I demonstrate below, the risk-based explanation is not supported by the data.

Following Greenwood and Hanson (2015), one can decompose the rate of return on shipping required by diversified investors as a product of three components: (1) the hedging value of shipping investments, i.e. the correlation between shipping returns and the welfare of a representative investor, (2) future shipping

risk, and (3) the economy-wide risk premium.

First, it is difficult to imagine that time variation in the hedge value of ships could account for the magnitude of documented return predictability in recessions. For that to be the case, performance of shipping investments made at the bottom of the cycle, i.e. when ship prices, earnings, and investment are low, must be particularly strongly linked to future developments in the rest of the economy. To investigate whether this is likely the case, I divide observations in recession months into two groups: when ship prices are high and low, or, more precisely, are above and below the median real ship price we observe in recessions. The notion that, during recessions, ships have particularly low hedge value at the bottom of the cycle is in sharp contrast with what we see in the data: the correlation between the future 12-months excess returns on ships and the 12-months excess stock market returns[22] is only 0.2 when prices are low in recessions, but 0.7 when ship prices are high in recession periods.[23] We see a similar picture when splitting the recession months observations by high/low ship earnings, or by high/low investment, and also when looking at the longer future horizons. Moreover, the pattern remains the same when instead of excess stock returns I use future world industrial production growth.[24] Therefore, shipping investments appear to have particularly low hedge value when ship prices, earnings, and investment are high when entering recessions, not at the bottom of the cycle. This is the exact opposite of what we would expect to observe if the predictability of shipping returns in recessions was related to time variation in the hedge value of ships.

To test the hypothesis that future shipping risk (and thus the required rate of

[22]Specifically, I use cumulative Mkt-Rf from Ken French's database.

[23]Noting that ship prices, earnings, and investment tend to be high at the start of the recessions, and fall during their course, I also repeat this analysis by dividing each recession into equal first and second halves. Correlations between the future shipping and market returns are substantially lower in the second halves of the recessions.

[24]Hamilton (2021) shows that measures based on world industrial production (specifically, OECD + 6 non-member economies) provide a good indicator of global economic activity. The data is from James Hamilton's website.

return) is negatively associated with current ship prices, earnings, and investment particularly during recessions, I regress a measure of future shipping volatility on the current values of predictive regressors, allowing coefficients to differ between recessions and expansions. I estimate the following regime-dependent regression:

$$\sigma_{t+K} = \mathbb{1}_t^R(\alpha_1 + \gamma_1 X_t) + (1 - \mathbb{1}_t^R)(\alpha_2 + \gamma_2 X_t) + \varepsilon_{t+K}, \tag{2.8}$$

where on the left hand side I use the standard deviation of monthly ship net excess capital gains[25] over the following K-months window. Table 2.4 presents the results. Firstly, volatility of capital gains appears to be more predictable in expansions, rather than recessions. Secondly, all the slope coefficients are positive across both the regimes, indicating that low ship prices, net earnings, and investment forecast low future volatility, in stark contrast to what we would expect if the return predictability during recessions was linked to shipping risk. I obtain similar results when using future K-months volatility of ship earnings, prices, or approximated 1-month excess returns[26] on the left hand side.

Finally, following the previous literature, I investigate whether controlling for variation over time in the economy-wide risk premium impacts the ability of earnings and asset prices to forecast future returns in dry bulk shipping during recessions. Specifically, I estimate predictive regressions of the following form:

$$rx_{t+K} = \mathbb{1}_t^R(a_1 + b_1 X_t) + (1 - \mathbb{1}_t^R)(a_2 + b_2 X_t) + \beta \mathbf{C_t} + v_{t+K}, \tag{2.9}$$

where $\mathbf{C_t}$ is a vector of controls intended to proxy for the economy-wide risk

[25] One month net excess capital gain is defined as $\Delta px_{t+1} = \ln\left(\frac{P_{t+1}^N - Depreciation_t^N}{P_t^N}\right) - \ln(1 + r_t^f)$, i.e. net of depreciation and in excess of the risk-free rate.

[26] Because the excess returns data is based on the 12-months earnings derived from the 12-months time charter rates, I can only approximate 1-month ahead excess returns as $rx_{t+1} = \ln\left(\frac{P_{t+1}^N + \Pi_t^N/12}{P_t^N}\right) - \ln(1 + r_t^f)$. The correlation between this measure of monthly excess returns and the net excess capital gains used in the main specification is 0.997, highlighting that ship returns are to a large extent driven by capital gains.

Table 2.4: Forecasting Capital Gains Volatility

	X = net real earnings Π			X = used ship price P			X = investment I		
K	12m	24m	36m	12m	24m	36m	12m	24m	36m
γ_1	0.007^{*}	0.002	0.002	0.001	0.000	0.000	0.442^{***}	0.013	0.010
	(0.003)	(0.002)	(0.002)	(0.001)	(0.000)	(0.000)	(0.063)	(0.020)	(0.063)
	[0.033]	[0.309]	[0.497]	[0.128]	[0.384]	[0.560]	[0.000]	[0.537]	[0.878]
γ_2	0.004^{**}	0.005^{*}	0.001	0.001^{+}	0.002^{**}	0.001^{*}	0.263^{***}	0.434^{***}	0.095
	(0.001)	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.062)	(0.084)	(0.103)
	[0.002]	[0.037]	[0.136]	[0.085]	[0.004]	[0.034]	[0.001]	[0.001]	[0.407]
F-test $\gamma_1 = \gamma_2$	0.095	0.277	0.927	0.216	0.056	0.267	0.004	0.001	0.397
T_1 obs.	237	225	217	237	225	217	115	103	95
T_2 obs.	291	291	287	291	291	287	162	162	158
R^2	0.397	0.195	0.032	0.290	0.257	0.065	0.570	0.407	0.028
Adj. R^2	0.394	0.190	0.027	0.286	0.253	0.059	0.565	0.401	0.016

Notes. This table reports the estimates of slope coefficients and F-test statistics for regime switching time series regressions of the form $\sigma_{t+K} = \mathbb{1}_t^R(\alpha_1 + \gamma_1 X_t) + (1 - \mathbb{1}_t^R)(\alpha_2 + \gamma_2 X_t) + \varepsilon_{t+K}$, where $\mathbb{1}_t^R$ is the OECD global recession indicator, X_t is one of the forecasting variables, K is the forecasting horizon in months, and σ_{t+K} is the volatility of monthly ship net excess capital gains over the period from t to $t+K$. Standard errors are reported in parentheses, and p-values are reported in brackets, while the F-test row reports p-value for testing $H_0 : \gamma_1 = \gamma_2$, all obtained using OS HAR inference. $^{+}$,*,**, and *** indicate significance at the 0.1, 0.05, 0.01, and 0.001 levels, respectively.

premium at time t. Specifically, I include the smoothed earnings yield on S&P500 stocks from Bob Shiller's database, the credit spread between Moody's BAA and AAA bond yields, and the term spread on US government bonds.[27]

The results in Table 2.5 show that, if anything, net earnings, ship prices and investment appear to be *stronger* predictors of future excess returns during recessions when macro risk premium controls are included in the regressions. In addition, most of the controls have individually insignificant estimated coefficients, with the exception of the credit spread when investment is used as the predictor. When testing their joint significance, we cannot reject the null that coefficients on all controls are equal to zero in eight out of nine regressions. This evidence is inconsistent with the hypothesis that ship earnings and prices somehow proxy for the economy-wide risk premium in the results reported in Table

[27] Specifically, the difference between the yields on 10-year and 1-year zero coupon Treasury bonds from Gürkaynak et al. (2007).

Table 2.5: Controlling for Economy-Wide Risk Premia

	X = net real earnings Π			X = used ship price P			X = investment I		
K	12m	24m	36m	12m	24m	36m	12m	24m	36m
b_1	-0.027***	-0.048*	-0.056*	-0.008***	-0.015*	-0.019*	-1.352***	-1.473**	-1.348*
	(0.005)	(0.018)	(0.022)	(0.001)	(0.005)	(0.005)	(0.311)	(0.310)	(0.550)
	[0.000]	[0.034]	[0.042]	[0.000]	[0.031]	[0.014]	[0.001]	[0.003]	[0.050]
b_2	0.008	0.008	0.019	-0.000	-0.000	0.002	-0.312	-0.613	0.871
	(0.011)	(0.018)	(0.016)	(0.002)	(0.004)	(0.006)	(0.439)	(0.733)	(1.049)
	[0.466]	[0.677]	[0.280]	[0.829]	[0.918]	[0.777]	[0.497]	[0.435]	[0.438]
Stocks $\frac{E_{10}}{P}$	0.000	0.870	1.815	0.262	1.449	2.269	-1.701	-9.183	-37.059
	(1.557)	(4.149)	(3.510)	(2.141)	(4.501)	(3.829)	(10.379)	(14.092)	(18.560)
	[1.000]	[0.844]	[0.632]	[0.909]	[0.764]	[0.585]	[0.875]	[0.550]	[0.117]
Credit spread	-0.378	-6.985	-17.191	-0.786	-7.979	-17.841	30.085**	45.732+	50.384+
	(10.923)	(23.835)	(20.087)	(11.866)	(24.491)	(17.797)	(8.304)	(21.075)	(21.480)
	[0.973]	[0.777]	[0.417]	[0.948]	[0.756]	[0.355]	[0.003]	[0.062]	[0.057]
Term spread	-0.531	2.956	9.436	-0.826	2.179	8.176	-2.183	-0.295	14.395
	(3.997)	(7.744)	(9.745)	(4.097)	(7.372)	(8.876)	(8.992)	(13.831)	(14.776)
	[0.899]	[0.722]	[0.388]	[0.847]	[0.782]	[0.409]	[0.820]	[0.984]	[0.385]
F-test $b_1 = b_2$	0.002	0.126	0.036	0.009	0.075	0.017	0.043	0.348	0.059
F-test controls	1.000	0.984	0.771	0.996	0.985	0.765	0.073	0.254	0.328
T_1 obs.	237	225	217	237	225	217	115	103	95
T_2 obs.	291	291	287	291	291	287	162	162	158
R^2	0.108	0.115	0.180	0.137	0.192	0.259	0.159	0.129	0.190
Adj. R^2	0.098	0.105	0.170	0.127	0.182	0.250	0.141	0.109	0.170

Notes. This table reports the estimates of slope coefficients and F-test statistics for regime switching time series regressions of the form $rx_{t+K} = \mathbb{1}_t^R(a_1 + b_1 X_t) + (1 - \mathbb{1}_t^R)(a_2 + b_2 X_t) + \beta \mathbf{C_t} + v_{t+K}$, where $\mathbb{1}_t^R$ is the OECD global recession indicator, X_t is one of the forecasting variables, K is the forecasting horizon in months, rx_{t+K} is cumulative log excess return between months t and $t+K$, and $\beta \mathbf{C_t}$ is a vector of macro risk premium controls. Standard errors are reported in parentheses, and p-values are reported in brackets, all obtained using OS HAR inference. The first F-test row reports p-values for testing $H_0 : b_1 = b_2$, while the second reports p-values for testing the joint significance of controls. $^+$,*,**, and *** indicate significance at the 0.1, 0.05, 0.01, and 0.001 levels, respectively.

2.3.[28]

To sum up, the empirical evidence does not favour the risk-based explanation that excess returns are predictable in recessions due to variation over time in the required rate of return.

[28] I obtain similar results when allowing the controls to have different effects during recessions and expansions, i.e. adding interaction terms between $\mathbb{1}_t^R$ and the variables in $\mathbf{C_t}$.

2.4 Liquidity constraints and industry slumps during recessions

What can account for the countercyclical return predictability documented in the previous section, if not time variation in the required rate of return? I argue that liquidity constraints faced by the industry participants during recessions result in the cash-in-the-market pricing for assets sold by distressed firms, as well as slumps in the industry investment. Those who manage to buy assets in this climate end up earning high excess returns, which are larger the deeper the recession is.

2.4.1 Financial frictions and return predictability in dry bulk shipping

Greenwood and Hanson (2015) suggest a behavioural explanation for the return predictability in competitive, capital-intensive industries like shipping. According to their theory, firms tend to over-extrapolate current demand conditions, as well as underestimate investment responses of their competitors. Biased expectations of firms about future conditions lead to excess volatility in ship prices, as well as industry overinvestment in new ships during booms and underinvestment in downturns. Ship earnings, prices, and industry investment should then negatively forecast future returns. While this explanation has an intuitive appeal and is echoed in many narrative accounts of shipping cycles, it is not consistent with my finding that the return predictability emerges only during recessions.

I argue that the observed counter-cyclical predictability arises due to liquidity constraints faced by shipping firms and the industry overall. When earnings fall and cash flows dry up in recessions, many firms are forced to sell their assets in order to repay outstanding debts. At the same time, the potential buyers, – mostly other firms in the industry, – are also likely to be tight on cash, and may find

it difficult to borrow funds. Their ability to absorb second-hand ships from the distressed firms is therefore limited: prices of second-hand ships have to fall below their fundamental value to equilibrate the distressed supply with the constrained demand. The worse is the recession, the more constrained the buyers are, and so the deeper prices fall, resulting in higher excess returns to the firms that manage to buy ships.

In his book, Martin Stopford corroborates the conjecture of financial troubles and forced selling during recessions:

> Obviously selling a ship at the bottom of a market cycle is disastrous for its owner and a great bargain for the buyer. No shipping company follows this suicidal course of action by choice. 'Distress' sales during market troughs are generally forced on companies by cashflow pressures such as bunker bills or a banker who has foreclosed and taken possession of the fleet. (Stopford, 2009, p. 203)

This view is also supported by industry practitioners. For example, during the 2015-2016 slump in dry bulk shipping, one chief executive commented to the *Financial Times:* "Everybody who's selling, especially good-quality assets, these days they really sell them because they need to [bring in cash]" (Wright, 2016). Similarly, writing about the events of the 2008 Financial Crisis in his blog, the prominent ship broker and finance advisor Basil Karatzas notes:

> There were so many ships to be bought cheaply from distressed shipowners, from distressed shipping banks, from bankrupt KG funds in Germany, from shipbuilders that got stuck with contracts in default and ships at their slipways. (Karatzas, 2017)

On the other hand, lack of financing and competition on the constrained demand side is often cited as the primary reason for why assets sell at such depressed prices during recessions. Discussing financial distress in the shipping sector in 2012, prominent maritime expert Randee Day said that "banks have been unwilling to finance the purchase of second-hand vessels, which is creating a stalled sale and purchase market", adding that "this has and will continue to put tremendous pressure on second hand values" (Financier Worldwide, 2012).

Despite a wave of private equity money flowing into shipping in the aftermath of the Financial Crisis, mainly attracted by rock-bottom asset prices, one shipping banker interviewed by the *Financial Times* had estimated that the industry was still facing a $100bn credit gap at the time (Odell and Makan, 2013). Basil Karatzas blames the "complete dearth of debt financing" for the free fall in vessel values in 2016 (Karatzas, 2016), and writes:

> In 2016, there were days that good quality dry bulk ships for sale could not get an offer or an interested party to send an inspection for a pre-purchase inspection. Ten-year old dry bulk vessels in 2016 were selling at a multiple of their scrap value (Karatzas, 2017).

"The big issue is finance. It's tougher than ever to raise bank debt" succinctly put it one manager of a shipping company in a comment to *Reuters* regarding low ship prices, even as the industry was already starting to show signs of recovery in 2017 (Wallis, 2017).

Financing woes in shipping during downturns is not a uniquely post-2008 phenomenon either. Stopford (2009) gives an example of Bulk Transport, a self-liquidating ship fund that successfully purchased ships at distressed prices during the mid-1980s industry slump, when the traditional sources of equity and debt financing became mostly unavailable to shipping companies. The scheme proved very successful, as ships appreciated four times their purchase price in the following four years.

It is only natural to expect financial constraints to play a critical role in an industry like dry bulk shipping, which, according to Stopford (2009), is dominated by many small players with specialist knowledge, often informal corporate structures, and traditional reliance on bank financing. Combined with high volatility of revenue flows and asset prices, it is not surprising that the sector is classified as "exotic" finance by Moody's. This also explains why traditional lenders to the shipping companies typically possess considerable industry expertise. For example, shipping bank loans are usually managed by a separate department,

staffed with marketing and credit officers who know and understand the business (Stopford, 2009).

For these very reasons, industry outsiders may find it difficult to jump into shipping and buy assets profitably at the bottom of the market, explaining the apparent lack of competition among ship buyers. There is indeed anecdotal evidence that while new investors are attracted to ships during distressed times, often in partnership with established players, it is often difficult for them to carry out a profitable operation. Paraphrasing eminent shipping expert and executive Randee Day, shipping is hard: inexperienced newcomers can lose money by buying ships of inferior quality, spending a lot of time and resources on negotiations and transaction fees, totally missing the cycle, and partnering up with a second-tier ship owner who would make money at their expense through 'leakage' on the cost side, and has every incentive to drag on an unprofitable joint venture in order to continue earning their management fees (Day, 2019). The *Financial Times* journalist John Dizard also highlights the classic agency problems faced by 'smart money' investors who depend on dry bulk operators. He concludes that, in shipping, "old-fashioned family dictators are able to react much more deftly than even the most educated and experienced institutional investors" (Dizard, 2015).

It is straightforward to infer the effect of financially constrained firms and depressed asset values on the industry investment in recessions. First, the lack of financing makes it difficult for the firms to order new ships: even though the full price of a newbuilt ship may not be payable until the delivery in a couple of years, shipbuilders still rely on substantial stage payments throughout the project as their working capital Stopford (2009). Second, as prices of second hand tonnage fall, firms with spare liquidity may prefer used ships to new. "It does make more financial sense to buy second-hand at the moment than order new-builds, given the very attractive prices," commented one ship broker to *Reuters* in 2016 (Wallis, 2016). The orders for new ships are therefore also likely to plunge during

recessions, negatively forecasting future excess returns, in line with my empirical results.

Let us now turn to formalizing these insights in a model of a competitive industry that suffers from liquidity constraints.

2.4.2 The model

In this section I develop a model of a financially constrained industry to formally show how return predictability arises during recessions. The intuition here is related to the idea in Shleifer and Vishny (1992), namely that: "When a firm in financial distress needs to sell assets, its industry peers are likely to be experiencing problems themselves, leading to asset sales at prices below value in best use" (p. 1343).[29] In the model, I allow for the aggregate shock to be continuous. I am then able to demonstrate explicitly that the deeper the recession is, the further below its fundamental value the equilibrium asset price falls, so the future return on assets increases.

