THREE ESSAYS IN FINANCIAL ECONOMICS

Martin Christoph Schmalz

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF ECONOMICS

Adviser: Markus K. Brunnermeier

JUNE 2012
Abstract

This dissertation is comprised of three chapters that cover topics in corporate finance, behavioral finance, and financial economics. In “Managing Human Capital Risk,” I argue that adjustment costs and firm-specific human capital make labor behave like an asset of the firm, rather than like an externally supplied service “flow,” as extant work suggests. I provide empirical support by showing that firms manage human capital risk similar to the way they manage risk from physical assets. In response to an exogenous increase of labor adjustment costs, firms reduce leverage and increase the liquidity of their on-balance sheet asset structure.

“Anxiety and Overconfidence in the Face of Risk,” joint with Thomas Eisenbach, is the first attempt in the literature to study dynamically inconsistent preferences with respect to risk tradeoffs. We show that higher risk-aversion for more imminent risks can explain a set of well-documented asset pricing patterns, such as earnings announcement anomalies, as well as systematic belief distortions, such as the underestimation of risks known as overconfidence.

“Revealing Downturns,” joint with Sergey Zhuk, presents a model of Bayesian learning about multiple parameters of firms’ fundamentals. We show that the market can better tell apart components of operating performance due to idiosyncratic “goodness” or skill from correlation with market-wide factors when the macroeconomy declines. Therefore, stock picking earns higher returns and boards tend to fire CEOs more frequently in downturns rather than in upturns.
Acknowledgments

Many people contributed to this dissertation. I am deeply grateful for their advice, support, encouragement, and friendship.

My adviser, Markus Brunnermeier, guided my steps in academic and non-academic matters with infinite patience and whenever needed. As the dissertation progressed, David Sraer assumed a more and more engaged role. He most generously shared ideas, his time, and advice. Without David Lee and Alexandre Mas, who shared not only advice, but also their data, my job market paper would not exist. Hyun Song Shin’s comments and questions helped me think further and substantially improve that paper as well.

I owe thanks to many more faculty members at Princeton’s Department of Economics, the Bendheim Center for Finance, and elsewhere, as well as my classmates at Princeton. For their helpful comments, enlightening discussions, time, encouragement, and patience, I want to thank Roland Bénabou, Smita Brunnermeier, Sylvain Chaissang, Harry DeAngelo, Thomas Eisenbach, Harrison Hong, Oleg Itskhoki, Jakub Jurek, Konstantin Milbradt, Benjamin Moll, Stephen Morris, Ulrich Müller, Martin Oehmke, Wolfgang Pesendorfer, José Scheinkman, and Wei Xiong. Others are separately acknowledged at the beginning of the chapters.

For their encouragement and support, I am grateful to my parents, family, and friends. Among the latter, Francesco Cellarosi, Ryan Peckner, and Judith Sieber stand out.
Contents

Abstract ................................................................. iii
Acknowledgments ....................................................... iv

1 Managing Human Capital Risk .................................. 1
   1.1 Introduction .................................................. 3
   1.2 Model .......................................................... 11
   1.3 Institutional Background, Research Design, and Data Set Assembly 26
   1.4 Empirical Results ........................................... 32
   1.5 Conclusion .................................................... 40

2 Anxiety and Overconfidence .................................... 59
   2.1 Introduction .................................................. 61
   2.2 Model .......................................................... 69
   2.3 Finance Applications ........................................ 71
   2.4 Overconfidence .............................................. 83
   2.5 Conclusion and Plans for Future Research ................. 89

3 Revealing Downturns ............................................. 95
   3.1 Introduction .................................................. 97
   3.2 Model .......................................................... 102
Chapter 1

Managing Human Capital Risk
CHAPTER 1. MANAGING HUMAN CAPITAL RISK

Abstract

This paper argues that corporate finance should treat labor as an asset of the firm, rather than as an externally supplied service “flow.” Labor adjustment costs make it optimal to retain hard-to-replace employees in bad times, and thus cause an “implicit liability” to pay their wages. Firms optimally hold equity-financed cash to insure against the risk of being unable to follow that optimal labor policy.