Because of commitment problems in the model, the maximum repayment that firms can promise to creditors is limited by the market value of collateral. In a boom, many firms generate cash flows in excess of maximum repayments they could possibly promise to creditors, so they have spare liquidity they can use to purchase more assets. Even though some firms are hit by bad idiosyncratic shocks and need to sell capital, unconstrained firms have enough spare cash to be able to buy assets at the full fundamental value (and the competition between them ensures that is indeed the equilibrium price). Assets simply earn the required rate of return to their buyers, so there is no predictability.

In a recession, however, even firms with good idiosyncratic shocks have limited

[29]Unlike Shleifer and Vishny (1992), I do not consider the possibility of misallocation, i.e. assets being sold to less efficient users. This is consistent with both the anecdotal evidence in the previous section, and also with Franks et al. (2021a), who do not find evidence of misallocation during fire sales in the context of airline industry.

spare liquidity, and are unable to borrow at this point. At the same time, the amount of capital that firms with bad shocks have to sell to repay their debts increases. The worse is the aggregate shock, the more constrained the buyers are, and so the deeper asset price falls below the fundamental value in order for constrained buyers to absorb all assets from forced sales. Those who do manage to buy assets in this climate end up earning high excess returns, which are larger the deeper the recession is.

The framework

The model economy lasts for three periods $t \in \{0, 1, 2\}$. There is a unit mass of identical firms in the industry we are considering, and also a large number of potential creditors with deep pockets. All agents are risk-neutral, and do not discount time, so the required rate of return is zero.

Each firm starts with an endowment of $\$n$ in period 0, but also has access to a competitive financial market where it can borrow $\$b$ (lend if $b < 0$). Firms have a constant return to scale technology, and can invest in capital in period 0, which we assume to be homogenous and perfectly divisible. One unit of capital costs \$1, and generates non-negative cash flows in both periods 1 and 2. Firms can also store cash between periods. We assume that owners of each firm maximize period 2 funds available for their consumption.

Period 1 cash generated by each unit of capital is subject to aggregate and idiosyncratic firm shocks, both uniformly distributed on $[0, 1]$. Specifically, a firm's cash flow in period 1 is $k_1 \cdot [\tilde{Z} + \tilde{\varepsilon}]$, where k_1 is the firm's capital stock held from period 0 to period 1, $\tilde{Z} \sim \mathcal{U}(0, 1)$ is the aggregate shock common to all firms, and $\tilde{\varepsilon} \sim \mathcal{U}(0, 1)$ is the firm-specific shock.[30] Period 2 cash flow is, for simplicity, non-stochastic and fixed at \$1 per unit of capital held from period 1 to period 2.

Capital is irreversible and industry specific: only firms in the industry know

[30] Note that the resulting distribution of $\tilde{Z} + \tilde{\varepsilon}$ is symmetric triangular, with mean 1 and support $[0, 2]$.

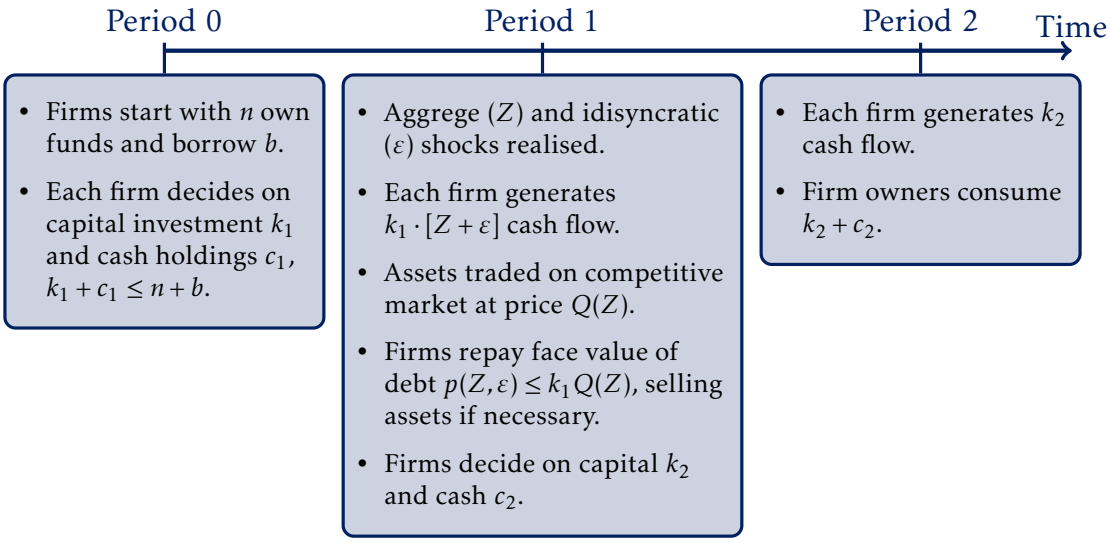


Figure 2.6: Timeline

how to operate it productively. There is a competitive secondary market for assets in period 1, in which, therefore, only firms in the industry can act as buyers. Capital becomes worthless in period 2.

Limited commitment and financial contracts

Financial contracts are subject to limited enforcement and commitment on behalf of the firms, who can default on existing liabilities specified in the contracts. In case of a default, creditors cannot enforce repayments out of firm's cash holdings, or restrict the firm from using that cash to acquire new assets. They can, however, be contractually given control rights over all or some of the firm's existing assets. In addition, the financial contract can specify the bargaining protocol following a default, in the spirit of Harris and Raviv (1995), and assign bargaining powers to the parties. There is no problem with commitment on behalf of the creditors, and contracts can, in principle, be conditioned on both aggregate and idiosyncratic states. Because of the latter, we can without loss of generality restrict our attention to default-free contracts.

Since capital is worthless in period 2, creditors cannot expect to get any positive repayment from the firms in the last period, even if given full control rights over

assets and maximum bargaining power. In period 1, on the other hand, positive repayment can be supported by the thread of creditors seizing the firm's assets and selling them on the secondary capital market. However, even if the creditors are given full control over existing firm's assets and maximum bargaining power, they cannot negotiate repayment higher than the market value of the underlying collateral. This is because the firm always has an outside option to walk away from the negotiations table and keep the cash, which it then can use to buy new assets on the secondary capital market.

To sum up, therefore, financial contracts in period 0 can specify any state-contingent repayment $p(Z,\varepsilon) \leq k_1 Q(Z)$ in period 1, where $k_1 Q(Z)$ is the market value of firm's capital in the aggregate state Z. Note that $p(Z,\varepsilon)$ can in principle be negative if a firm decides to get insurance against particular states, as there is no commitment problem on behalf of the creditors.[31] We will see, however, that no firm will use insurance in equilibrium. Figure 2.6 summarizes the timing of events.

Firm's problem

Each firm solves:

$$\max \; v = \int_0^1 \int_0^1 \Big[k_2(Z,\varepsilon) + c_2(Z,\varepsilon)\Big] \,\mathrm{d}\varepsilon \,\mathrm{d}Z \tag{2.10}$$

$$\text{s.t.} \quad k_1 + c_1 \leq n + b \tag{2.11}$$

$$b = \int_0^1 \int_0^1 p(Z,\varepsilon) \,\mathrm{d}\varepsilon \,\mathrm{d}Z \tag{2.12}$$

$$k_2(Z,\varepsilon)Q(Z) + c_2(Z,\varepsilon) \leq k_1 \cdot [Z+\varepsilon] + c_1 + k_1 Q(Z) - p(Z,\varepsilon) \quad \forall \varepsilon, Z \tag{2.13}$$

$$p(Z,\varepsilon) \leq k_1 Q(Z) \quad \forall \varepsilon, Z \tag{2.14}$$

$$k_1 \geq 0,\, c_1 \geq 0,\, k_2(Z,\varepsilon) \geq 0,\, c_2(Z,\varepsilon) \geq 0 \quad \forall \varepsilon, Z$$

[31] Note that firms could also in principle contract insurance to be paid out to them in period 2, but because firms can freely store cash, this option is irrelevant.

The objective function (2.10) is simply the expectation of total funds available for consumption to the firm owners in period 2, which consist of cash flow generated by capital k_2 held into period 2, as well as cash c_2 carried over from period 1 into period 2. We explicitly indicate the variables that may potentially depend on the aggregate and idiosyncratic firm shocks. (2.11) is the budget constraint in period 0, (2.12) is the break-even condition for the creditors, (2.13) is the budget constraint in period 1, while (2.14) constraints the maximum feasible repayment due to the financial frictions discussed above.

We can make several immediate observations:

- Firms do not carry any cash from period 0 into period 1, so $c_1 = 0$. Holding cash over capital can potentially be beneficial only in states where $Z + \varepsilon + Q(Z) < 1$, but because state-contingent contracts are available, the firm would be better off buying insurance against those states instead.

- In any equilibrium, $Q \leq 1$, otherwise no-one would choose to hold capital in period 1, a contradiction. Similarly, $Q(Z) > 0 \; \forall Z$, otherwise the demand for capital would be infinite in period 1 in some states.

- Trivially, budget constraint (2.13) must bind at the optimum. This allows us to substitute it directly into the objective function. Note that in states where $Q(Z) = 1$, $k_2 + c_2$ simply equals to the right hand side of the constraint. When $Q(Z) < 1$, the firm will optimally only buy capital, so $c_2 = 0$ and k_2 is equal to right hand side of the constraint (2.13) divided by $Q(Z)$. Either way, we can substitute the right hand side of the constraint (2.13) divided by $Q(Z)$ instead of $k_2 + c_2$.

With these observations, the problem can be simplified and rewritten as:

$$\max \; v = k_1 + \int_0^1 \frac{k_1 \cdot [Z + \frac{1}{2}] - \int_0^1 p(Z, \varepsilon) \, \mathrm{d}\varepsilon}{Q(Z)} \, \mathrm{d}Z$$

$$\begin{aligned} \text{s.t.} \quad & k_1 \leq n + \int_0^1 \int_0^1 p(Z,\varepsilon)\,\mathrm{d}\varepsilon\,\mathrm{d}Z \\ & p(Z,\varepsilon) \leq k_1 Q(Z) \quad \forall \varepsilon, Z \\ & k_1 \geq 0 \end{aligned}$$

The Lagrangian is:

$$\begin{aligned} \mathcal{L} = k_1 + & \int_0^1 \frac{k_1 \cdot [Z + \frac{1}{2}] - \int_0^1 p(Z,\varepsilon)\,\mathrm{d}\varepsilon}{Q(Z)}\,\mathrm{d}Z \\ & - \lambda \left[k_1 - n - \int_0^1 \int_0^1 p(Z,\varepsilon)\,\mathrm{d}\varepsilon\,\mathrm{d}Z \right] \\ & - \int_0^1 \int_0^1 \mu(Z,\varepsilon)[p(Z,\varepsilon) - k_1 Q(Z)]\,\mathrm{d}\varepsilon\,\mathrm{d}Z \\ & + \nu k_1 \end{aligned} \tag{2.15}$$

Kuhn-Tucker first order conditions with respect to k_1 and $p(Z,\varepsilon)$ are, respectively:

$$1 + \int_0^1 \frac{Z + \frac{1}{2}}{Q(Z)}\,\mathrm{d}Z - \lambda + \int_0^1 \int_0^1 \mu(Z,\varepsilon) Q(Z)\,\mathrm{d}\varepsilon\,\mathrm{d}Z + \nu = 0 \tag{2.16}$$

and

$$\mu(Z,\varepsilon) = \lambda - \frac{1}{Q(Z)}, \tag{2.17}$$

where the Lagrangian multipliers λ, μ and ν must be non-negative. Combining (2.16) and (2.17), we also get:

$$\lambda = \frac{\int_0^1 \frac{Z+\frac{1}{2}}{Q(Z)}\,\mathrm{d}Z + \nu}{1 - \int_0^1 Q(Z)\,\mathrm{d}Z} > 0, \tag{2.18}$$

so, unsurprisingly, the period 0 budget constraint (2.11) binds.

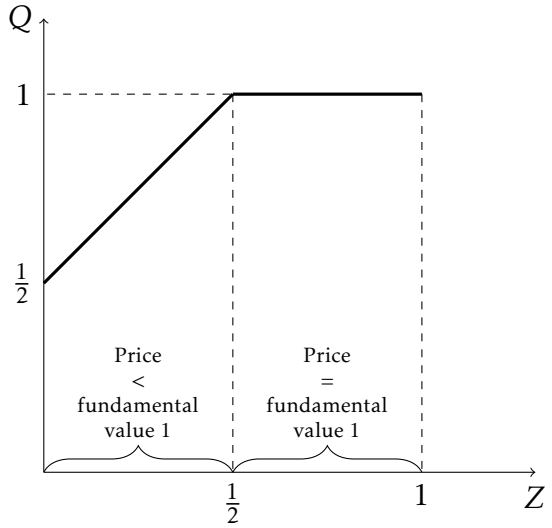


Figure 2.7: Asset price in period 1 as a function of aggregate shock

Equilibrium

Proposition 2.1. *There is a unique, symmetric equilibrium in which all firms exhaust their borrowing capacity, and invest all own and borrowed funds into capital in period 0. Therefore,* $p(Z,\varepsilon) = k_1 Q(Z)\ \forall \varepsilon, Z$*, and optimal financial contracts can effectively take the form of simple non-recourse debt secured by resale value of capital.*

The equilibrium price of capital in period 1 is

$$Q_1^*(Z) = \begin{cases} 1 & \text{if } Z \geq \frac{1}{2} \\ Z + \frac{1}{2} & \text{if } Z < \frac{1}{2}. \end{cases} \tag{2.19}$$

Firms borrow $\$7n$ *and thus invest in* $8n$ *assets in equilibrium.*

Proof. First, we verify that individual firm's optimality conditions are satisfied. As $k_1 > 0$, $\nu = 0$. Plugging (2.19) into (2.18), we obtain $\lambda = 9$. Using (2.17), this implies that $\mu > 0\,\forall Z$, since μ increases in Z, and $\mu = 7$ even when $Z = 0$. Therefore, it is indeed optimal for each firm to borrow until constraint (2.14) binds in all states, and invest all funds in capital.

Second, as all firms face the same constraints, this proposed equilibrium is

symmetric, and each firm invests the same amount of capital. Substituting (2.19) into binding constraints (2.14), (2.12), and (2.11), it is easy to verify that $k_1 = 8n$ and hence $b = 7n$.

Third, we need to verify that Q^* in (2.19) is indeed the equilibrating price on the market for assets in period 1 in every state Z. When $Q = 1$, unconstrained firms are indifferent between selling and buying capital, since each unit earns \$1 in period 2 with certainty. However, there are firms that must sell some assets, since their period 1 cash flow is not enough to cover debt repayments, in this case $p = k_1$. A marginal firm that has just enough cash to repay debt in state Z is one with $Z + \varepsilon = 1$, so every firm with $\varepsilon < 1 - Z$ has to sell at least $k_1 \cdot [1 - Z - \varepsilon]$ assets. This means that the market must have enough liquidity to absorb at least

$$\underline{K}^S = \int_0^{1-Z} k_1 \cdot [1 - Z - \varepsilon] \, \mathrm{d}\varepsilon = \frac{1}{2} k_1 [1 - Z]^2$$

assets. On the other hand, firms with $\varepsilon \geq 1 - Z$ can spend at most $k_1 \cdot [Z + \varepsilon - 1]$ on buying assets, so the total market demand when $Q = 1$ is at most:

$$\overline{K}^D = \int_{1-Z}^1 k_1 \cdot [Z + \varepsilon - 1] \, \mathrm{d}\varepsilon = \frac{1}{2} k_1 Z^2.$$

It is easy to verify that $\overline{K}^D \geq \underline{K}^S$ as long as $Z \geq \frac{1}{2}$, as required.

When $Q < 1$, on the other hand, all firms with spare liquidity want to buy assets. However, the firms whose cash flow falls below the market value of their capital, i.e. those with $\varepsilon < Q - Z$, have no choice but to sell $k_1 \cdot [Q - Z - \varepsilon]$ worth of assets in order to make debt repayments. The total supply of capital on the market is therefore:[32]

$$K^S(Q) = \int_0^{Q-Z} k_1 \frac{Q - Z - \varepsilon}{Q} \, \mathrm{d}\varepsilon = k_1 \frac{(Q - Z)^2}{2Q}.$$

[32] This is "net" supply, since, in principle, firms could sell more capital than they have to sell, but then buy some of it back. This would make no difference to the market price.

All other firms spend their spare cash, $k_1 \cdot [Z + \varepsilon - Q]$, on purchasing assets, so the aggregate demand for assets is:[33]

$$K^D(Q) = \int_{Q-Z}^{1} k_1 \frac{Z + \varepsilon - Q}{Q} \, \mathrm{d}\varepsilon = k_1 \frac{(1 - Q + Z)^2}{2Q}.$$

Market clearing condition $K^S(Q) = K^D(Q)$ yields the equilibrium price $Q^* = Z + \frac{1}{2}$, which is indeed smaller than 1 when $Z < \frac{1}{2}$, as required.

Lastly, to prove uniqueness, we note that $Q^*(Z)$ in (2.19) is, in fact, the lower bound on values that equilibrium asset price could possibly take in state Z. Suppose the total amount of physical capital in the industry is K. In state Z, it generates the aggregate cash flow of $K \cdot [Z + \frac{1}{2}]$. Even if the entire capital stock is traded on the market in order to repay outstanding debts (which would always be enough, since debt cannot exceed the market value of capital), there is enough liquidity for firms to be able to pay at least $Z + \frac{1}{2}$ per unit of capital. Competition for assets will ensure that $Z + \frac{1}{2} \leq Q \leq 1$ when $Z < \frac{1}{2}$, and $Q = 1$ otherwise. Using (2.18) we then can show that possible values of λ are bounded below by 8.[34] In turn, using (2.17), the lower bounds on Q and λ imply that μ is at least 7, i.e. the constraint (2.14) on firms' repayments always binds in equilibrium. But there is only one equilibrium in which all firms fully exhaust their borrowing constraints: the symmetric equilibrium we have found above. The equilibrium is necessarily symmetric because all firms face the same binding constraints and asset prices in period 1, and we only needed symmetry to derive the equilibrium prices. *Q.E.D.*

Corollary 2.1. *When aggregate shock is $Z < \frac{1}{2}$, period 2 return $\frac{1}{Q^*(Z)} - 1$ is negatively related to industry earnings and asset prices in period 1. However, period 2 return is constant, and hence there is no predictability when $Z \geq \frac{1}{2}$.*

[33] Likewise, this is "net" demand.