I distinguish my model from existing models of the interaction between corporate finance and labor by identifying the corporate finance response to unionization with a regression discontinuity design. Increased labor adjustment costs due to unionization cause higher cash-to-asset ratios and lower net leverage in financially unconstrained firms. Firms that cannot raise cash “save on risk management.” They decrease cash-to-assets and increase net leverage.\(^1\)

\(^1\)I am deeply grateful to my adviser Markus Brunnermeier, as well as to David Sraer and Alexandre Mas for invaluable support, ideas, and advice, and to David Lee and Alexandre Mas for generously providing data. For helpful discussions and comments, I would like to thank José Azar, Thomas Eisenbach, Henry Farber, Edoardo Grillo, Harrison Hong, Bo Honoré, Oleg Itskhoki, Jakub Jurek, Martin Kanz, David Lee, Thomas Mertens, Benjamin Moll, Ulrich Müller, Juan Ortner, Filipppos Papakonstantinou, José Scheinkman, Hyun Song Shin, and Sergey Zhuk as well as seminar participants at Bocconi University, Boston University SoM, Columbia GSB, Harvard Business School, ICEF Moscow, IESE, INSEAD, London School of Economics, Michigan Ross, Norges Bank Investment Management, Ohio State Fisher, Princeton, Rochester Simon, SMU Cox, UC Irvine, University of Vienna and VGSF, USC Marshall, and USI Lugano. Laurien Gilbert and Alexis Furuichi provided excellent research assistance. I gratefully acknowledge support through a Fellowship of Woodrow Wilson Scholars.
1.1 Introduction

Managers often think of employees as assets of the firm, and thereby look beyond what is visible on the balance sheet. For example, the software company SAP states that “The assets that are the basis for our [...] success do not appear on the Consolidated Statements of Financial Position. [...] They include [...] employees and their knowledge and skills.”\(^2\) Given that employees cannot be owned, what do firms really mean when they refer to employees as assets of the firm? How is that asset financed? How does the firm manage the ensuing risk?

These questions have been difficult to answer, as neither the asset value of the firm-worker relationship nor all respective liabilities are systematically recorded on the balance sheet. As a consequence, arrived corporate finance models often do not recognize employees as either an asset or a risk factor. In practice, they seem to be both. Specifically, researchers often model employees as external suppliers of the firm – a stark contrast also to newer theories of the firm (e.g., Rajan and Zingales (1998)). These distinctions are not merely semantic. Even the principal direction of optimal corporate financial policy can depend on whether labor is an externally supplied “flow” or an internal asset, or “stock,” of the firm. Risk management is one example. If labor has the characteristics of a valuable, off-balance sheet asset that requires continuous cash flows to be maintained, but is viewed by managers as a flow that can be turned on and off at will, the firm will keep too little cash and use too much debt. One adverse consequence of such a flawed policy is the loss of firm-specific human capital through inefficient firing of key employees in bad times.

CHAPTER 1. MANAGING HUMAN CAPITAL RISK

Such constraint-driven firing protects short-term accounting profits, but destroys shareholder value.\(^3\)

This paper argues and provides empirical support that labor is best treated as a stock that is a part of the firm to the extent to which it is unique to the firm or otherwise hard to replace. First, I define what is meant by “the human capital of the firm.” Then, I provide (i) a simple theory that generates basic answers to the motivating questions above, (ii) a regression discontinuity design to identify the firm’s risk management response to an increase in labor adjustment costs (and thus human capital risk), and (iii) an intuitive interpretation in terms of implicit assets and liabilities, thus bridging some of the gap between empirical corporate finance and implicit contract-based theories of the firm.

It is the firm’s relationship to its employees that gives the firm de facto control rights over it’s employees’ human capital. This relationship represents the productive asset that is commonly referred to as the “human capital of the firm,” and that the above questions apply to. The liability side of this relationship is the implicit promise of the firm to retain valuable employees in bad times and to thus assume a portion of the risk from fluctuations in the value of its employees’ human capital. Thus, the firm is exposed to human capital risk, although it does not legally own any human capital. The human capital risk the firm bears increases with how costly it is to replace the employee. The cost of replacing an employee can be due to the loss of firm-specific human capital, or due to explicit firing or hiring costs. In standard parlance, the human capital that “belongs” to the firm can be

\(^3\)Similar in spirit to Shleifer and Summers (1988), it may thus in the firm’s best long-term interest to engage in employee-friendly policies. Passov (2003) gives a practitioner’s perspective that expands the case to other intangible assets and associated “intangible liabilities”. Rajan (2012) re-iterates that the same frictions that cause the firm to exist also impacts optimal financial policy, of which labor adjustment costs are one.
regarded as a firm-specific asset whose value fluctuates over time, has low market liquidity, and no funding liquidity. It is financed with a promise of a quasi-fixed stream of coupon payments. The ensuing human capital risk is managed not unlike the way that risk stemming from debt-financed physical assets is managed (e.g., Bolton, Chen, and Wang (2011)): with equity-financed cash. Thus, for first-order risk management predictions, the firm’s human capital can be thought of as an off-balance sheet lease with adjustment costs.

A need to manage human capital risk arises because the firm needs liquid funds to pay wages in the “bad” states of nature, but may be unable to raise cash precisely then. Specifically, I assume that loss-making firms have no access to outside capital because, among other reasons, their productive asset – the firm’s relationship to its employees – is not verifiable, let alone collateralizable. As a result, shareholder value maximizing managers respond with higher equity-financed cash to an increase in labor adjustment costs, as long as they have access to external capital markets or can simply retain earnings.\(^4\) In contrast, financially constrained firms cannot raise additional cash by definition. They prefer to “invest” their available funds in the human capital in the firm in form of wage payments, rather than put it aside to ensure future wage payments. As a mechanical result, they passively eat through their cash cushion in bad times. Some of them will exhaust their liquid funds, be unable to pay wages, and thus effectively fire their employees. This, by assumption, is more costly if adjustment costs are higher. Thus, in contrast to unconstrained firms, the average constrained firm reduces cash-to-asset ratios and increases net

\(^4\)Financing any portion of that cash with debt would lead to inefficiently early firing of employees in a recession. Indeed, firms increasingly issue shares for cash savings, as David McLean (2011) documents. There may be other, similarly suitable, tools that firms use to manage that risk. For the purposes of this paper, I concentrate on cash and negative debt as they are easily observable.
leverage when adjustment costs increase. In sum, I predict higher cash-to-asset ratios and lower net leverage for firms with low financial constraint risk when it becomes more costly to replace workers, and the opposite for firms that are likely to be financially constrained.

The obvious difficulty for an empirical evaluation of my theory is finding exogenous variation of an unmeasured variable – the liability implied by labor adjustment costs, and the thus ensuing “human capital risk.” The solution I offer is to identify the corporate finance response to unionization, an actively discussed but still unresolved question in itself. Unionization changes the firm-worker relationship by increasing labor adjustment costs, therefore the firm’s incentives to retain employees in bad times, and thus the firm’s exposure to human capital risk. Indeed, as recently in the case of Goldman Sachs, unions may form in response to a decline in workplace security (Tabuchi (2012)). This underscores that unionization takes a similar role as other labor characteristics that make employees costly to replace. The causal effect of unionization on corporate financial policy should therefore be an increase in cash-to-asset ratios and a decrease in net leverage.

I use the data and regression discontinuity design from Lee and Mas (2011) (hereafter, LM), and estimate that unionization causes 5% higher cash-to-asset ratios and 1.8% lower net leverage (compared to a hypothetical, not unionized, but otherwise identical firm) in firms that are likely to be able to freely choose their

---

5Constrained firms therefore appear to “save on risk management” (see also Rampini, Sufi, and Viswanathan (2011)).