[34] Examining equation (2.18), we see that the nominator is at least $\int_0^1 [Z + \frac{1}{2}]] \, \mathrm{d}Z = 1$ (when setting $\nu = 0$ and $Q(Z) = 1 \; \forall Z$), while the denominator is at most $1 - \int_0^1 Q^*(Z) \, \mathrm{d}Z = \frac{1}{8}$ (when setting $Q(Z) = Q^*(Z)$ in (2.19)). Together, these imply that λ is at least 8.

This is the central result, in accordance with the empirical finding. Note that predictability kicks in once aggregate earnings fall below average. Thus, perhaps contrary to common intuition, there does not need to be a particularly deep recession for asset prices to drop below the fundamental value.

Slumps in industry investment

It is feasible to extend the model to also incorporate industry investment, which would only require a sacrifice of simplicity and tractability. One way to do it is adding a competitive capital producing sector that faces convex production costs. This would imply an upward sloping supply of new assets, in addition to sales of existing capital. [35] It is not difficult then to imagine what happens to industry investment. In an expansion, there is enough liquidity to absorb the combined supply of existing and new assets at the price equal to their fundamental value, and the industry achieves the first best level of investment. In a recession, however, asset price must fall below the fundamental value due to liquidity constraints faced by the buyers, as before. The lower is the price, the lower industry investment falls below its first best level, and the gap between the marginal cost of producing capital and return on assets widens. In this case, while sales of second hand assets at depressed prices simply represents a transfer from the sellers to the buyers, a plunge in industry investment is a manifestation of industry inefficiencies due to financial constraints. This mechanism generates a negative association between industry investment and future returns, in line with my empirical findings.

2.5 Conclusion

In this paper I study the link between earnings, asset prices, industry investment, and the subsequent returns on capital in the context of the dry bulk shipping

[35] In fact, assuming quadratic production costs, it is still possible to obtain closed-form solutions.

industry. I document that future excess returns on owning ships are strongly and negatively related to the current level of industry activity during global economic recessions, but there is no evidence of such predictability during expansions. I discuss the explanations of predictability previously suggested in the literature, and find them to be at odds with the empirical evidence. Consistent with accounts from industry practitioners, I propose an explanation that links the countercyclical predictability of returns on ships to the liquidity constraints faced by the firms in the industry. Specifically, due to these liquidity constraints, ship prices and industry investment plunge excessively during recessions. This highlights the important role that financial frictions play in competitive, capital-intensive industries like shipping.

Bibliography

Bernanke, B. S., Gertler, M., and Gilchrist, S. (1999). Chapter 21 the financial accelerator in a quantitative business cycle framework. volume 1 of *Handbook of Macroeconomics*, pages 1341–1393. Elsevier.

Brunnermeier, M. K. and Sannikov, Y. (2014). A Macroeconomic Model with a Financial Sector. *American Economic Review*, 104(2):379–421.

Campbell, J. Y., Giglio, S., and Pathak, P. (2011). Forced sales and house prices. *American Economic Review*, 101(5):2108–31.

Cochrane, J. H. (2011). Presidential Address: Discount Rates. *Journal of Finance*, 66(4):1047–1108.

Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106(1):157 – 181.

Day, R. (2019). Private equity and shipping: From common and preferred equity to secondary loan purchases. Presentation slides. Retreived June

28, 2021 from https://www.marinemoney.com/system/files/media/2019-09/255%20PM%20-%20Randee%20Day%20-%20Goldin%20Maritime.pdf.

Dizard, J. (2015). The expensive shipping news for Wall Street's smart money. *Financial Times*. February 27, 2015. Retreived June 28, 2021 from https://www.ft.com/.

Evans, J. (1988). The elasticity of supply of transport. *Maritime Policy & Management*, 15(4):309–313.

Financier Worldwide (2012). Financial distress in the shipping sector. Retreived June 28, 2021 from https://www.financierworldwide.com/financial-distress-in-the-shipping-sector#.YNof7ehKiUk.

Franks, J. R., Seth, G., Sussman, O., and Vig, V. (2021a). Revisiting the Asset Fire Sale Discount: Evidence from Commercial Aircraft Sales. Finance Working Paper 722/2021, European Corporate Governance Institute.

Franks, J. R., Seth, G., Sussman, O., and Vig, V. (2021b). The Privatization of Bankruptcy: Evidence from Financial Distress in the Shipping Industry. Finance Working Paper 505/2017, European Corporate Governance Institute.

Gertler, M. and Karadi, P. (2011). A model of unconventional monetary policy. *Journal of Monetary Economics*, 58(1):17–34.

Greenwood, R. and Hanson, S. G. (2015). Waves in Ship Prices and Investment. *The Quarterly Journal of Economics*, 130(1):55–109.

Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). The U.S. Treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54(8):2291–2304.

Hamilton, J. D. (2021). Measuring global economic activity. *Journal of Applied Econometrics*, 36(3):293–303.

Hansen, B. E. and Racine, J. S. (2018). Bootstrap Model Averaging Unit Root Inference. Department of Economics Working Papers 2018-09, McMaster University.

Harri, A. and Brorsen, W. B. (2009). The overlapping data problem. *SSRN Electronic Journal*, 3:78–115.

Harris, M. and Raviv, A. (1995). The Role of Games in Security Design. *Review of Financial Studies*, 8(2):327–367.

Henkel, S. J., Martin, J. S., and Nardari, F. (2011). Time-varying short-horizon predictability. *Journal of Financial Economics*, 99(3):560 – 580.

Kalouptsidi, M. (2014). Time to Build and Fluctuations in Bulk Shipping. *American Economic Review*, 104(2):564–608.

Karatzas, B. (2016). Dry bulk vessel values are a subject of concern, once again. March 08, 2016. Retreived June 28, 2021 from https://shippingfinance.wordpress.com/2016/03/08/dry-bulk-vessel-values-are-a-subject-of-concern-once-again-basil-m-karatzas/.

Karatzas, B. (2017). The asset appreciation play has yet to leave port. July 06, 2017. Retreived June 28, 2021 from https://shippingfinance.wordpress.com/2017/07/09/the-asset-appreciation-play-has-yet-to-leave-the-port-basil-m-karatzas/.

Kiefer, N. M. and Vogelsang, T. J. (2002). Heteroskedasticity-Autocorrelation Robust Standard Errors Using The Bartlett Kernel Without Truncation. *Econometrica*, 70(5):2093–2095.

Kiefer, N. M., Vogelsang, T. J., and Bunzel, H. (2000). Simple robust testing of regression hypotheses. *Econometrica*, 68(3):695–714.

Kiyotaki, N. and Moore, J. (1997). Credit cycles. *Journal of Political Economy*, 105(2):211–248.

Kostakis, A., Magdalinos, T., and Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *The Review of Financial Studies*, 28(5):1506–1553.

Lazarus, E., Lewis, D. J., Stock, J. H., and Watson, M. W. (2018). HAR Inference: Recommendations for Practice. *Journal of Business & Economic Statistics*, 36(4):541–559.

Magdalinos, T. and Phillips, P. C. (2011). Econometric inference in the vicinity of unity.

Odell, M. and Makan, A. (2013). Wave of private equity money flows into shipping. *Financial Times*. October 27, 2013. Retreived June 28, 2021 from `https://www.ft.com/`.

OECD (2019). OECD Composite Leading Indicators: Reference Turning Points and Component Series. Retreived Apr 07, 2019 from `http://www.oecd.org/sdd/leading-indicators/`.

Phillips, P. C. and Lee, J. H. (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics*, 177(2):250 – 264. Dynamic Econometric Modeling and Forecasting.

Pulvino, T. C. (1998). Do asset fire sales exist? an empirical investigation of commercial aircraft transactions. *The Journal of Finance*, 53(3):939–978.

Rajan, R. and Ramcharan, R. (2016). Local financial capacity and asset values: Evidence from bank failures. *Journal of Financial Economics*, 120(2):229–251.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2009). Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *The Review of Financial Studies*, 23(2):821–862.

Shleifer, A. and Vishny, R. (2011). Fire sales in finance and macroeconomics. *Journal of Economic Perspectives*, 25(1):29–48.

Shleifer, A. and Vishny, R. W. (1992). Liquidation values and debt capacity: A market equilibrium approach. *The Journal of Finance*, 47(4):1343–1366.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, 54(3):375 – 421.

Stopford, M. (2009). *Maritime Economics*. Routledge, 3 edition.

Wallis, K. (2016). Bargain prices, rising cargo rates lift second-hand ship sales from the shallows. *Reuters*. November 29, 2016. Retreived June 28, 2021 from https://www.reuters.com/.

Wallis, K. (2017). Cheap as ships: Second-hand ship sales reach 10-year high. *Reuters*. December 21, 2017. Retreived June 28, 2021 from https://www.reuters.com/.

Wright, R. (2016). Shipowners rocked by sinking charter rate. *Financial Times*. January 3, 2016. Retreived June 28, 2021 from https://www.ft.com/.

Ye, X. and Sun, Y. (2018). Heteroskedasticity- and autocorrelation-robust F and t tests in Stata. *Stata Journal*, 18(4):951–980(30).

Appendices

2.A Alternative ship types

Let us now verify that the results are not limited to the Panamax class ships, and are indeed representative of the broader industry. I estimate predictive regressions (2.7) for smaller 32k DWT Handysize ships, and larger 150k DWT Capesize vessels.

Table 2.A.1: Regime-Dependent Predictability: Alternative Ship Sizes

Panel A: Handysize ships

	X = net real earnings Π			X = used ship price P			X = investment I		
K	12m	24m	36m	12m	24m	36m	12m	24m	36m
b_1	-0.045*	-0.083*	-0.108+	-0.014**	-0.025+	-0.034+	-1.043***	-0.933***	-1.118
	(0.018)	(0.025)	(0.046)	(0.004)	(0.011)	(0.016)	(0.185)	(0.213)	(0.774)
	[0.033]	[0.011]	[0.057]	[0.010]	[0.085]	[0.093]	[0.000]	[0.001]	[0.199]
b_2	0.031	0.039	0.057	0.001	-0.005	-0.005	0.134	-0.536	0.515
	(0.022)	(0.042)	(0.045)	(0.006)	(0.004)	(0.009)	(0.381)	(0.393)	(0.614)
	[0.186]	[0.392]	[0.246]	[0.910]	[0.349]	[0.612]	[0.732]	[0.244]	[0.449]
F-test $b_1 = b_2$	0.002	0.098	0.053	0.055	0.130	0.062	0.011	0.391	0.033
T_1 obs.	241	229	217	241	229	217	119	107	95
T_2 obs.	291	291	291	291	291	291	162	162	162
R^2	0.097	0.105	0.125	0.134	0.154	0.196	0.117	0.066	0.083
Adj. R^2	0.092	0.100	0.120	0.129	0.149	0.191	0.107	0.055	0.072

Panel B: Capesize ships

	X = net real earnings Π			X = used ship price P			X = investment I		
K	12m	24m	36m	12m	24m	36m	12m	24m	36m
b_1	-0.015***	-0.028	-0.036	-0.005**	-0.010+	-0.014*	-0.747*	-0.971*	-1.674+
	(0.003)	(0.015)	(0.019)	(0.001)	(0.005)	(0.006)	(0.242)	(0.343)	(0.841)
	[0.000]	[0.119]	[0.109]	[0.006]	[0.096]	[0.042]	[0.021]	[0.018]	[0.094]
b_2	0.006	0.008	0.016	0.001	0.001	0.001	-0.124	-0.420	0.571
	(0.005)	(0.010)	(0.012)	(0.002)	(0.003)	(0.004)	(0.420)	(0.477)	(0.990)
	[0.301]	[0.477]	[0.248]	[0.731]	[0.836]	[0.729]	[0.774]	[0.404]	[0.585]
F-test $b_1 = b_2$	0.012	0.164	0.110	0.009	0.096	0.049	0.271	0.295	0.044
T_1 obs.	241	229	217	241	229	229	119	107	95
T_2 obs.	252	252	252	252	252	252	162	162	162
R^2	0.152	0.105	0.148	0.159	0.170	0.228	0.099	0.038	0.073
Adj. R^2	0.146	0.099	0.142	0.154	0.165	0.223	0.089	0.027	0.062

Notes. This table reports the estimates of slope coefficients and F-test statistics for regime switching time series regressions of the form $rx_{t+K} = \mathbb{1}_t^R(a_1 + b_1 X_t) + (1 - \mathbb{1}_t^R)(a_2 + b_2 X_t) + v_{t+K}$, where $\mathbb{1}_t^R$ is the OECD global recession indicator, X_t is one of the forecasting variables, K is the forecasting horizon in months, and rx_{t+K} is cumulative log excess return between months t and $t+K$. Standard errors are reported in parentheses, and p-values are reported in brackets, while the F-test row reports p-value for testing $H_0 : b_1 = b_2$, all obtained using OS HAR inference. $^+$,*,**, and *** indicate significance at the 0.1, 0.05, 0.01, and 0.001 levels, respectively.

The steps to construct the datasets for these types are identical to the construction of our main dataset.

Table 2.A.1 presents the results. Overall, they are fully in line with the results

Table 2.B.1: Regime-Dependent Predictability in Shipping: NBER US Recessions

	X = net real earnings Π			X = used ship price P			X = investment I		
K	12m	24m	36m	12m	24m	36m	12m	24m	36m
b_1	-0.025*	-0.027***	-0.031	-0.007*	-0.009***	-0.011	-1.662**	-1.595***	-2.847***
	(0.009)	(0.006)	(0.018)	(0.003)	(0.002)	(0.006)	(0.466)	(0.269)	(0.507)
	[0.014]	[0.001]	[0.116]	[0.033]	[0.001]	[0.107]	[0.004]	[0.000]	[0.000]
b_2	0.004	-0.013	-0.008	-0.003	-0.009	-0.010	0.324	-0.611	0.288
	(0.009)	(0.016)	(0.029)	(0.002)	(0.005)	(0.008)	(0.474)	(0.635)	(0.578)
	[0.684]	[0.484]	[0.786]	[0.156]	[0.133]	[0.260]	[0.510]	[0.390]	[0.644]
F-test $b_1 = b_2$	0.041	0.443	0.516	0.416	0.959	0.908	0.005	0.140	0.019
T_1 obs.	56	56	56	56	56	56	26	26	26
T_2 obs.	472	460	448	472	460	448	251	239	227
R^2	0.057	0.044	0.052	0.077	0.116	0.135	0.101	0.047	0.086
Adj. R^2	0.052	0.039	0.047	0.071	0.111	0.130	0.091	0.036	0.075

Notes. This table reports the estimates of slope coefficients and F-test statistics for regime switching time series regressions of the form $rx_{t+K} = \mathbb{1}_t^R(a_1 + b_1 X_t) + (1 - \mathbb{1}_t^R)(a_2 + b_2 X_t) + v_{t+K}$, where $\mathbb{1}_t^R$ is the NBER US recession indicator, X_t is one of the forecasting variables, K is the forecasting horizon in months, and rx_{t+K} is cumulative log excess return between months t and $t+K$. Standard errors are reported in parentheses, and p-values are reported in brackets, while the F-test row reports p-value for testing $H_0 : b_1 = b_2$, all obtained using OS HAR inference. $^+$,*,**, and *** indicate significance at the 0.1, 0.05, 0.01, and 0.001 levels, respectively.

I obtained for Panamax class vessels: for both ship types, coefficients on ship prices, earnings, and investment are all negative and statistically significant for most regressors and forecasting horizons during recessions. Among the only three exceptions, two of the p-values are only marginally above the 10% threshold. Similarly, in the majority of cases we can reject $H_0 : b_1 = b_2$.

2.B NBER recessions

As the OECD global recession indicator is based on growth cycle approach and an automatic dating procedure, it may tend to call recessions "too often". To check the robustness of our findings, it is therefore desirable to use a more conservative indicator of recessions. Alas, there are no other official monthly indicators for world recessions I am aware of. As a robustness exercise, I estimate regressions

(2.7) using NBER indicator for recessions in the US. Due to the sheer size of the US economy, one would expect it to correlate with periods of global downturns.

Table 2.B.1 presents the results, which are similar to the ones presented in Table 2.3. While there are a lot less recessionary observations when using NBER indicator, the predictive slope coefficients on net earnings and prices during recessions are statistically significant at 12 and 24 months horizons, at the 5% and 0.1% levels respectively. At 36 months horizon, the p-values are just short of clearing the 10% bar. The evidence of industry investments being able to predict future returns during NBER recessions is surprisingly strong despite limited number of observations: all coefficient estimates are significant at least at the 1% level.

As before, there is no indication there being any return predictability during expansions. However, some of the point estimates in expansions are now closer to those in recessions, likely because a few global recessions are now classified as expansions. This makes it difficult to formally reject the null of $b_1 = b_2$.

2.C Excluding the Financial Crisis

Figure 2.2 shows that the shipping industry has experienced a super boom-bust cycle during 2006-2009. This raises a concern that our results could potentially be driven by this unprecedented boom followed by the Global Financial Crisis. To address it, I estimate regime-switching predictive regressions using data through 2006. Specifically, I compute forward returns only with earnings and prices available up to 2006. This way I only use predictive regressors up to 2005, 2004, and 2003 to forecast forward returns at 12, 24, and 36 months, respectively. Since there is less than 10 years worth of observations on industry investment up to 2005, I exclude investment from this robustness check.

Table 2.C.1 present the results. We can safely conclude that the main finding

Table 2.C.1: Regime-Dependent Predictability in Shipping: 1976-2006

	X = net real earnings Π			X = used ship price P		
K	12m	24m	36m	12m	24m	36m
b_1	-0.035	-0.115**	-0.154*	-0.010*	-0.025***	-0.033**
	(0.018)	(0.024)	(0.038)	(0.003)	(0.003)	(0.006)
	[0.127]	[0.008]	[0.015]	[0.017]	[0.001]	[0.004]
b_2	0.022^{+}	0.032	0.005	-0.001	-0.001	-0.010
	(0.010)	(0.018)	(0.059)	(0.007)	(0.007)	(0.009)
	[0.100]	[0.122]	[0.940]	[0.896]	[0.937]	[0.312]
F-test $b_1 = b_2$	0.076	0.003	0.107	0.256	0.014	0.051
T_1 obs.	169	169	169	169	169	169
T_2 obs.	191	179	167	191	179	167
R^2	0.115	0.273	0.310	0.159	0.364	0.450
Adj. R^2	0.107	0.267	0.304	0.152	0.359	0.445

Notes. This table reports the estimates of slope coefficients and F-test statistics for regime switching time series regressions of the form $rx_{t+K} = \mathbb{1}_t^R(a_1 + b_1 X_t) + (1 - \mathbb{1}_t^R)(a_2 + b_2 X_t) + v_{t+K}$, where $\mathbb{1}_t^R$ is the OECD global recession indicator, X_t is one of the forecasting variables, K is the forecasting horizon in months, and rx_{t+K} is cumulative log excess return between months t and $t+K$. Standard errors are reported in parentheses, and p-values are reported in brackets, while the F-test row reports p-value for testing $H_0 : b_1 = b_2$, all obtained using OS HAR inference. $^{+}$,*,**, and *** indicate significance at the 0.1, 0.05, 0.01, and 0.001 levels, respectively.

of return predictability being concentrated during recessions is not driven by the Great Recession. Estimates of predictive slope coefficients are in fact slightly larger than those obtained in the whole sample, and all but one are statistically significant. In the majority of cases we can also comfortably reject $H_0 : b_1 = b_2$.

2.D Potential upward bias in returns during recessions

One potential concern with my measure of shipping returns is that it assumes that an owner is always able to lease their ship out at the prevailing time charter rates. However, some ships may end up idle during recessions due to the lack of demand and difficulty finding business for them. Because of this, the measures of earnings and returns based on time charter rates may tend to overstate the true

Table 2.D.1: Forecasting net excess capital gains

	X = net real earnings Π			X = used ship price P			X = investment I		
K	12m	24m	36m	12m	24m	36m	12m	24m	36m
b_1	-0.050***	-0.080*	-0.094*	-0.012***	-0.022*	-0.027*	-1.865***	-2.079***	-2.769**
	(0.007)	(0.029)	(0.032)	(0.003)	(0.006)	(0.007)	(0.312)	(0.419)	(0.748)
	[0.000]	[0.050]	[0.026]	[0.001]	[0.018]	[0.018]	[0.001]	[0.001]	[0.006]
b_2	-0.008	-0.022	-0.016	-0.004	-0.008^{+}	-0.009	-1.040^{+}	-2.108*	-1.288
	(0.010)	(0.017)	(0.015)	(0.002)	(0.004)	(0.005)	(0.454)	(0.473)	(0.638)
	[0.458]	[0.249]	[0.358]	[0.096]	[0.099]	[0.134]	[0.051]	[0.011]	[0.114]
F-test $b_1 = b_2$	0.001	0.171	0.110	0.006	0.136	0.065	0.253	0.948	0.092
T_1 obs.	237	225	217	237	225	217	115	103	95
T_2 obs.	291	291	287	291	291	287	162	162	158
R^2	0.189	0.230	0.248	0.229	0.327	0.382	0.176	0.260	0.249
Adj. R^2	0.184	0.226	0.243	0.225	0.323	0.378	0.167	0.252	0.240

Notes. This table reports the estimates of slope coefficients and F-test statistics for regime switching time series regressions of the form $\Delta px_{t+K} = \mathbb{1}_t^R(a_1 + b_1 X_t) + (1 - \mathbb{1}_t^R)(a_2 + b_2 X_t) + v_{t+K}$, where $\mathbb{1}_t^R$ is the OECD global recession indicator, X_t is one of the forecasting variables, K is the forecasting horizon in months, and Δpx_{t+K} is cumulative log net excess capital gain between months t and $t+K$. Standard errors are reported in parentheses, and p-values are reported in brackets, while the F-test row reports p-value for testing $H_0 : b_1 = b_2$, all obtained using OS HAR inference. $^{+}$,*,**, and *** indicate significance at the 0.1, 0.05, 0.01, and 0.001 levels, respectively.

earnings and returns to owning ships especially during recessions. Furthermore, it is likely that the number of idle ships increases as the demand for shipping services falls, and ship earnings, prices, and investment decrease. The predictive variables therefore would be negatively correlated with this bias in the measure of returns, generating an impression of regime-dependent predictability.

To address this concern, I test whether the net excess capital gains on ships are also predictable during recessions. Specifically, I estimate regressions of the form:

$$\Delta px_{t+K} = \mathbb{1}_t^R(a_1 + b_1 X_t) + (1 - \mathbb{1}_t^R)(a_2 + b_2 X_t) + v_{t+K}, \tag{2.D.1}$$

where Δpx_{t+K} is the capital gain over the following K months, net of depreciation and in excess of risk free rate, computed as detailed in the footnote 25. We are effectively setting the ship earnings to zero, which should help eliminate the above-mentioned bias.

Table 2.E.1: Regime-Dependent Predictability: Annual Sampling

	X = net real earnings Π			X = used ship price P			X = investment I		
K	1y	2y	3y	1y	2y	3y	1y	2y	3y
b_1	-0.020***	-0.029^{+}	-0.039^{+}	-0.006***	-0.010*	-0.014*	-1.095***	-0.974***	-1.234^{+}
	(0.004)	(0.014)	(0.020)	(0.002)	(0.004)	(0.006)	(0.181)	(0.191)	(0.592)
	[0.000]	[0.056]	[0.069]	[0.001]	[0.019]	[0.024]	[0.000]	[0.000]	[0.071]
b_2	0.039***	0.027	0.029	0.004	0.001	-0.001	1.456***	0.590	1.256
	(0.008)	(0.024)	(0.042)	(0.004)	(0.005)	(0.008)	(0.386)	(0.597)	(0.878)
	[0.000]	[0.255]	[0.501]	[0.403]	[0.775]	[0.947]	[0.001]	[0.347]	[0.183]
F-test $b_1 = b_2$	0.000	0.149	0.155	0.042	0.059	0.160	0.000	0.020	0.008
T_1 obs.	19	18	17	19	18	17	9	8	7
T_2 obs.	25	25	25	25	25	25	14	14	14
R^2	0.212	0.113	0.121	0.163	0.144	0.166	0.283	0.046	0.067
Adj. R^2	0.153	0.044	0.052	0.100	0.078	0.101	0.170	-0.113	-0.097

Notes. This table reports the estimates of slope coefficients and F-test statistics for regime switching time series regressions of the form $rx_{t+K} = \mathbb{1}_t^R(a_1 + b_1 X_t) + (1 - \mathbb{1}_t^R)(a_2 + b_2 X_t) + v_{t+K}$, where $\mathbb{1}_t^R$ is the OECD global recession indicator, X_t is one of the forecasting variables, K is the forecasting horizon, and rx_{t+K} is cumulative log excess return between months t and $t+K$. The data used in estimation is annually sampled in June. Standard errors are reported in parentheses, and p-values are reported in brackets, while the F-test row reports p-value for testing $H_0 : b_1 = b_2$, all obtained using OS HAR inference. $^{+}$,*,**, and *** indicate significance at the 0.1, 0.05, 0.01, and 0.001 levels, respectively.

Table 2.D.1 presents the results, which are very similar to the ones reported in Table 2.3. All coefficient estimates are statistically significant in recessions at least at the 5% level, and in the majority of cases we either reject $H_0 : b_1 = b_2$, or are close to rejecting it. Therefore, it is unlikely that it is a bias in the measure of returns that drives my main results.

2.E Annual sampling

As already mentioned, the returns data I use has high degree of overlapping. While I take utmost care to account for that by relying on OS HAR inference, concern remains that apparent statistical significance of predictive coefficients is driven by high degree of autocorrelation in the error term. One immediate

response to this concern is that it is not clear why this problem would manifest itself only during recessions.

To gain further confidence that my results are not driven by overlapping nature of observations, I estimate (2.7) using only annually sampled data.[36] This, of course, significantly reduces the number of observations, but completely eliminates the overlapping data problem at the 12 months horizon, and significantly reduces it at the 24 and 36 months horizons. Table 2.E.1 reports the results. Despite the limited number of observations, all my results survive both qualitatively and quantitatively.

2.F Stambaugh bias and near-unit-root inference

Shipping earnings and prices are both highly persistent at monthly frequency and endogenous. In finite samples this gives rise to an amplifying bias in the estimates of slope coefficients, and may lead to incorrect inference in favour of predictability, as was first demonstrated by Stambaugh (1999). Moreover, in the local-to-unity case, the bias is still present in the limit, and correcting this asymptotic bias is generally not possible, since it depends on the localizing coefficient that is not consistently estimable (Phillips and Lee, 2013). As with overlapping data, however, it is not clear why this would become a problem only during recessions, and not expansions.

To further check that my results are not driven by the Stambaugh bias, I employ the IVX estimation method developed by Magdalinos and Phillips (2011) for econometric inference in the vicinity of unity. It was adapted to the context of return predictability testing by Phillips and Lee (2013) and Kostakis et al. (2015), and is considered state of the art in the predictability literature. The basic idea behind the IVX approach is to remove the endogeneity by means of data-filtered

[36] Specifically, I pick only the observation in June from each year, such there is a 12 months gap between the observations.

Table 2.F.1: Regime-Dependent Predictability: IVX approach

Panel A: Forecasting net excess capital gains, monthly data									
	X = net real earnings Π			X = used ship price P			X = investment I		
K	1m	12m	24m	1m	12m	24m	1m	12m	24m
A_1	-0.001	-0.006***	-0.008***	-0.001*	-0.001***	-0.001***	-0.167**	-0.239**	-0.323
	[0.373]	[0.000]	[0.000]	[0.029]	[0.000]	[0.000]	[0.007]	[0.009]	[0.130]
A_2	0.002*	0.004*	0.006**	0.000	0.001	0.001+	0.054	0.089	0.259
	[0.045]	[0.035]	[0.005]	[0.715]	[0.212]	[0.087]	[0.403]	[0.463]	[0.410]
χ^2-test $A_1 = A_2$	0.049	0.000	0.000	0.099	0.001	0.000	0.014	0.078	0.257

Panel B: Forecasting excess returns, annual sampling			
	X = net real earnings Π	X = used ship price P	X = investment I
K	1y	1y	1y
A_1	-0.020+	-0.006*	-1.032
	[0.054]	[0.029]	[0.105]
A_2	0.043*	0.005	1.798*
	[0.017]	[0.344]	[0.033]
χ^2-test $A_1 = A_2$	0.002	0.057	0.008

Notes. Panel A reports the IVX estimates of slope coefficients in time series regressions of the form $\Delta px_{t+K} = \mathbb{1}_t^R(c_1 + A_1^K X_t) + (1 - \mathbb{1}_t^R)(c_2 + A_2^K X_t) + u_{t+K}$, where X_t is one of the three forecasting variables, K is the forecasting horizon in months, and $\Delta px_{t+K}(K)$ is log net excess capital gain between periods t and $t + K$. Panel B reports the IVX estimates based on the data annually sampled in June, using 1 year excess returns instead of the capital gains as the dependent variable. p-values are reported in brackets, and the χ^2-test row of each panel reports the p-values for testing $H_0 : A_1 = A_2$. $^+$,*,**, and *** indicate significance at the 0.1, 0.05, 0.01, and 0.001 levels, respectively.

instruments, the persistence of which can be explicitly controlled, so they remain in the class of near-stationary processes.

The IVX methodology assumes that the true data generating process has only one lag, i.e.:

$$y_{t+1} = c + AX_t + u_{t+1}, \tag{2.F.1}$$

where y_t is a dependent variable and $X_t = \mu + \rho X_{t-1} + v_t$ is a persistent predictor

variable. While the IVX method allows one to estimate so-called long-horizon predictive regressions to test for predictability of the dependent variable cumulated over future K periods, it still requires one period ahead data for the computation of variance-covariance matrix.

For this reason, I cannot use my series of excess returns when applying the IVX method to the monthly data, because I can only compute 12-months ahead returns. Instead, I first test for predictability of the future net excess capital gains, as in 2.D, and estimate the following IVX regressions:

$$\Delta px_{t+K} = \mathbb{1}_t^R(c_1 + A_1^K X_t) + (1 - \mathbb{1}_t^R)(c_2 + A_2^K X_t) + u_{t+K} \tag{2.F.2}$$

over the 1, 12, and 24 months horizons, using monthly data. Panel A in Table 2.F.1 presents the results. Firstly, we see that most slope coefficients in recessions remain highly statistically significant over the horizons of 1 months to 2 years. Secondly, the $H_0 : A_1 = A_2$ is rejected in most cases, often with extremely low p-values. Note, however, that when we estimate "misspecified" long-horizon model using the cumulative gains over K months, the IVX estimator $\tilde{A}_{IVX}^K$ is designed to still consistently estimate the true parameter A in the 1-month ahead DGP. This, alas, makes the IVX point estimates at the 12 and 24 months horizons significantly smaller in absolute value, and not comparable to the estimates in Table 2.D.1.

To be able to compare the magnitude of IVX estimates with my previous results, I therefore also estimate the model using annual sampling. Effectively I am assuming that the DGP has one lag, as required by the IVX method, but on annual frequency. Specifically, I estimate:

$$rx_{t+12} = \mathbb{1}_t^R(c_1 + A_1 X_t) + (1 - \mathbb{1}_t^R)(c_2 + A_2 X_t) + u_{t+12}, \tag{2.F.3}$$

using only data sampled 12 months apart. Note also that I am now able to use excess returns instead of capital gains. Panel B of Table 2.F.1 presents the results.

Table 2.G.1: Predicting 12-Months Returns: Lagged Regressors

	X = net real earnings Π			X = used ship price P		
h lag	3m	6m	12m	3m	6m	12m
b_1	-0.025**	-0.019*	-0.019	-0.007**	-0.006*	-0.006+
	(0.007)	(0.009)	(0.014)	(0.002)	(0.002)	(0.003)
	[0.002]	[0.048]	[0.207]	[0.002]	[0.039]	[0.087]
b_2	0.001	-0.000	0.001	-0.002	-0.002	0.001
	(0.010)	(0.009)	(0.006)	(0.003)	(0.004)	(0.003)
	[0.952]	[0.979]	[0.865]	[0.444]	[0.530]	[0.849]
F-test $b_1 = b_2$	0.066	0.203	0.234	0.183	0.460	0.127
T_1 obs.	237	237	237	237	237	237
T_2 obs.	288	285	279	288	285	279
R^2	0.105	0.089	0.085	0.134	0.121	0.117
Adj. R^2	0.100	0.084	0.079	0.129	0.116	0.112

Notes. This table reports the estimates of slope coefficients and F-test statistics for regime switching time series regressions of the form $rx_{t+12} = \mathbb{1}_t^R(a_1 + b_1 X_{t-h}) + (1 - \mathbb{1}_t^R)(a_2 + b_2 X_{t-h}) + v_{t+12}$, where $\mathbb{1}_t^R$ is the global recession indicator, X_t is one of the three forecasting variables, h is the regressor lag, and rx_{t+12} is cumulative log excess return between months t and $t+12$. Standard errors are reported in parentheses, and p-values are reported in brackets, while the F-test row reports p-value for testing $H_0 : b_1 = b_2$, all obtained using OS HAR inference. $^+$,*,**, and *** indicate significance at the 0.1, 0.05, 0.01, and 0.001 levels, respectively.

We see that point estimates in recessions are very close to the estimates at the 12-months horizon reported in Table 2.3. They are statistically significant, except for investment, and $H_0 : A_1 = A2$ is rejected at least at the 10% level.

Taken together, the evidence indicates that Stambaugh bias is unlikely to constitute a large fraction of OLS estimates and drive our main findings.

2.G Measurement errors in earnings and price series

Apparent statistical significance of prices and earnings in forecasting future excess returns could be potentially driven by large, correlated and persistent measurement errors in these series.[37] To investigate whether this is this the case, I estimate

[37] As the data is gathered based on a sample of ships that Clarkson's are able to get access to.

the following predictive regressions at 12 months horizon:

$$rx_{t+12} = \mathbb{1}_t^R(a_1 + b_1 X_{t-h}) + (1 - \mathbb{1}_t^R)(a_2 + b_2 X_{t-h}) + v_{t+12}, \tag{2.G.1}$$

where h is a number by which we lag the regressor. I.e. at any time t, we forecast cumulative excess returns over the next 12 months using X_{t-h} rather than X_t. The idea is that as we increase h, the impact of measurement errors is likely to diminish.

Table 2.G.1 shows the results. While the evidence for return predictability in recessions becomes weaker statistically as the regressor lag increases, as expected, the results do not vanish and remain in line with those in Table 2.3

Moreover, the story about measurement errors in price and earnings series is inconsistent with the fact that industry investment also negatively predict future returns, as well as with the regime-dependent nature of predictability: it is difficult to imagine why measurement errors would be larger during recessions.

2.H Using unadjusted data

Lastly, we verify that our results are not driven by simple linear adjustments of historical data we made in order to have consistent measures for the fixed ship size. Table 2.H.1 reports the results when instead we use raw data as obtained from Clarksons. At 12-months horizon, the results are nearly identical to those reported in Table 2.3. The results become weaker at 24 and 36 month horizons, although we note that p-value are typically not far from clearing the 10% threshold. This is expected, given that raw data likely contains larger measurement errors than adjusted series.

Table 2.H.1: No Adjustments to Earnings and Prices Data

	X = net real earnings Π			X = used ship price P			X = investment I		
K	12m	24m	36m	12m	24m	36m	12m	24m	36m
b_1	-0.028***	-0.047	-0.059	-0.009**	-0.016	-0.021	-0.778*	-0.682	-1.154
	(0.006)	(0.028)	(0.033)	(0.002)	(0.008)	(0.010)	(0.316)	(0.402)	(0.895)
	[0.000]	[0.164]	[0.124]	[0.004]	[0.107]	[0.108]	[0.039]	[0.121]	[0.245]
b_2	0.007	0.005	0.016	-0.001	-0.002	-0.001	-0.188	-0.529	0.432
	(0.011)	(0.017)	(0.020)	(0.002)	(0.003)	(0.004)	(0.402)	(0.474)	(0.610)
	[0.502]	[0.771]	[0.443]	[0.818]	[0.441]	[0.781]	[0.653]	[0.327]	[0.518]
F-test $b_1 = b_2$	0.007	0.196	0.137	0.020	0.183	0.111	0.265	0.759	0.087
T_1 obs.	237	225	217	237	225	217	115	103	95
T_2 obs.	291	291	287	291	291	287	162	162	158
R^2	0.099	0.087	0.111	0.135	0.153	0.189	0.061	0.052	0.056
Adj. R^2	0.094	0.082	0.106	0.130	0.148	0.184	0.050	0.041	0.045

Notes. This table reports the estimates of slope coefficients and F-test statistics for regime switching time series regressions of the form $rx_{t+K} = \mathbb{1}_t^R(a_1 + b_1 X_t) + (1 - \mathbb{1}_t^R)(a_2 + b_2 X_t) + v_{t+K}$, where $\mathbb{1}_t^R$ is the OECD global recession indicator, X_t is one of the forecasting variables, K is the forecasting horizon in months, and rx_{t+K} is cumulative log excess return between months t and $t+K$. Standard errors are reported in parentheses, and p-values are reported in brackets, while the F-test row reports p-value for testing $H_0 : b_1 = b_2$, all obtained using OS HAR inference. $^{+}$,*,**, and *** indicate significance at the 0.1, 0.05, 0.01, and 0.001 levels, respectively.