6At the same time, empirically, there is no causal effect of unionization on wages. Thus, the costs of labor are only affected by unionization in the rare case that the firm is forced to fire workers because a liquidity constraint binds. I discuss and rule out other effects of unionization in the following sections.
financial policy. As expected, the response for “constrained” firms is more noisy, but points to lower cash-to-asset ratios and higher net leverage.\(^7\)

It is an established insight in the economics literature that labor adjustment costs can make it optimal for the firm to retain its workforce, even when wages temporarily exceed the marginal product of labor. It does so to avoid current firing costs, future hiring costs, and to retain firm-specific human capital (Oi (1962), Fay and Medoff (1985), and Dixit (1997)). The insights that the capital structure of firms includes obligations to non-financial stakeholders (Fama (1990)), and that firms with large implicit or explicit off-balance sheet obligations should operate with conservative financial policies (e.g., Cornell and Shapiro (1987)), are not new either. This paper’s contribution is to empirically investigate the corporate finance consequences of the optimal labor retention policy in the presence of financial constraints. It thereby provides quantitative support that implicit contracts with non-financial stakeholders matter for financial policy in predictable ways. The paper shares the intuition with Berk, Stanton, and Zechner (2010) and Agrawal and Matsa (2011) that labor-related costs of financial distress make more conservative financial policies optimal. Their papers focus on debt and the motivation to avoid bankruptcy, a default on explicit liabilities. My paper focuses on cash management.

\(^7\)The results are robust to the choice of regression model, conditioning of the data, the ex ante measure of financial constraints I use, and other variations. LM and DiNardo and Lee (2004) (hereafter, DL) have used the same discontinuity to investigate the impact of unionization on wages and other economic variables (DL) and on stock market performance (LM). Consistent with ideas as proposed in Acharya, Baghai, and Subramanian (2010), DL and LM find little causal impact on most real variables, including wages, productivity, and market equity. In an online appendix, LM also use quarterly data for an event study of a selection of Compustat variables. They find little evidence of a causal response for capital structure choice. In contrast to their approach, I use annual data (which are less noisy) and different variables, and I focus on the RD implementation rather than the time-series analysis. Finally, only the distinction between financially constrained and unconstrained firms reveals that unionization has a causal impact on both risk management and capital structure.
and the motivation to avoid default on implicit liabilities, which does not lead to bankruptcy, but to a decline in shareholder value. Financial distress – adverse economic consequences of an insufficient liquidity position – can arise even when the firm has a net cash position. The existing models, by construction, cannot address financial distress in zero-leverage firms. Also, the framework in Berk, Stanton, and Zechner (2010) makes opposite, if any, predictions, how firms would respond to unionization.

Four aspects distinguish this work from existing papers on the corporate finance response to unionization. First, existing theories focus on capital structure choice rather than risk management; they do not jointly predict and measure the leverage and cash response to unionization. Second, existing empirical approaches do not make the distinction between unconstrained and constrained firms, which is crucial in my paper.

Third, while not challenging the results of the “strategic security design” literature, my paper may convey a distinct intuition, at least for a subset of firms. Specifically, unionized firms have been associated with lower cash (Klasa, Maxwell, and Ortiz-Molina (2009)) and higher leverage (Bronars and Deere (1991), Matsa (2010)) before. The same correlations obtain from my dataset: driven by financially constrained firms, the average unionized firm has lower cash-to-assets and higher leverage also in my sample. However, I show that the causal effect of unionization goes in the opposite direction for both variables.\(^8\) That does not imply that strategic motives such as securing a better bargaining position by way of an aggressive financial strategy are unimportant. In fact, they seem to be the dominating force

\(^8\)The previous papers’ identification strategies allow for the alternative interpretation of their estimates as the response to a union threat, rather than to actual unionization. Another difference is that I focus on book leverage instead of market leverage to avoid confounding a market response with a conscious choice of the firm.
for firms that get unionized by a wide margin. Only the effect of managing human capital risk seems to dominate at the marginal firm. The present paper’s objective is to identifying the causal effect of an increase of labor adjustment costs on the firm’s financial strategy, and not to study the interaction between corporate finance and unionization more generally. The causal effects are therefore the critical result of my paper. Documenting the non-causal correlations merely serves to rule out that the contrast to the existing literature stems from the difference of data sets.