Chapter 3

Executive compensation in high-growth no-dividend firms

3.1 Introduction

In 2012, about 65% of stocks in the US did not pay dividends (Atmaz and Basak, 2019). While the propensity of firms to pay dividends has generally been declining, it is well known that small firms with high growth opportunities tend to not pay dividends (Fama and French, 2001; Farre-Mensa et al., 2014). Because investments of such firms tend to be constrained by internal finance (Carpenter and Petersen, 2002b), they often have to retain earnings for expansion of their businesses, and thus hardly generate any net cash flow. In addition, these firms are usually also characterised by a high degree of opaqueness and asymmetric information. Indeed, because their investments frequently embody new knowledge and are difficult to evaluate, management often have far better information about their firms' growth prospects (Cornell and Shapiro, 1988; Carpenter and Petersen, 2002a). The combination of asymmetric information and the lack of reliable performance indicators that are hard to manipulate, like net cash flow or dividends, may create acute agency problems in such firms. In this

paper I explore the tradeoffs that shareholders face when designing managerial compensation contracts in high-growth no-dividend companies, as well as their economic consequences.

Specifically, I study a financially constrained, publicly traded startup firm that undergoes a period of rapid growth. The firm has not yet reached its full productive scale and, therefore, it can profitably reinvest its operating profits instead of paying out dividends. However, the firm's shareholders face several agency problems on behalf of the firm's manager. First, the manager must exert unobserved costly effort in order to maximize the firm's growth potential. Second, he is naturally much better informed about the firm's prospects compared to its investors. Together, these two issues result in a non-trivial problem of designing managerial compensation. On the one hand, the contract should provide strong enough incentives for the manager to work hard and pursue growth. However, asymmetric information about growth prospects together with the absence of reliable performance indicators create a danger that the manager may continue to imitate strong growth even after the growth opportunities are exhausted, if this benefits his compensation. This may result in significant overinvestment and the firm stretching beyond its productive scale. I show that popular stock-based compensation contracts incentivize the manger to exert effort, but can also encourage him to conceal the end of the high growth phase from investors and continue rampant investment in the firm in order to maintain the pretence. I then consider alternative compensation schemes that may perform better, notably contracts that delay payments from stock sales.

I build on the framework developed by Benmelech et al. (2010), who study a similar trade-off of incentives in the context of an established, dividend-paying firm. While my analysis is similar to theirs, there are several important differences. First, in their model the firm already operates at the maximum productive capacity. This maximum capacity grows over time, but the firm generates enough profits to

both finance the required investments in order to keep up with it, and also pay out dividends to its shareholders. This is consequential for their results, because dividend payments are verifiable and play an important role in the optimal contract design. Second, Benmelech et al. (2010) study the "investment behaviour of the firm over a long haul," in their own words. Indeed, their quantitative analysis focuses on a firm that undergoes a prolonged period of moderately high growth,[1] rather than a startup that grows rapidly over a relatively short period of time. This affects the set of viable contracts. Lastly, the presence of dividends in Benmelech et al.'s setting makes it trickier for the manager to conceal the lack of growth opportunities from investors, who expect dividends to grow accordingly. In order to run the pretence, therefore, the manager must continue to pay ever higher dividends, eventually leading to an inevitable capital shortfall. In contrast, in the context of high-growth startups studied here, the opposite is true: investors expect to see continued reinvestment of profits into the company and no dividends while the high growth phase continues. The ability of the manager to mislead investors is, therefore, only limited by his ability to manipulate the reported performance indicators. His choice to deceive investors, then, leads to overinvestment in the firm's assets, rather than depletion of capital. The contribution of this paper, therefore, is the analysis of the trade-off between incentives provision and information revelation in the context of rapidly growing startups that do not pay dividends.

I begin by considering a stock-based compensation contract. As time passes and the firm continues to exhibit strong growth, rational investors positively update their expectation of the firm's prospects. This results in an increasing stock price, which directly benefits the manager. Once the rapid growth phase is over, the firm should stop expanding and instead begin distributing profits to the shareholders. However, the initiation of dividend payments, in fact, provides a

[1] In their calibration of the model, the high-growth regime entails a gradual expansion in the upper bound on the firm's productively deployable capital at 7% per year and lasts 15 years on average. After the growth slows down, the manager can run the pretence for further 17.5 years, making the entire episode potentially last more than 30 years.

negative signal to shareholders that there are no more new positive NPV projects left for the firm to invest in. This leads to a sharp drop in the share price and hurts the manager. It is therefore possible that stock-based compensation may induce the manager to initially work hard and push the firm's growth prospects, yet eventually conceal bad news and overinvest in the firm's assets, effectively destroying economic value. This can happen in equilibrium even though the investors are fully rational, know that the manager will eventually overinvest, and discount the stock price accordingly. Nevertheless, the absence of any bad news while the firm appears to grow strongly leads shareholders to improve their outlook of the firm's value, resulting in rising stock price and benefiting the manager. This bubble-like surge in the share price continues until the manager is not able to maintain the pretence any longer, at which point the firm's stock crashes. I characterise the conditions under which such equilibrium obtains. Since pure strategy equilibria do not always exist, I also characterise the mixed strategy equilibria in which the manager conceals with a positive probability.

These results shed light on the tech boom of the 1990s. This was the decade when stock-based compensation schemes increased dramatically in popularity, being perceived as an effective way to align managerial incentives with interests of shareholders (Hall and Liebman, 1998). This was also the decade that ended with the bursting of the Dot-com bubble and the collapses of several high profile telecom companies such as WorldCom, Global Crossing, and Adelphia. The eventual stock market crash is only part of the story, however. Inflated stock prices benefited the companies' managers, who sold their shares for hundreds of millions of dollars (Beresford et al., 2003; Fabrikant and Romero, 2002). To maintain the pretence of strong growth, the firms were making new, but increasingly unprofitable investments. Supported by a popular myth that Internet traffic was doubling every 100 days, for example, the telecom companies collectively spent trillions of dollars on expanding fibre optic communication lines to satisfy the

seemingly rapidly growing demand. This growth statistic has later been found to be severely overstated, as no more than 2.5-3% of the developed telecom networks were utilised at the time (Dreazen, 2002; Brenner, 2003). Notably, none of the big telecom names mentioned above – WorldCom, Global Crossing, or Adelphia – had ever paid dividends on their common stocks prior to their bankruptcies. (WorldCom Inc., 2001; Global Crossing Ltd., 2001; Adelphia Communications Corporation, 2001). Indeed, not paying dividends was perceived as a badge of honour during the boom, as tech companies and investors were focused on growth (Coombs, 2015).

Following the previous literature, I also explore the theoretical consequences in the present setting of using several alternative compensation schemes popular in practice. While contracts that specify flat wage and/or bonuses based on future dividend payments (i.e. future net cash flows) always effectively induce truth-telling, I find that they never incentivise high effort, and may even induce the manager to stop investing prematurely. This is a stronger result than in Benmelech et al. (2010), who find that net cash flow based bonuses may incentivise high effort in some cases. The difference arises because in the present setting the firm does not generate any net cash flows during the growth phase, so the impatient manager is willing to forego growth to start receiving payments earlier. On the other hand, granting the manager stock options creates even stronger incentives to conceal information from investors compared to simply awarding him stocks, because options put even higher weight on stock price growth.

I find, however, that delaying payments from otherwise standard stock-based compensation by a sufficient time interval can always deter the manager from concealing, while not compromising the incentives to exert effort. Because the manager is fired upon discovery of misconduct and receives no further payments, delaying payments acts as an effective financial penalty for cheating. This suggests that stock-based contracts with delayed payments could work well in the context

of high-growth startups, in contrast with Benmelech et al. (2010)'s setting, where delaying payments significantly reduces incentives to work hard.

Lastly, I investigate the performance of a mixed compensation contract suggested by Benmelech et al. (2010). Inspired by the capacity of stock-based compensation to incentivise effort, and by the ability of cash flow-based bonuses to deter cheating, they propose that mixing the two schemes might be able to achieve the first-best, high effort-reveal equilibrium.[2] A startup firm studied in this paper, however, does not generate any net cash flows during the high growth phase. A similar contract can be implemented, nevertheless, if it specifies future bonuses that will accrue to the manager once the firm starts distributing cash to the shareholders. I formally show that such a contract is always able to achieve the desired high effort-reveal outcome in the present context, whenever stock-based compensation on its own induces high effort, which is a stronger result to previously reported.

I also explore the model quantitatively and calibrate it based on the boom of the late 1990s. The calibration is meant to capture a startup that grows rapidly over the course of several years. I verify that under stock-based compensation, a high effort-conceal equilibrium obtains. Moreover, I also find that since the firm matures quickly, the shareholders strongly prefer the manager to exert high effort, even if they know that he will cheat and overinvest in equilibrium. This suggests that if, for whatever reason, investors cannot induce the manager to both work hard and reveal bad news, they would indeed willingly offer a stock-based compensation contract. However, I also demonstrate that the feasible contracts discussed above can effectively incentivise the manager to reveal bad news to investors, while still exerting high effort. First, a delay of payments from the stock-based compensation by just under two years effectively deters the manager from concealing. Vesting periods of several years are common in practice, highlighting

[2] They derive formal, though unintuitive conditions when such a contract works in their setting, and show that it is indeed the case numerically.

the viability of this contract. Second, similarly to Benmelech et al. (2010), I find that a mixed compensation scheme that puts around 20-30% weight on stocks, and the rest on future dividend-based bonuses also effectively resolves the agency problems.

Although Benmelech et al. (2010) is the most closely related work, this paper contributes to the large literature on managerial compensation. Edmans and Gabaix (2016) and Edmans et al. (2017) provide a comprehensive overview of the recent theoretical and empirical research. Since Stein (1988, 1989), it has been understood that short-termism may lead a CEO to forgo good investments in order to boost current earnings and share prices. The present paper argues that in some contexts the manager's concern with share prices can, in fact, lead to an overinvestment rather than underinvestment problem. In this regard, it is similar to Kedia and Philippon (2007), who show that misreporting low productivity firms may hire and invest excessively in order to pool with high productivity firms. Goldman and Slezak (2006), Peng and Roell (2008, 2014) also investigate optimal stock-based compensation in the presence of information manipulation. My paper adds to this research by focusing on the dynamic contracting issues specific to high-growth startups, where the focus on growth and absence of good performance indicators like net cash flows make the danger of stock price manipulation particularly salient.

3.2 The model

In this section I present a model of agency problems in a growth firm that is financially constrained and does not (yet) pay dividends. Since the model is based on Benmelech et al. (2010), I maintain their notation and simplifying assumptions whenever possible for comparison purposes.

3.2.1 Setup

Time is continuous. Consider a firm ran by a manager, and owned by a large number of investors (shareholders). The firm's operating profit at time t is given by

$$Y_t = z \cdot \min\{K_t, J\}, \tag{3.1}$$

where z is the rate of return on capital, K_t is the capital stock at time t, and J is an upper bound on deployable productive capital. The return on capital z is sufficiently high to compensate for the cost of capital:

$$z > r, \tag{3.2}$$

where r is the rate of return required by investors. For simplicity, I assume that there is no depreciation: $\delta = 0$. The firm's capital stock evolves according to

$$\dot{K}_t = I_t, \tag{3.3}$$

where $I_t \geq 0$ is investment that is irreversible.[3]

Time $t = 0$ is set a point when it is common knowledge that the firm's capital stock is $K_0 = 1$ (by normalisation), and the firm's full potential has not yet been reached, so $K_0 < J$. The maximum productive size of the firm J is a random, Pareto distributed variable with the probability density function

$$f_J(j) = \frac{\alpha}{j^{\alpha+1}}, \quad \alpha > 1, \tag{3.4}$$

and support $[1, \infty)$. This assumption reflects empirical studies on firm sizes (Axtell, 2001; Gabaix, 2009), and also represents the idea that optimal size of the firm can

[3] I assume the capital is highly firm-specific (e.g. R&D), and so is unrecoverable.

potentially be very large. Crucially, neither the manager nor the shareholders of the firm can observe the realization of J until the firm reaches it. I let τ^* denote the moment in time when the firm achieves its maximum productive size.

The main difference compared to Benmelech et al. (2010) here lies in the assumptions about the firm relative to the upper bound J. In their model, the firm is assumed to always have enough resources to invest and stay at the technological frontier (s.t. $K_t = J_t \; \forall t$), while also paying dividends. J_t grows over time, and the authors focus on a period when its growth accelerates. In contrast, in the present setting the firm starts below the maximum productive scale J. Until J is reached, therefore, the firm is in the position to make profitable investments. I assume that the firm is financially constrained and cannot raise additional capital, so the only potential source of funds for investment is retained earnings: $I_t \leq Y_t$.[4] For simplicity, here I assume that the upper bound J itself does not grow, though this assumption straightforward to relax.

There are two agency conflicts between the shareholders and the manager of the firm. First, the manager decides on the unobserved effort $e \in \{e^H, e^L\}$, $e^H > e^L$. High effort increases the expected maximum productive size of the firm. Specifically, high effort lowers the shape parameter α in (3.4), and thus increases the expected value of J:

$$\alpha^H \equiv \alpha(e^H) < \alpha^L \equiv \alpha(e^L) \quad \Longleftrightarrow \quad \mathbb{E}[J|e^H] > \mathbb{E}[J|e^L]. \tag{3.5}$$

It is costly for him to exert high effort, however. At any moment in time, the cost of high effort is

$$c^H \equiv c(e^H) > c^L \equiv c(e^L) = 0, \tag{3.6}$$

[4]The assumption that the firm is financially constrained is taken as given here. It is realistic given the extent of agency problems and information asymmetries in the context of high growth startups.

whereas the cost of low effort c^L is normalised to zero. To achieve high effort, therefore, the shareholders must design a compensation contract that provides sufficient incentives to the manager.

The second agency problem arises because the manager is much better informed about the prospects of the firm than outside investors. Specifically, I assume that he observes J the moment the firm exhausts its growth opportunities at time τ^*. Shareholders, on the other hand, do not directly observe J, τ^*, firm's profits or other performance metrics, and, instead, rely on the firm's reports. There is therefore a danger that the manager does not reveal τ^* when it arrives and continues to invest in the firm beyond its productive scale, if he stands to benefit from doing so. Although shareholders eventually discover the misconduct (more on this later), this would result in overinvestment and destruction of economic value. As in Benmelech et al. (2010), I assume that the manager is fired upon discovery of the misconduct, and receives no further payments from the shareholders.[5]

I follow Benmelech et al. (2010) and assume linear preferences of the manager:

$$U_t = \mathbb{E}_t \left[\int_t^T \mathrm{e}^{-\beta(s-t)} w_s [1 - c(e)] \, ds \right], \tag{3.7}$$

where T is the manager's departure date, β is his discount rate, and w_s is his compensation in period s. The planning horizon is without end: $T \to \infty$, unless the manager is fired. The manager is impatient: $\beta > z$, implying also $\beta > r$. This assumption captures the fact that the manager bears a lot more risk associated with the firm's performance than diversified investors; there is also substantial experimental evidence suggesting that individual discount rates are greater than the market rate of return (Frederick et al., 2002). Consequently, the manager's valuation of the firm is significantly lower than that of investors, so the sharehold-

[5] While it may be difficult for the shareholder to claim the payments that have already been made back, it is realistic to include a clause in the contract that prevents further payments if the manager is fired.

ers cannot solve the incentive problem by "selling" the firm to the manager in this setting.[6] Note that the cost of effort is specified in the multiplicative fashion: effort is costly because it does not allow the manager to enjoy his pay to the fullest. The resulting scale-invariance greatly simplifies the analysis. I also maintain the simplifying assumption of Benmelech et al. (2010) that the resources to pay the manager are coming form outside of the model, i.e. higher pay does not affect the dividends or the firm's value. This is justifiable if the manager's compensation is a relatively small fraction of the economic value of the firm.

Note that the manager incurs the cost of effort at each point of time until the limit J is reached. Yet the random variable J is drawn from the Pareto distribution with a shape parameter α, which in turn depends on the level of effort. This creates a potential logical inconsistency: what happens to J if the manager changes the amount of effort before the firm reaches its maximum productive capacity?[7] To avoid such problems, one of the two assumptions can be made. First, we can assume that the manager commits to the level of effort he chooses at time $t = 0$. This assumption is justifiable because, as we will see later, the manager's problem of choosing effort is largely independent of time. Second, we can assume that J remains a random variable until it is reached. Then, if at any point in time the manager changes his effort, the distribution of a yet unrealised random variable J also changes. Given effort level e from time t onwards, the pdf of J at time t is

$$f_{J,t}(j|e) = \frac{\alpha(e)K_t^{\alpha}}{j^{\alpha(e)+1}}, \tag{3.8}$$

with support $[K_t, \infty)$.[8]

[6] Since impatience may lead the owner-manager to also exert low effort and even disinvest from the firm, the difference in valuations goes beyond the simple difference in discount rates.

[7] The model in Benmelech et al. (2010) features a similar minor inconsistency.

[8] If effort e is constant, (3.8) is simply a Bayesian posterior of (3.4), after using the information that $J > K_t$.