Fourth, my theoretical and empirical results are conceptually different from the idea of using operating leverage as a measure of human capital risk – an interpretation that the existing literature would allow. Similar in spirit to my results, Simintzi, Vig, and Volpin (2010) estimate a negative effect of labor bargaining power on financial leverage. The authors attribute the effect to higher operating leverage “in the form of higher labor cost,” which crowds out financial leverage. As they investigate the effect of labor protection on a more general level than unionization, our respective results are not directly comparable. Yet, it is noteworthy that I neither predict nor find a causal effect of unionization on either the wage bill or on operating leverage – in particular not for the firms that respond to unionization with higher cash-to-assets and lower net leverage. I therefore propose that unionization increases firm risk by way of increasing the fixity of the quasi-fixed labor cost share, rather than instantaneously increasing the labor-related fixed cost, which would be captured by traditional measures of operating leverage. In sum, this paper complements, rather than challenges, the existing literature on the corporate finance response to unionization.9

---

9This paper builds on but does not contribute to the existing results on the employment consequences of financial constraints, such as in Caggese and Cuñat (2008) or Benmelech, Bergman, and Seru (2011), who expand the intuition of Hubbard (1997), Stein (2003), Almeida, Campello, and Weisbach (2004) to the labor dimension, or Bertrand, Schoar, and Thesmar (2007). Also, I do not discuss asset
The “Hidden Balance Sheet” helps to illustrate the distinction between traditional concepts of operating leverage and the risk factor I have in mind. It also provides an informal link between implicit contract-based theories of the firm and empirical corporate finance. The value of the firm-worker relationship can be regarded as an implicit off-balance sheet asset, and the implicit obligation to insure employees’ human capital risk as an implicit off-balance sheet liability. The value of that liability can increase with the duration of the obligation even when the per-period quasi-fixed payments – and thus operating leverage – remain constant. This is, I argue, the impact of an increase of labor adjustment costs on the firm, proxied for by unionization. Managing human capital risk is about preventing default on these implicit liabilities.

While very basic and theoretically expandable, The Hidden Balance Sheet is pragmatically useful as it is: it predicts the qualitative empirical results. Establishing an employment relation (as opposed to contracting with externally supplied labor on the spot market) leverages the physical assets and the economic balance sheet of the firm. Higher adjustment costs, e.g., by unionization, further increase the implicit liabilities. As the firm holds precautionary cash not only against the liabilities visible on the balance sheet, but against its total liabilities, an increase in implicit liabilities triggers an on-balance sheet response in the form of higher cash holdings and lower net leverage. ¹⁰

The economic rationale behind the answers to the motivating questions is also illustrated intuitively. The liability corresponding to the asset value of the firm-pricing implications of the firm-worker relation, as in Eisfeldt and Papanikolaou (2009) or Chen, Kacperczyk, and Ortiz-Molina (2010). For a more general overview of the interaction of labor and finance, see Pagano (2010).

¹⁰The “Hidden Balance Sheet” also illustrates on-balance sheet variations of financial policy in response to variations of explicit off-balance sheet assets and liabilities, such as in Shivdasani and Stefanescu (2010).
worker relationship is the insurance claim that hard-to-replace employees hold. That promise is only credible, and the insurance claim only valuable, if it is collateralized with liquid funds. An increase in the implicit liability is therefore met with an increase in liquid funds. This view opens up the possibility of a strategic use of corporate financial policy that contrasts the previously discussed propositions. If firms compete for workers, a firm with plenty of liquid funds and low debt may be able to attract valuable employees from an otherwise identical firm that employs a more aggressive financial policy. It is a comparatively attractive employer not because it offers a higher wage, but because it can credibly signal that it will insure the worker commensurate with the economic value of his specific skills. According to this view, a prudent financial policy is a necessary condition to induce workers to make firm-specific investments in their human capital, see Zingales (2000).