3.2.2 Full information benchmark

First, consider the benchmark case in which the shareholders observe all the firm's financials and τ^*. The assumption that $z > r$ implies that it is economically optimal for the firm to invest up to its fullest potential, i.e. until $K = J$. To maximise the value of the firm, the manager must thus reinvest all profits zK_t every period, until the firm reaches its maximum productive scale. This optimal investment policy implies that the firm's capital is

$$K_t = \begin{cases} \mathrm{e}^{zt} & \text{for } t < \tau^* \\ J = \mathrm{e}^{z\tau^*} & \text{for } t \geq \tau^*. \end{cases} \tag{3.9}$$

Using (3.9), it is straightforward to show via the transformation of the random variable J that the time τ^* at which the firm reaches its maximum productive size is exponentially distributed with the rate parameter $\lambda = \alpha z$ and pdf:

$$f_{\tau^*}(\tau^* | \tau^* > t) = \lambda \mathrm{e}^{-\lambda(\tau^* - t)}. \tag{3.10}$$

This result is similar to the assumption in Benmelech et al. (2010), where the duration of the accelerated growth period is also exponentially distributed. Here, however, the exponential distribution follows naturally from the empirically established distribution of firm sizes, and the optimal investment strategy. To keep things as close to Benmelech et al. as possible, in what follows I mostly focus the random variable τ^* instead of J. Note that the effect of the manager exerting high effort in (3.5) could equivalently be restated as delaying the expected time τ^* at which the firm reaches its full potential if it reinvests all its profits:

$$\lambda^H \equiv \lambda(e^H) < \lambda^L \equiv \lambda(e^L) \quad \Longleftrightarrow \quad \mathbb{E}[\tau^* | e^H] > \mathbb{E}[\tau^* | e^L]. \tag{3.11}$$

Starting from τ^*, the firm generates cash flow that cannot be productively reinvested, and thus should be paid out to the shareholders as dividends. If the firm follows the optimal investment plan, its dividend stream should be, therefore:

$$D_t = zK_t - I_t = \begin{cases} 0 & \text{for } t < \tau^* \\ zJ = ze^{z\tau^*} & \text{for } t \geq \tau^*, \end{cases} \tag{3.12}$$

i.e. the firm only begins dividend payments after the growth phase is over. Having established the dividend process, we can now derive the stock price, i.e. the expected value of the firm to its shareholders:[9]

Proposition 3.1. *Under the full information (indicated by "fi"), the price of the firm after τ^*, when it reaches its maximum productive capacity is*

$$P_{\text{fi},t}^{\text{after}} = \frac{ze^{z\tau^*}}{r} \quad \text{for } t \geq \tau^*. \tag{3.13}$$

Given the manager's effort e and thus the rate parameter $\lambda = z\alpha(e)$, the value of the firm before τ^* is

$$P_{\text{fi},t}^{\text{before}} = A_\lambda^{\text{fi}} e^{zt} \quad \text{for } t < \tau^*, \tag{3.14}$$

where

$$A_\lambda^{\text{fi}} \equiv \frac{\lambda z}{r(r+\lambda-z)}. \tag{3.15}$$

Proof. Once τ^* arrives, the value of the firm to the shareholders is simply the

[9] I normalize the amount of outstanding shares to 1, so the terms "stock price" and "firm value" are used interchangeably.

present value of the constant dividend stream given in (3.12):

$$P_{\text{fi},t}^{\text{after}} = \int_{s=t}^{\infty} \mathrm{e}^{-r(s-t)} D_s \, ds = \frac{zJ}{r} = \frac{z\mathrm{e}^{z\tau^*}}{r} \quad \text{for } t \geq \tau^*. \tag{3.16}$$

Prior to τ^*, and taking the manager's effort and the rate parameter $\lambda = z\alpha(e)$ as given, the valuation of the firm by rational investors under full information takes into account the uncertainty regarding time τ^* when the firm will start paying dividends, as well as the expected size of the future dividend stream:

$$P_{\text{fi},t}^{\text{before}} = \mathbb{E}_t\left[\mathrm{e}^{-r(\tau^*-t)} P_{\text{fi},\tau^*}^{\text{after}}\right] = \int_{s=t}^{\infty} \mathrm{e}^{-r(s-t)} \frac{z\mathrm{e}^{zs}}{r} \lambda \mathrm{e}^{-\lambda(s-t)} \, ds = A_\lambda^{\text{fi}} \mathrm{e}^{zt} \quad \text{for } t < \tau^*. \tag{3.17}$$

Q.E.D.

Prior to τ^*, the firm's stock price grows at the rate z because investors positively update their beliefs about the firm's growth prospects as time passes. At τ^*, however, when the firm exhausts its growth opportunities, the stock price falls by

$$P_{\text{fi},\tau^*}^{\text{before}} - P_{\text{fi},\tau^*}^{\text{after}} = \frac{z(z-r)\mathrm{e}^{z\tau^*}}{r(r+\lambda-z)}, \tag{3.18}$$

and remains constant thereafter. The stock price thus exhibits a bubble-like behaviour, i.e. rapid growth followed by a crash. This happens even under the full information, because the arrival of τ^* constitutes a negative shock to the investors' expectations about future prospects of the firm.

Corollary 3.1. *Prior to τ^* under full information, the firm value is always higher under high managerial effort than under low effort:*

$$P_{\text{fi},t}^{\text{before}}(\lambda^H) > P_{\text{fi},t}^{\text{before}}(\lambda^L) \quad \text{for } t < \tau^*. \tag{3.19}$$

Proof. The pricing multiplier A_λ^{fi} defined in (3.15) decreases in λ since $r < z$. The

inequality follows because $\lambda^H < \lambda^L$, so $A^{\text{fi}}_{\lambda^H} > A^{\text{fi}}_{\lambda^L}$. *Q.E.D.*

Because high managerial effort increases the firm value prior to τ^*, and the shareholders always prefer the manager to exert high effort.[10] They would specify high effort in the contract if it was verifiable.

3.3 Stock-based compensation and overinvestment under asymmetric information

3.3.1 Stock-based compensation with unobserved effort

If the manager's effort is unobservable, how can the shareholders incentivise the manager to exert high effort? Similarly to Benmelech et al. (2010) and much previous literature on managerial compensation, I show that stock-based compensation may provide the necessary incentives to work hard in the present setting. Suppose that investors have full information about the firm's performance (profits, capital stock, and τ^* when it arrives etc.), with the exception of the manager's effort. Consider a simple compensation scheme in which the manager receives η shares of the company every instant. Because the manager is impatient, he always sells the shares immediately,[11] so his effective compensation at time t is

$$w_t = \eta P_t. \tag{3.20}$$

Proposition 3.2. *Suppose investors have full information about the firm's performance, and only the manager's effort is unobserved. An equilibrium in which the manager*

[10] In the present model it is not possible to infer the social first best when taking the manager's welfare into account. Firstly, the resources to pay the manager are coming form outside of the model. Secondly, because of the multiplicative specification of the manager's preferences, the absolute cost of high effort depends on the overall level of the manager's compensation.

[11] In this model managerial sell-off thus provides no signal to the market.

exerts high effort under the stock-based compensation (3.20) *occurs iff*

$$\frac{\lambda^L - \lambda^H}{\beta + \lambda^L - z}(1 - \Theta_{\lambda^H}) > c^H, \tag{3.21}$$

where

$$\Theta_\lambda \equiv \frac{(\beta - z)(r + \lambda - z)}{\beta\lambda} \in (0, 1). \tag{3.22}$$

Proof. Plugging compensation (3.20) into the manager's utility (3.7), and taking market valuations (3.13) and (3.14) as given, the manager's expected utility at time $t < \tau^*$ as a function of effort is:[12]

$$\begin{aligned} U_t(e) &= \mathbb{E}_t\left[\int_t^{\tau^*} \mathrm{e}^{-\beta(s-t)}\eta A_\lambda^{\mathrm{fi}}\mathrm{e}^{zs}(1 - c(e))\, ds + \int_{\tau^*}^{\infty} \mathrm{e}^{-\beta(s-t)}\eta\frac{z\mathrm{e}^{z\tau^*}}{r}\, ds\right] \\ &= \frac{\eta\mathrm{e}^{zt}}{\beta + \lambda(e) - z}\left[A_\lambda^{\mathrm{fi}}(1 - c(e)) + \frac{z}{\beta r}\lambda(e)\right]. \end{aligned} \tag{3.23}$$

The manager will exert high effort as long as $U_t(e^H) > U_t(e^L)$, which yields

$$\frac{\beta + \lambda^L - z}{\beta + \lambda^H - z} > \frac{A_\lambda^{\mathrm{fi}} + \frac{z}{\beta r}\lambda^L}{A_\lambda^{\mathrm{fi}}(1 - c^H) + \frac{z}{\beta r}\lambda^H}. \tag{3.24}$$

It follows that a high-effort equilibrium occurs iff (3.24) is satisfied for $A_{\lambda^H}^{\mathrm{fi}}$,[13] which after substitution yields the condition (3.21). *Q.E.D.*

Examining the condition (3.21), we see that the high-effort equilibrium occurs as long as the cost of effort is not too high, and the impact that high effort makes on expected τ^* is sufficiently large. To keep the model interesting, I assume that this is always satisfied.

[12] Of course, the manager will optimally not exert high effort after τ^*.

[13] A low-effort equilibrium occurs iff (3.24) is not satisfied for $A_{\lambda^L}^{\mathrm{fi}}$.

3.3.2 Conceiling and overinvestment under stock based compensation and asymmetric information

I now show, however, that if investors do not directly observe τ^* and also do not have access to reliable performance indicators, stock-based compensation may induce the manager to conceal τ^* when it arrives, and overinvest in the firm's assets. This happens despite the assumption that the manager is fired upon discovery of the misconduct.

If the manager chooses to conceal at τ^*, he will try to run the pretence for as long as possible. This is because he is fired and loses any further payments when the shareholders learn about his misconduct. I assume that he can run the pretence for a fixed period h^{**}, such that the shareholders discover the overinvestment at $T^{**} = \tau^* + h^{**}$. This assumption could be microfounded in several ways. For example, suppose the manager can only run the pretence at a cost, which is proportional to the size of the "hole" in the firm's accounts, i.e. the difference between the reported and actual operating profit at $t > \tau^*$. The manager then optimally times his exit after h^{**} periods, when the marginal cost of running the pretence outweighs the marginal benefit.[14] Note that the concealing manager has to report growing operating profits ze^{zt} after τ^* to prevent the investors from realising that the growth phase is over. The manager also cannot initiate dividends payments without revealing information to the market, so, assuming he cannot easily loot the firm's operating profits, he continues to (unproductively) reinvest them into new projects.

If investors expect the manager to conceal information, they will rationally take this into account when valuing the firm. Specifically, their valuation prior to

[14]Formally, suppose the manager's utility is $U_t = \mathbb{E}_t\left[\int_t^T e^{-\beta(s-t)} w_s(1-c(e))(1-c_t^p)\,ds\right]$, where the cost of misleading the investors c_t^p is proportional to the discrepancy between the firm's reported and actual operating profit: $c_t^p = \chi\left(e^{z(t-\tau^*)}-1\right)$, $\chi > 0$. The manager will optimally exit at T^{**} when $c_{T^{**}}^p = 1$, which yields $h^{**} \equiv T^{**} - \tau^* = \frac{1}{z}\ln\frac{\chi+1}{\chi}$.

T^{**} is discounted compared to the full information case, because investors know that they will learn about J and τ^* only with a delay h^{**}, and they cannot rule out that the growth phase is already over when pricing the shares at $t < T^{**}$.

Proposition 3.3. *Take the manager's effort e and thus the rate parameter λ as given. If information is asymmetric (indicated by "ai") and investors expect the manager to conceal the bad news about growth, the stock price at any $t \in [h^{**}, T^{**})$ is*[15]

$$P_{\mathrm{ai},t} = A_\lambda^{\mathrm{ai}} \mathrm{e}^{zt}, \tag{3.25}$$

where

$$A_\lambda^{\mathrm{ai}} \equiv \frac{\lambda z \mathrm{e}^{-zh^{**}}}{r(r+\lambda-z)} = A_\lambda^{\mathrm{fi}} \mathrm{e}^{-zh^{**}}. \tag{3.26}$$

Proof. As under the full information, investors take into account the uncertainty regarding τ^* and the size of the future dividends. However, under asymmetric information, they also know that they will only begin to receive dividends starting from $\tau^* + h^{**}$. In addition, their belief about τ^* at any $t \in [h^{**}, T^{**})$ has support $(t - h^{**}, \infty)$. Therefore,

$$P_{\mathrm{ai},t} = \mathbb{E}_t \left[\mathrm{e}^{-r(\tau^*+h^{**}-t)} \frac{z\mathrm{e}^{z\tau^*}}{r} \right] = \int_{t-h^{**}}^{\infty} \mathrm{e}^{-r(s+h^{**}-t)} \frac{z\mathrm{e}^{zs}}{r} \lambda \mathrm{e}^{-\lambda(s+h^{**}-t)} \, ds = A_\lambda^{\mathrm{ai}} \mathrm{e}^{zt}. \tag{3.27}$$

Q.E.D.

The stock price also exhibits a bubble-like behaviour when the manager is expected to conceal bad news: it grows at the high rate z prior to T^{**}, and then crashes, similarly to the full information case. Because of the asymmetric information discount, the stock price is initially smaller compared to the case in which

[15] The case $0 < t < h^{**}$ is more complicated and does not yield additional economic intuition. The integral starts at 0 rather than $t - h^{**}$, yielding $P_{\mathrm{ai},t} = A_\lambda^{\mathrm{ai}} \mathrm{e}^{(z-r)h^{**}} \mathrm{e}^{rt}$ for $t < h^{**}$. The price prior to h^{**} grows slower, because investors do not learn any new information as time passes; yet it is also relatively higher, because the manager did not yet have much time to cheat: $A_\lambda^{\mathrm{ai}} \mathrm{e}^{(z-r)h^{**}} \mathrm{e}^{rt} > A_\lambda^{\mathrm{ai}} \mathrm{e}^{zt}$.

investors receive full information ($A_\lambda^{\text{ai}} < A_\lambda^{\text{fi}}$). However, it grows for longer before the crash, and this surge is accompanied by overinvestment in the firm's assets.

Let us define reveal and conceal equilibria as equilibria in which the manager either reveals or conceals the bad news from investors, respectively. To determine which equilibrium will prevail under the stock-based compensation (3.20), we can plug the pricing functions (3.13), (3.14), and (3.27) into the manager's utility (3.7), to get his expected utilities at τ^* from choosing the reveal and conceal strategies, respectively. The following result follows:

Proposition 3.4. *Take the manager's effort e and thus the rate parameter λ as fixed, and define*

$$A^{\text{mixed}} \equiv \frac{z}{\beta r} \frac{\beta - z}{1 - \mathrm{e}^{-(\beta - z)h^{**}}}. \tag{3.28}$$

The conceal equilibrium obtains under the stock-based compensation (3.20) *for $\tau^* \geq h^{**}$ iff*[16]

$$A_\lambda^{\text{ai}} > A^{\text{mixed}}. \tag{3.29}$$

The reveal equilibrium obtains iff

$$A_\lambda^{\text{fi}} < A^{\text{mixed}}. \tag{3.30}$$

Because $A_\lambda^{\text{fi}} > A_\lambda^{\text{ai}}$, the two equilibria are mutually exclusive. Finally, a mixed strategy equilibrium obtains under the stock-based compensation contract iff

$$A^{\text{mixed}} \in \left[A_\lambda^{\text{ai}}, A_\lambda^{\text{fi}}\right]. \tag{3.31}$$

[16] The case $\tau^* < h^{**}$ is far more complicated and does not yield additional economic intuition. Nevertheless, because $P_{\text{ai},t} > A_\lambda^{\text{ai}} \mathrm{e}^{zt}$ for $t < h^{**}$, the manager's incentive to conceal is even greater, so (3.29) remains a sufficient condition for the conceal equilibrium.

In the mixed strategy equilibrium, the manager conceals with probability

$$\pi = \frac{1}{1 - \mathrm{e}^{-zh^{**}}} \left(1 - \frac{\Theta_\lambda}{1 - \mathrm{e}^{-(\beta - z)h^{**}}} \right) \in [0, 1], \tag{3.32}$$

where Θ_λ *is defined in* (3.22)*, and the firm's value prior to the revelation is*

$$P_{\text{mixed},t} = A^{\text{mixed}} \mathrm{e}^{zt} \quad \text{for } t \geq h^{**}. \tag{3.33}$$

Proof. Take the level of effort and λ as given. Note that $(1 - c(e))$ is thus a constant that we can leave out of the expression for the manager's utility for the sake of conciseness. First, the manager's utility from choosing to reveal τ^* when investors (correctly) expect him to reveal is

$$U^{\text{reveal}}_{\text{Stock},\tau^*} = \int_{\tau^*}^{\infty} \mathrm{e}^{-\beta(s-\tau^*)} \eta \frac{z\mathrm{e}^{z\tau^*}}{r} \, ds = \frac{\eta z \mathrm{e}^{z\tau^*}}{\beta r}. \tag{3.34}$$

Second, the manager's utility from choosing to conceal τ^* when investors (correctly) expect him to conceal is

$$U^{\text{conceal, ai}}_{\text{Stock},\tau^*} = \int_{\tau^*}^{\tau^*+h^{**}} \mathrm{e}^{-\beta(s-\tau^*)} \eta A^{\text{ai}}_\lambda \mathrm{e}^{zs} \, ds = \eta A^{\text{ai}}_\lambda \mathrm{e}^{z\tau^*} \left(\frac{1 - \mathrm{e}^{-(\beta - z)h^{**}}}{\beta - z} \right). \tag{3.35}$$

Lastly, the manager's utility from choosing to conceal τ^* when investors (incorrectly) expect him to reveal is

$$U^{\text{conceal, fi}}_{\text{Stock},\tau^*} = \int_{\tau^*}^{\tau^*+h^{**}} \mathrm{e}^{-\beta(s-\tau^*)} \eta A^{\text{fi}}_\lambda \mathrm{e}^{zs} \, ds = \eta A^{\text{fi}}_\lambda \mathrm{e}^{z\tau^*} \left(\frac{1 - \mathrm{e}^{-(\beta - z)h^{**}}}{\beta - z} \right). \tag{3.36}$$

For a given λ, the conceal equilibrium obtains iff $U^{\text{conceal, ai}}_{\text{Stock},\tau^*} > U^{\text{reveal}}_{\text{Stock},\tau^*}$, which yields the condition (3.29). Similarly, the reveal equilibrium obtains iff $U^{\text{conceal, fi}}_{\text{Stock},\tau^*} < U^{\text{reveal}}_{\text{Stock},\tau^*}$, which gives (3.30). Because $A^{\text{fi}}_\lambda > A^{\text{ai}}_\lambda$, the two equilibria are mutually exclusive.