In the next section, I present a continuous-time model of human capital risk management, discuss its limitations and possible extensions. Section 1.3 describes the data. Section 1.4 gives the empirical results and robustness checks, and discusses limitations of the empirical approach. Section 3.4 briefly summarizes the insights of the paper and suggests a direction for future research.

1.2 Model

This section outlines a simple dynamic model of risk management in continuous time. The conceptual insight is that labor adjustment costs make labor behave like an asset of the firm that is valuable but risky, illiquid, and requires periodic payments to be maintained. It has similar corporate finance implications as an off-balance sheet lease with adjustment costs: if there are financing frictions in states of the world in which internal cash flows are low, firms optimally manage human
capital risk similar to the way they optimally manage the risk associated with physical assets that have limited technological liquidity; very low, if any, market liquidity; and no funding liquidity: with equity-financed cash.\textsuperscript{11} This view of human capital in the firm is sufficient to generate basic empirical predictions about human capital risk management. It is too narrow in different contexts, such as attracting and investing in human capital, as opposed to merely retaining it. Extensions that capture some of these differences between human and physical capital are discussed at the end of this section.

**Setup**

There are three necessary ingredients to the model: (i) non-negative labor adjustment costs, (ii) the wage has a fixed component, (iii) a financing friction for off-balance sheet implicit assets in states in which internal cash flows are low. For the sake of clarity and simplicity of the model, I use stronger assumptions: labor adjustment costs are positive, wages are perfectly fixed, and outside capital markets are closed when operating cash flows are negative. The qualitative predictions do not depend on these stronger assumptions.

For the sake of simplicity, and following the labor search literature (e.g., Mortensen and Pissarides (1994)), I assume that the only input to production is one worker. More general production functions can be easily accommodated. A firm’s revenues are determined by the productivity of its worker’s human capital, $\nu_t$. $\nu_t$ follows a Markov process such that $\nu_t \in \{\nu_L, \nu_H\}$, $\nu_L < w < \nu_H$, where $w$ is the wage level,\textsuperscript{11}The reason for a preference for equity, and thus a wedge between negative debt and cash, is that the firm’s debt capacity is ill-defined in my model: the availability of debt financing fluctuates over time. Empirically, debt covenants often include requirements for positive operating cash flows and profits, making a comparatively liquid asset structure the optimal tool to survive prolonged downturns of the business; see, e.g., DeAngelo, DeAngelo, and Wruck (2002).
and with switching probabilities $\lambda_{H,L}dt$ from $\nu_H$ to $\nu_L$ and $\lambda_{L,H}dt$ from $\nu_L$ to $\nu_H$.

Financing is costless, but carrying cash in the firm is costly, as interest rates are positive, $r > 0$.

Cash flow at time $t$, $dc_t$, is composed of cash flow from operations, $1(\nu_t - w)$ (where $1 = 1$ indicates that there is an employee in the firm), plus cash flows from financing activities, $dc_{fin,t}$, minus dividend payments, $dc_{div,t}$, minus labor adjustment costs, $dc_{adj,t}$, i.e., cash paid out for firing, hiring, or training employees.

$$dc_t = 1(\nu_t - w)dt + dc_{fin,t} - dc_{div,t} - dc_{adj,t}$$

The firm’s problem is to solve for the shareholder value maximizing (i) employment policy $1 \in \{0, 1\}$, (ii) financing policy $dc_{fin,t}$, and (iii) dividend policy $dc_{div,t}$. Labor adjustment costs $dc_{adj,t}$ are implied by the employment policy $1$. The non-trivial part of the problem amounts to determining the optimal precautionary cash level $\bar{c}$ and how it is financed, subject to the following three assumptions.