What if neither condition is satisfied? It is straightforward to show that for

the intermediate case $U^{\text{reveal}}_{\text{Stock},\tau^*} \in \left[U^{\text{conceal, ai}}_{\text{Stock},\tau^*}, U^{\text{conceal, fi}}_{\text{Stock},\tau^*}\right]$, which after substitution is equivalent to (3.31), no pure strategy equilibrium obtains. Suppose a mixed strategy equilibrium obtains in which the manager conceals with some probability $\pi \in [0,1]$, and investors value the firm at $P_{\text{mixed},t} = A^{\text{mixed}}\mathrm{e}^{zt}$ at $t \geq h^{**}$ prior to the revelation. The manager's utility from choosing to conceal under such valuation at $\tau^* \geq h^{**}$ is

$$U^{\text{conceal, mixed}}_{\text{Stock},\tau^*} = \int_{\tau^*}^{\tau^*+h^{**}} \mathrm{e}^{-\beta(s-\tau^*)}\eta A^{\text{mixed}}\mathrm{e}^{zs}\,ds = \eta A^{\text{mixed}}\mathrm{e}^{z\tau^*}\left(\frac{1-\mathrm{e}^{-(\beta-z)h^{**}}}{\beta-z}\right). \quad (3.37)$$

For the manager to be indeed willing to randomise, it must be the case that $U^{\text{conceal, mixed}}_{\text{Stock},\tau^*} = U^{\text{reveal}}_{\text{Stock},\tau^*}$, which immediately implies that A^{mixed} must be as defined in (3.28).

For the rational investors with equilibrium-consistent beliefs to indeed value the firm at the price given in (3.33), it must be the case that

$$A^{\text{mixed}} = \pi A^{\text{ai}}_{\lambda} + (1-\pi)A^{\text{fi}}_{\lambda}, \quad (3.38)$$

which we can solve for π to obtain (3.32). Using (3.31), we can see that $\pi \in [0,1]$, as required. *Q.E.D.*

Let us know turn to the issue of effort provision by the manager.

3.3.3 Choice of effort under asymmetric information

Following the conceal strategy does not mean that the manager supplies low effort. Indeed, exerting high effort allows him to enjoy high and rapidly growing compensation for longer. I show this formally in the following proposition:

Proposition 3.5. *Let $t \geq h^{**}$ and suppose that $A^{\text{fi}}_{\lambda^H} \geq A^{\text{mixed}}$. Then, under the stock-based compensation, there is an equilibrium in which the manager initially exerts high*

effort, but then conceals the bad news or mixes conceal/reveal strategies iff

$$\frac{\lambda^L - \lambda^H}{\beta + \lambda^L - z} e^{-(\beta - z)h^{**}} > c^H. \tag{3.39}$$

Proof. Let us separately consider the conceal and mixed equilibria cases. Similarly to the full information case, we can express the manager's utility in the conceal equilibrium as a function of effort:

$$\begin{aligned} U_t(e) &= \mathbb{E}_t \left[\int_t^{\tau^*} e^{-\beta(s-t)} \eta A^{\text{ai}}_{\lambda^H} e^{zs} (1 - c(e))\, ds + e^{-\beta(\tau^* - t)} U^{\text{conceal, ai}}_{\text{Stock},\tau^*} \right] \\ &= \frac{\eta A^{\text{ai}}_{\lambda^H} e^{zt}}{\beta + \lambda(e) - z} \left(1 - c(e) + \lambda(e) \frac{1 - e^{-(\beta - z)h^{**}}}{\beta - z} \right). \end{aligned} \tag{3.40}$$

The manager then indeed supplies high effort in the conceal equilibrium iff $U_t(e^H) > U_t(e^L)$, which yields the condition (3.39).

Turning to the mixed equilibrium case, note that the manager's utility as a function of effort is similar to (3.23) in the symmetric information case, if we replace A^{fi}_{λ} with A^{mixed}. This is because the manager is indifferent between revealing and concealing at τ^*. The manager thus exerts high effort as long as

$$\frac{\beta + \lambda^L - z}{\beta + \lambda^H - z} > \frac{A^{\text{mixed}} + \frac{z}{\beta r} \lambda^L}{A^{\text{mixed}}(1 - c^H) + \frac{z}{\beta r} \lambda^H}, \tag{3.41}$$

which after substituting in A^{mixed} from (3.28) and simplifying yields (3.39) once again. *Q.E.D.*

Examining condition (3.39), we see that the manager exerts high effort in conceal and mixed equilibria as long as the cost of effort is not too high, and the impact that high effort makes on expected τ^* is not too small. In addition, the manager will not exert high effort if the deception period h^{**} is sufficiently long. Intuitively, since the manager is impatient, he cares less and less about additional benefit from exerting high effort as h^{**} increases, since, even if τ^*

arrives immediately, he still receives compensation for h^{**} periods if he chooses to conceal.

To summarise, stock-based compensation may induce the manager to exert high effort, but also not disclose bad news about the firm's growth prospects in equilibrium. Indeed, it is intuitive that stock-based compensation is likely to lead the manager to conceal: the manager is effectively "punished" for revealing τ^* by the immediate drop in share price.

3.4 Alternative compensation schemes

In this section I explore the implications in the present setting of alternative popular compensation schemes that are widely discussed in the literature and applied in practice. I follow Benmelech et al. (2010) and include flat wage, deferring compensation, options, cash flow-based bonuses, and also mixing different compensation schemes. Notably, I find that delaying payments from stock-based compensation can deter the manager from concealing, without compromising the incentives to exert effort.

3.4.1 Flat wage

Suppose the manager receives a fixed payment w every period, so his compensation is

$$w_t = w \quad \forall t. \tag{3.42}$$

Proposition 3.6. *Under the flat wage compensation scheme, the manager truthfully reveals τ^* when the firm exhausts its growth opportunities. However, he never exerts high effort.*

Proof. If the manager chooses to conceal τ^*, he continues to receive exactly the

same compensation as if he reveals for a while, but is fired at T^{**}. Therefore, revealing is strictly preferred, since it allows the manager to keep receiving the wage forever. Clearly, the manager has no incentive to exert high effort, since it is costly, but does not affect his compensation. *Q.E.D.*

Would the shareholders always prefer truthful revelation to high effort, i.e. flat wage to stock-based compensation explored in the previous section? The answer is no.

Corollary 3.2. *Investors prefer high effort to truthful revelation as long as*

$$\frac{1-\frac{z-r}{\lambda^L}}{1-\frac{z-r}{\lambda^H}} > \mathrm{e}^{zh^{**}}, \tag{3.43}$$

i.e. if high effort has a sufficiently large effect on expected τ^* (and thus expected J), and the deception period h^{**} is not too long.

Proof. Investors prefer high effort to truthful revelation as long as $A^{\mathrm{ai}}_{\lambda^H} > A^{\mathrm{fi}}_{\lambda^L}$. Condition (3.43) follows. *Q.E.D.*

Therefore, similarly to Benmelech et al. (2010), in the present setting investors might willingly prefer to compensate the manager with company shares to incentivize high effort, even if this eventually leads to cheating and overinvestment, if they are unable to induce the desired high effort-reveal behaviour.

3.4.2 Deferred compensation: vesting

Suppose the manager is awarded η shares per period, as under the stock-based compensation (3.20), but these shares have a vesting period k, following which he can sell them at the market price. Suppose also that if the manager conceals, he loses all non-vested shares when the shareholders fire him at T^{**}.

How does vesting affect incentives? First, if $\tau^* > k$, the manager has already accumulated shares that become eligible for vesting every instant at that point.

This implies that his problem at τ^* is in fact exactly the same as under standard stock-based compensation, so vesting does not reduce his incentives to conceal bad news. If $\tau^* < k$, the manager does not yet have shares vesting at τ^*, and his incentive to conceal is reduced, as desired. However, this result relies on (i) the manager having no shares eligible for vesting at $t = 0$, which may not be the case, and (ii) τ^* being sufficiently small ($< k$), which is against the desire of the shareholders to push the manager to exert high effort and delay τ^* as far as possible. For these reasons, similarly to Benmelech et al. (2010), managerial compensation based on vesting is unlikely to resolve both agency problems also in the present setting.

3.4.3 Deferred compensation: delayed payments

A popular variation of deferred compensation is to let the manager sell his shares immediately, but delay the payment for k periods, so his effective compensation at t is

$$w_t = \eta P_{t-k}. \tag{3.44}$$

Proposition 3.7. *The desired high effort/reveal equilibrium obtains under the delayed payments compensation scheme for $t \in [k, \tau^*]$ iff*[17]

$$\frac{\lambda^L - \lambda^H}{\beta + \lambda^L - z} \mathrm{e}^{-(\beta - z)h^{**}} > c^H. \tag{3.45}$$

The minimum delay period should be[18]

$$\underline{k} \equiv h^{**} + \frac{\ln(1 - \Theta_{\lambda^H})}{\beta - z}. \tag{3.46}$$

[17] The case t<k is considerably more complex, and is unlikely to yield additional economic intuition. In addition, it is different from the case $t \geq k$ considered here only if we assume that the manager is not awarded any shares prior to time $t = 0$.

[18] Note that because $\Theta_{\lambda^H} \in (0, 1)$, $\underline{k} < h^{**}$.

Proof. First, let us suppose that the manager exerts high effort and thus take the rate parameter λ^H as given, and focus on the manager's reveal vs. conceal decision under the delayed payments compensation scheme. The manager's utility from revealing τ^* at the time it arrives is

$$\begin{aligned} U^{\text{reveal}}_{\text{Delayed},\tau^*} &= \int_{\tau^*}^{\tau^*+k} \mathrm{e}^{-\beta(s-\tau^*)} \eta A^{\text{fi}}_{\lambda^H} \mathrm{e}^{z(s-k)}\, ds \;+\; \int_{\tau^*+k}^{\infty} \mathrm{e}^{-\beta(s-\tau^*)} \eta \frac{z\mathrm{e}^{z\tau^*}}{r}\, ds \\ &= \eta \mathrm{e}^{z(\tau^*-k)} \left[A^{\text{fi}}_{\lambda^H} \frac{1-\mathrm{e}^{-(\beta-z)k}}{\beta-z} + \frac{z}{\beta r} \mathrm{e}^{-(\beta-z)k} \right]. \end{aligned} \tag{3.47}$$

His utility from concealing at τ^* if investors (incorrectly) expect him to reveal is

$$U^{\text{conceal,fi}}_{\text{Delayed},\tau^*} = \int_{\tau^*}^{\tau^*+h^{**}} \mathrm{e}^{-\beta(s-\tau^*)} \eta A^{\text{fi}}_{\lambda^H} \mathrm{e}^{z(s-k)}\, ds = \eta \mathrm{e}^{z(\tau^*-k)} A^{\text{fi}}_{\lambda^H} \frac{1-\mathrm{e}^{-(\beta-z)h^{**}}}{\beta-z}. \tag{3.48}$$

The manager reveals iff $U^{\text{reveal}}_{\text{Delayed},\tau^*} \geq U^{\text{conceal,fi}}_{\text{Delayed},\tau^*}$, which becomes

$$\frac{z}{\beta r} \geq A^{\text{fi}}_{\lambda^H} \frac{1-\mathrm{e}^{-(\beta-z)(h^{**}-k)}}{\beta-z}. \tag{3.49}$$

Since the right-hand side decreases in k, the smallest k that induces the manager to reveal is the one that makes (3.49) hold with equality. Substituting in $A^{\text{fi}}_{\lambda^H}$ and solving for k yields $\underline{k}$ as defined in (3.46).

Now, let us suppose that k is set sufficiently high so that reveal equilibrium obtains, and focus on the effort provision. Suppose the investors expect the manager to exert high effort. The manager's expected utility as a function of effort is

$$\begin{aligned} U_t(e) &= \mathbb{E}_t \left[\int_t^{\tau^*} \mathrm{e}^{-\beta(s-t)} \eta A^{\text{fi}}_{\lambda^H} \mathrm{e}^{z(s-k)} (1-c(e))\, ds + \mathrm{e}^{-\beta(\tau^*-t)} U^{\text{reveal}}_{\text{Delayed},\tau^*} \right] \\ &= \frac{\eta A^{\text{fi}}_{\lambda^H}(1-c(e))}{\beta-z} \mathrm{e}^{z(t-k)} + \frac{\eta\lambda(e)\mathrm{e}^{z(t-k)}}{\beta+\lambda(e)-z} \left[A^{\text{fi}}_{\lambda^H} \frac{c(e)-\mathrm{e}^{-(\beta-z)k}}{\beta-z} + \frac{z}{\beta r} \mathrm{e}^{-(\beta-z)k} \right] \end{aligned} \tag{3.50}$$

Replacing $\frac{z}{\beta r}$ with $A^{\text{fi}}_{\lambda^H} \frac{1-\mathrm{e}^{-(\beta-z)(h^{**}-\underline{k})}}{\beta-z}$ and rearranging, this simplifies to

$$U_t(e) = \frac{\eta A^{\text{fi}}_{\lambda^H} \mathrm{e}^{z(t-k)}}{\beta + \lambda(e) - z} \left(1 - c(e) + \frac{1 - \mathrm{e}^{-(\beta-z)(h^{**}-\underline{k}+k)}}{\beta - z} \right) \tag{3.51}$$

The manager indeed provides high effort iff $U(c^H) > U(c^L)$, which yields

$$\frac{\lambda^L - \lambda^H}{\beta + \lambda^L - z} \mathrm{e}^{-(\beta-z)(h^{**}-\underline{k}+k)} > c^H. \tag{3.52}$$

Setting k as small as possible while still ensuring the reveal equilibrium, i.e. $k = \underline{k}$, we obtain the condition (3.45). *Q.E.D.*

Note that the condition (3.45) is exactly the same as the condition (3.39) for high effort provision under the standard stock-based compensation (3.20) in the conceal/mixed equilibria. Therefore, the shareholders can always achieve the desired outcome by simply delaying the payments from stock-based compensation sufficiently long. The intuition is that while such delay does not affect incentive to provide effort,[19] it acts as an effective deterrent from cheating. Indeed, if the manager conceals and is fired, he effectively looses his compensation for working between time $T^{**} - k$ and T^{**}. This is acts like a financial penalty for concealing, although it is likely easier to implement compared to claiming money back from the manager. [20]

It is worth pointing out that the condition in (3.45) is significantly less restrictive than the one found in Benmelech et al. (2010) for the delayed payments compensation scheme to work their context. This provides hope that such compensation scheme can be effective in the present setting of a high-growth startup. I leave further analysis until Section 3.5.

[19] This is partly due to the multiplicative specification of the manager's utility (3.7), i.e. the assumotion that the cost of effort is proportional to the current pay.

[20] In principle, it is intuitive that clawback clauses can always resolve the agency problems in this model by imposing a sufficiently large penalty for the misconduct at T^{**}, if the overinvestment is verifiable. However, in real life the non-optimal investment strategy of the manager may be very hard verify and prove in court (Benmelech et al., 2010).

3.4.4 Option-based compensation

Option-based compensation creates even stronger incentive to conceal than pure stock-based compensation in this model. This is because stock options reward the manager for growth in the price of shares, rather than for their level. Formally, suppose the manager is awarded η stock options every period, their time to maturity is k, and strike price is the stock price at the time of issuance (i.e. options issued "at the money"). The manager's compensation is then

$$w_t = \eta \max(P_t - P_{t-k}, 0). \tag{3.53}$$

Proposition 3.8. *Given the managers effort e and thus the rate parameter λ, define*

$$k^*_{\text{fi}} \equiv \frac{1}{z} \ln\left(\frac{\lambda}{r+\lambda-z}\right). \tag{3.54}$$

The manager never reveals bad news in equilibrium if $k \leq k^*_{\text{fi}}$.

Proof. If the manager reveals bad news at τ^*, the stock price falls to $P^{\text{after}}_{\text{fi},t}$ in (3.13) and stops growing. Therefore, the manager is not going to receive any positive payments on the options granted after τ^*. His utility from revealing τ^* is thus:

$$U^{\text{reveal, fi}}_{\text{Option},\tau^*} = \int_{\tau^*}^{\tau^*+k} \mathrm{e}^{-\beta(s-\tau^*)} \eta \max\left[\frac{z\mathrm{e}^{z\tau^*}}{r} - A^{\text{fi}}_{\lambda}\mathrm{e}^{z(s-k)}, 0\right] ds. \tag{3.55}$$

The first term in the square brackets is less than 0 at the upper limit of the integral, so the manager receives no payment at $\tau^* + k$. For his utility from choosing to reveal to be positive, the payment must at least be positive at τ^*. For that we require

$$\frac{z\mathrm{e}^{z\tau^*}}{r} > A^{\text{fi}}_{\lambda}\mathrm{e}^{z(\tau^*-k)}, \tag{3.56}$$

which after substituting in A_λ^{fi} and rearranging yields

$$k > k_{\text{fi}}^* \equiv \frac{1}{z} \ln\left(\frac{\lambda}{r+\lambda-z}\right). \tag{3.57}$$

This is the necessary condition for the manager to even consider revealing bad news over concealing them, since the latter strategy yields him strictly positive utility.[21] *Q.E.D.*

The case $k > k_{\text{fi}}^*$ is very complex mathematically. It is clear, however, that $k > k_{\text{fi}}^*$ by no means guarantees the reveal equilibrium. Rather, as k increases, the manager's conceal/reveal problem converges to his problem under the standard stock-based compensation. Therefore, if stock-based compensation fails to achieve the first best, option-based compensation is even more likely to fail.

3.4.5 Cash flow-based (bonus) compensation

Let us now consider a compensation scheme in which the manager is remunerated based on the net cash flow (NCF) generated by the firm, i.e. cash flow after all expenses including investment. NCF is harder to manipulate than operating profit. Moreover, because there is no rationale for the firm to retain cash in this model, once the firm begins to generate positive NCF, it should distribute it to the shareholders. NCF, thus, coincides with the dividends paid by the firm, D_t. Suppose, therefore, that the manager's pay is:

$$w_t = \eta_D D_t. \tag{3.58}$$

Under such compensation the manager is not paid at all while the firm is growing and actively investing, and starts receiving compensation only when the firm initiates dividend payments. It immediately follows that the manager never

[21] Provided either that $\tau^* + h^{**} < k$, or that the manager receives options for some time prior to $t = 0$.

conceals bad news: if he does so, his total compensation is 0, because he is fired when investors discover misconduct.

Proposition 3.9. *Under the cash flow-based compensation scheme, the manager always truthfully reveals τ^*. However, the manager never exerts high effort.*

Proof. If the manager reveals and initiates dividend payments, he receives positive pay, and his utility is thus strictly positive. If the manager conceals, he does not receive any payments while running the pretence (since NCF is 0), and is fired the moment shareholder learn about his misconduct. Hence, his utility is 0. The first claim follows immediately.

Does this scheme incentivize high effort? The manager's expected utility as a function of effort is

$$U_t(e) = \mathbb{E}_t\left[\int_{\tau^*}^{\infty} \mathrm{e}^{-\beta(s-t)}\eta_D z\mathrm{e}^{z\tau^*}(1-c(e))\,ds\right] = \frac{\eta_D\lambda(e)(1-c(e))z\mathrm{e}^{zt}}{\beta(\beta+\lambda(e)-z)}. \tag{3.59}$$

The manager would only exert high effort iff $U_t(e^H) > U_t(e^L)$, which requires

$$c^H < -\frac{(\lambda^L-\lambda^H)(\beta-z)}{\lambda^H(\beta+\lambda^L-z)} < 0, \tag{3.60}$$

which is impossible. The second claim follows. *Q.E.D.*

In fact, if the manager could prematurely announce that τ^* has arrived, he would do so under this compensation scheme, since $\beta > z$. The manager prefers to start receiving lower pay now rather than higher pay later. Therefore, while the cash flow-based compensation provides strong incentives to reveal τ^*, it also strongly discourages any effort on behalf of the manager to extend the duration of high growth. This is a stronger result compared to Benmelech et al. (2010), where this compensation scheme may achieve the first best under some parametrizations.

3.4.6 Mixed compensation

Inspired by the potential of stock-based compensation to incentivise effort, and the power of net cash flow-based compensation to discourage cheating, Benmelech et al. (2010) suggest a mixed compensation of the form

$$w_t = \omega\eta P_t + (1-\omega)\eta_D D_t, \quad \omega \in [0,1], \tag{3.61}$$

that achieves the first best in their setting under a range of parameters. η_D here is normalized to η/r, such that the manager's expected utilities under purely stock-based compensation and purely cash flow-based compensation are the same.

Let us now investigate whether and under what conditions this compensation works in the present setting.

Proposition 3.10. *Under the mixed compensation scheme* (3.61)*, the high effort-reveal equilibrium is obtainable iff*

$$\frac{\lambda^L - \lambda^H}{\beta + \lambda^L - z} \mathrm{e}^{-(\beta - z)h^{**}} > c^H. \tag{3.62}$$

The weight on stock-based component should be set in the interval[22]

$$\omega \in \left(\frac{\Theta_{\lambda^H}}{1 - c^H \frac{\beta + \lambda^L - z}{\lambda^L - \lambda^H}}, \frac{\Theta_{\lambda^H}}{1 - \mathrm{e}^{-(\beta - z)h^{**}}} \right), \tag{3.63}$$

which has a nonempty intersection with $[0,1]$ *if* (3.62) *is satisfied.*

Proof. First, take it as given that the manager supplies high effort, and consider his conceal/reveal decision. Reusing the proof of the proposition 3.4, it is straightforward to show that $U^{\text{reveal}}_{\text{Mixed},\tau^*} = U^{\text{reveal}}_{\text{Stock},\tau^*}$ and $U^{\text{conceal, fi}}_{\text{Mixed},\tau^*} = \omega U^{\text{conceal, fi}}_{\text{Stock},\tau^*}$. The desired

[22] Closed-form bounds on ω are not possible to derive in Benmelech et al.'s setting.

reveal equilibrium obtains iff $U^{\text{conceal, fi}}_{\text{Mixed},\tau^*} < U^{\text{reveal}}_{\text{Mixed},\tau^*}$, which holds iff

$$\frac{\omega A^{\text{fi}}_{\lambda^H}}{\beta - z}\left(1 - \mathrm{e}^{-(\beta - z)h^{**}}\right) < \frac{z}{\beta r}, \tag{3.64}$$

which in turn is satisfied as long as

$$\omega < \frac{\Theta_{\lambda^H}}{1 - \mathrm{e}^{-(\beta - z)h^{**}}}. \tag{3.65}$$

This is the upper limit on ω in (3.63).

Now suppose that the manager reveals bad news in equilibrium, and focus on his effort provision decision. If investors believe that the rate parameter is λ, the manager's expected utility as a function of effort is[23]

$$\begin{aligned} U_t(e) &= \mathbb{E}_t\left[\int_t^{\tau^*} \mathrm{e}^{-\beta(s-t)}\omega\eta A^{\text{fi}}_{\lambda}\mathrm{e}^{zs}(1 - c(e))\, ds + \int_{\tau^*}^{\infty} \mathrm{e}^{-\beta(s-t)}\eta\frac{z\mathrm{e}^{z\tau^*}}{r}\, ds\right] \\ &= \frac{\eta \mathrm{e}^{zt}}{\beta + \lambda(e) - z}\left[\omega A^{\text{fi}}_{\lambda}(1 - c(e)) + \frac{z}{\beta r}\lambda(e)\right]. \end{aligned} \tag{3.66}$$

The manager thus supplies high effort iff

$$\frac{\beta + \lambda^L - z}{\beta + \lambda^H - z} > \frac{\omega A^{\text{fi}}_{\lambda} + \frac{z}{\beta r}\lambda^L}{\omega A^{\text{fi}}_{\lambda}(1 - c^H) + \frac{z}{\beta r}\lambda^H}. \tag{3.67}$$

The high effort equilibrium obtains iff (3.67) is satisfied for $A^{\text{fi}}_{\lambda^H}$. This yields the lower bound on ω:

$$\omega > \frac{\Theta_{\lambda^H}}{1 - c^H\frac{\beta + \lambda^L - z}{\lambda^L - \lambda^H}}. \tag{3.68}$$

The interval (3.63) must be non-empty for the first-best equilibrium to be obtain-

[23] Note that after τ^* the manager's compensation is $\omega\eta\frac{z\mathrm{e}^{z\tau^*}}{r} + (1 - \omega)\eta_D z\mathrm{e}^{z\tau^*} = \eta\frac{z\mathrm{e}^{z\tau^*}}{r}$, independent of ω.

able under mixed compensation. This is true as long as

$$\frac{\Theta_{\lambda^H}}{1-\mathrm{e}^{-(\beta-z)h^{**}}} > \frac{\Theta_{\lambda^H}}{1-c^H\frac{\beta+\lambda^L-z}{\lambda^L-\lambda^H}}, \tag{3.69}$$

which yields the condition (3.62). In addition, this interval must have a non-empty overlap with the interval $[0,1]$, in ω is required to lie. Since $0<\Theta_{\lambda^H}<1$, we need

$$\frac{\Theta_{\lambda^H}}{1-c^H\frac{\beta+\lambda^L-z}{\lambda^L-\lambda^H}} < 1. \tag{3.70}$$

After rearranging, this is the same as condition (3.21), which always holds by assumption. *Q.E.D.*

Observe that the condition (3.62) is the same as the condition (3.39) for high effort provision under the stock-based compensation in the conceal/mixed equilibria. Therefore, when the standard stock-based compensation provides incentives for the manager to work hard, but also induces him to conceal information, investors can fix the problem by adding a net cash flow-based component to the compensation package. The stock component continues to provide sufficient incentives for the manager to exert high effort, while the cash flow component effectively deters him from cheating and overinvesting.

3.5 Quantitative analysis

In this section I calibrate the model to explore the extent of the agency problems in the present setting, as well as quantitatively examine the performance of the stock-based and alternative compensation schemes.

Table 3.1 presents the calibration of the model parameters, which is based on the boom of the late 1990s. r is set at 10%, which is approximately the real cost of capital for the Telecommunications companies in the late 1990s according

Table 3.1: The alternative calibration: short period of rapid growth

Cost of capital	r	10%
Return on capital	z	30%
Manager's discount rate	β	35%
Shape parameter (high effort)	α^H	1.05
Implied expected firm's size	$\mathbb{E}[J\|e^H]$	21
Implied expected τ^*	$\mathbb{E}[\tau^*\|e^H] = 1/\lambda^H$	3.2 years
Shape parameter (low effort)	α^L	8.00
Implied expected firm's size	$\mathbb{E}[J\|e^L]$	1.14
Implied expected τ^*	$\mathbb{E}[\tau^*\|e^L] = 1/\lambda^L$	0.4 years
Manager's cost of high effort	c^H	50%
Duration of the concealment	h^{**}	3 years

to Aswath Damodaran's data.[24] z is set at 30%, which is roughly the average real return on NASDAQ composite index in the second half of 1990s.[25] The manager's discount rate is set at 35%. This is the median value of the average estimates of individual discount rates across 43 studies reviewed by Frederick et al. (2002).[26] While the number appears high, Frederick et al. point out that many of the estimates confound time preference with risk aversion. This calibration here is appropriate, given that the manager faces a lot of undiversified risk from the company's performance. I set $\alpha^H = 1.05$, such that the expected duration of the growth phase under high effort is realistic 3.2 years. This implies that the firm is expected to grow in size 21 times over that period. α^L is set equal to 8, such that the expected duration of the growth phase under low effort is 0.4 years, exactly $2/15^{th}$ of 3.2 years, to mimic the calibration in Benmelech et al. (2010). It is difficult to calibrate the cost of effort, so I set it at $c^H = 50\%$. The results remain similar, however, for a wide range of cost parameters. Lastly, I set $h^{**} = 3$ years.

I find that under this calibration, the standard stock-based compensation (3.20) induces high effort-conceal behaviour in equilibrium, since both conditions (3.29) and (3.39) are satisfied. In fact, investors strongly prefer this outcome

[24]Damodaran Online is available here: http://people.stern.nyu.edu/adamodar/
[25]Encon Corporation and the telecommunications sector had similar returns over that period.
[26]I owe this reference to Michaillat and Saez (2018).

to the low effort-reveal one, due to high return on capital and relatively short concealment period during which they receive no dividends. Shareholders thus would willingly compensate the manager with stocks, despite the inevitable misbehaviour and overinvestment, if they are unable to achieve the desired high effort-reveal outcome.

We know from the Proposition 3.7, however, that stock-based compensation with appropriately delayed payments can successfully deter the manager from concealing information, while not compromising his incentives to work hard. I find that the minimum required delay $\underline{k}$ should be 1.9 years. This is sensible, given that vesting periods of several years are common in practice. Delaying payments from stock sales thus appears to be a viable way of resolving agency problems in high-growth startups.[27]

In addition, a compensation scheme that mixes stock awards with net cash flow-based bonuses should also work, according to the Proposition 3.10, as long as the weights are set appropriately. I find that the weight on the stock-based component must lie within the range of 13% to 37%, which in fact includes the 30% number suggested by Benmelech et al. (2010) in their setting.

Given the relative simplicity of the above compensation schemes, the model suggests that investors should be able to successfully resolve the agency problems in startup firms. Episodes like the Dot-com bubble in the 1990s, which was accompanied by many instances of managerial misbehaviour and even outright fraud, may thus appear puzzling. One possibility is that the shareholders simply did not understand the drawbacks of stock-based compensation contracts when they surged in popularity during the 1990s. Parameter uncertainty, especially in relation to 'new economy' firms, is another possibility: the shareholders might lack enough knowledge about the innovative businesses to design effective compensa-

[27] This contrasts with Benmelech et al. (2010), who find that in the context of mature firms, delaying payments only works when the cost of effort is unrealistically low, and the delay is at least 11.8 years.

tion schemes. Since, as we have seen, they prefer high effort to truth telling, they might willingly err in the direction of overpowering the managerial incentives to work hard. Lastly, market irrationality and overoptimistic expectations may well also play their role during such episodes.

3.6 Conclusion

In this paper I study the design and economic consequences of managerial compensation contracts in financially constrained high-growth startups. I show that stock-based compensation can incentivize the manager to work hard to pursue further growth, but also to conceal from investors bad news about the firm's prospects in order to maintain high stock price growth. This leads to substantial overinvestment and bubble-like boom and bust in the stock price in equilibrium. This paper thus extends results of Benmelech et al. (2010) to high-growth, no-dividend firms, and sheds light on the tech boom of the 1990s. I also explore alternative compensation schemes popular in practice. Notably, I find that delaying payments from stock sales can discourage the manager from hiding bad news, while not compromising his incentives to work hard.

One potential shortcoming of the model is that a lot of the presented results rely on the manager having a high personal discount rate. While this assumption is justified in the present setting, the behaviour of the model changes significantly if the manager's discount rate falls below the rate of return on the firm's capital. This is because under the stock-based compensation, a patient manager would not continuously sell his shares to the market. Instead, he has incentive to accumulate stocks, and time selling off all his holdings at the peak market price. There is indeed some evidence that the Dot-com bubble was burst by informed insiders selling their shares (Griffin et al., 2011). It will be worthwhile to explore in the future research how different compensation schemes perform when the manager

is more patient.

Another limitation of this study is its focus on a single firm. The historical episodes that are accompanied by bubble-like behaviour in stock prices and apparent managerial misbehaviour and overinvestment in capital often involve entire industries: the Railway Mania, the Internet Bubble, and the Telecommunications Boom, for example (Chancellor, 2000). The model presented in this paper is not readily applicable to the case when there is more than one firm, unless one is willing to assume that the firms are very similar and run out of growth opportunities at the same time. Otherwise, the model is unable to explain why the entire industries experience simultaneous stock market crashes, rather than individual firms. A promising avenue for future research is, therefore, extending the model to include many startup firms with different, but correlated growth prospects, and aggregate uncertainty about the overall potential of the industry. Indeed, aggregate uncertainty likely plays a large role in driving the investors' expectations and stock prices in innovative sectors. At the same time, the presence of other firms that can reveal information about industry productivity to investors may substantially affect individual managers' incentives to conceal news.

Bibliography

Adelphia Communications Corporation (2001). 2000 10-K Form. Retrived from https://www.sec.gov/Archives/edgar/data/796486/000079648601000012/0000796486-01-000012.txt.

Atmaz, A. and Basak, S. (2019). Asset prices and no-dividend stock. SSRN Working Paper.

Axtell, R. L. (2001). Zipf Distribution of U.S. Firm Sizes. *Science*, 293(5536):1818–1820.

Benmelech, E., Kandel, E., and Veronesi, P. (2010). Stock-Based Compensation and CEO (Dis)Incentives. *The Quarterly Journal of Economics*, 125(4):1769–1820.

Beresford, D. R., deB. Katzenbach, N., and Rogers, Jr, C. (2003). Report of Investigation by The Special Investigative Committee of the Board of Directors of Worldcom, Inc. SEC.

Brenner, R. (2003). Towards the precipice. *London Review of Books*, 25(3).

Carpenter, R. E. and Petersen, B. C. (2002a). Capital Market Imperfections, High-Tech Investment, and New Equity Financing. *Economic Journal*, 112(477):54–72.

Carpenter, R. E. and Petersen, B. C. (2002b). Is The Growth Of Small Firms Constrained By Internal Finance? *The Review of Economics and Statistics*, 84(2):298–309.

Chancellor, E. (2000). *Devil Take the Hindmost: A History of Financial Speculation*. Plume.

Coombs, B. (2015). The surprising trend of tech stocks and dividends. Retrieved from https://www.cnbc.com/2015/03/08/the-surprising-trend-of-tech-stocks-and-dividends.html.

Cornell, B. and Shapiro, A. C. (1988). Financing corporate growth. *Journal of Applied Corporate Finance*, 1(2):6–22.

Dreazen, Y. J. (2002). Wildly optimistic data drove telecoms to build fiber glut. *The Wall Street Journal*. Retrieved from http://www.wsj.com/articles/SB1032982764442483713.

Edmans, A. and Gabaix, X. (2016). Executive compensation: A modern primer. *Journal of Economic Literature*, 54(4):1232–87.

Edmans, A., Gabaix, X., and Jenter, D. (2017). Chapter 7 - Executive Compensation: A Survey of Theory and Evidence. volume 1 of *The Handbook of the Economics of Corporate Governance*, pages 383–539. North-Holland.

Fabrikant, G. and Romero, S. (2002). How Executives Prospered as Global Crossing Collapsed. *The New York Times*. Retrieved from https://www.nytimes.com/2002/02/11/business/how-executives-prospered-as-global-crossing-collapsed.html.

Fama, E. F. and French, K. R. (2001). Disappearing Dividends: Changing Firm Characteristics Or Lower Propensity To Pay? *Journal of Applied Corporate Finance*, 14(1):6779.

Farre-Mensa, J., Michaely, R., and Schmalz, M. (2014). Payout Policy. *Annual Review of Financial Economics*, 6(1):75–134.

Frederick, S., Loewenstein, G., and O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40(2):351–401.

Gabaix, X. (2009). Power Laws in Economics and Finance. *Annual Review of Economics*, 1(1):255–294.

Global Crossing Ltd. (2001). 2000 10-K Form. Retrived from https://www.sec.gov/Archives/edgar/data/1061322/000095013001001603/0000950130-01-001603.txt.

Goldman, E. and Slezak, S. L. (2006). An equilibrium model of incentive contracts in the presence of information manipulation. *Journal of Financial Economics*, 80(3):603–626.

Griffin, J. M., Harris, J. H., Shu, T., and Topaloglu, S. (2011). Who drove and burst the tech bubble? *The Journal of Finance*, 66(4):1251–1290.

Hall, B. J. and Liebman, J. B. (1998). Are CEOs Really Paid Like Bureaucrats? *The Quarterly Journal of Economics*, 113(3):653–691.

Kedia, S. and Philippon, T. (2007). The Economics of Fraudulent Accounting. *The Review of Financial Studies*, 22(6):2169–2199.

Michaillat, P. and Saez, E. (2018). A New Keynesian Model with Wealth in the Utility Function. Working Paper.

Peng, L. and Roell, A. (2008). Manipulation and equity-based compensation. *American Economic Review*, 98(2):285–90.

Peng, L. and Roell, A. (2014). Managerial incentives and stock price manipulation. *The Journal of Finance*, 69(2):487–526.

Stein, J. C. (1988). Takeover threats and managerial myopia. *Journal of Political Economy*, 96(1):61–80.

Stein, J. C. (1989). Efficient Capital Markets, Inefficient Firms: A Model of Myopic Corporate Behavior*. *The Quarterly Journal of Economics*, 104(4):655–669.

WorldCom Inc. (2001). 2000 10-K Form. Retrived from https://www.sec.gov/Archives/edgar/data/723527/000091205701510527/0000912057-01-510527.txt.