The first assumption is that there are costs of replacing employees. These costs may be direct or indirect labor adjustment costs, or consist in the loss of firm-specific human capital that is costly and time-consuming to rebuild. Formally, the firm can fire the worker at any time, but at a cost $f$.\(^{12}\) In addition, or alternatively, there may be costs associated with hiring new workers in the future, $\gamma$. This cost captures more than search expenditures and administrative processes associated with hiring a new worker. It can also include costs for training that generates firm-specific human capital that is necessary for production. In the real world, the firm

\(^{12}\) One can think of severance pay $f$ as an enforceable commitment the firm makes at the beginning of an employment relationship. Alternatively, it is straightforward to endogenize the optimality for the firm to honor the commitment even if it is implicit, e.g., if it can secure profitable future employment relationships only by honoring implicit commitments with the current worker.
may “pay” that cost partly by investing experienced workers’ time to introduce a new hire to business processes, colleagues, and her particular task. Such investments temporarily reduce experienced workers’ output, and thus profits. Thus, \( \gamma \) cannot only be thought of as a hiring cost, but also as a rough measure of specific human capital.\(^{13}\)

**Assumption 1:** Firing costs are \( f \geq 0 \), hiring costs or specific human capital is \( \gamma \geq 0 \), and \( f + \gamma > 0 \).

Assumption 1 makes labor behave like a stock. For that stock to contribute a financial risk to the firm, some part of it needs to be financed with quasi-fixed payments. If there is any fixed component in the employees’ wage, this condition is met, and establishing an employment relation automatically transfers risk from the employee on the firm. For analytical convenience, I assume the wage to be perfectly fixed.

**Assumption 2:** Wages are fixed, at level \( w \).

Thirdly, an impediment to external financing is that there are no verifiable, let alone collateralizable, assets in the firm. The only hard information investors have to gauge the value of the firm are cash flows, which are negative. Empirically, that makes external financing prohibitively costly or impossible. Even if that asymmetric information problem could be resolved, firms will often try to prevent doing so,

\(^{13}\)\( \gamma \) is the share of the total costs of training that the firm bears. Including the pecuniary and non-pecuniary costs to the employee is not necessary for a first-order analysis of the firm’s management of human capital risk. A more important omission is on-the-job learning, i.e., a mechanical increase of worker productivity with tenure. Including that aspect would not materially alter the model predictions.
CHAPTER 1. MANAGING HUMAN CAPITAL RISK

for two reasons. First, revealing information about their human capital investments (e.g. research and development efforts) may lead to a competitive disadvantage. Second, key employees may leave the firm as a result of a change of ownership, destroying shareholder value, and exacerbating the asymmetric information problem. A last impediment may be self-imposed constraints to avoid external financing by owners such as family firms and majority stakeholders, e.g., to prevent a loss of control. As a result, the firm may be unable to access capital markets when cash flows are negative, i.e., when it need outside cash the most.\footnote{Phone interviews with a variety of companies confirmed the validity of this assumption. In particular, pioneering software companies such as SAP had little access to bond markets until an established and well-understood business model had generated stable cash flows over many years. Passov (2003) is another strong account from a practitioner’s view that funding for investment in intangibles dries up when product market conditions are unfavorable. Endogenizing these legal, agency and informational frictions is not the focus of this paper.}

**Assumption 3:** The firm cannot access capital markets when $\nu_t = \nu_L$, i.e., $dc_{fin,t} = 0$ when $\nu_t - w < 0$.

As mentioned above, in the strong form presented here, the assumption is for analytical convenience only. Restricted access to equity markets is sufficient to generate the qualitative predictions.

In sum, Assumption 1 creates an incentive to retain employees. Assumption 2 says that retaining employees necessitates periodic wage payments, even if they are not productive. Labor adjustment costs therefore impose an implicit liability to make periodic wage payments on the firm. If such implicit liabilities are not honored, the firm does not go into bankruptcy, as would be true in case of default on explicit liabilities, but its shareholder value erodes. Assumption 3 gives the motive to manage the risk created by assumptions 1 and 2 with cash.
Figure 1.2 illustrates the model dynamics. When the regime is “low,” the firm makes losses and cannot refinance, and thus has a negative cash flow. The firm is forced to fire workers if the regime does not switch back to “high” before cash hits the lower bound $c_t = f$. It can then wait for a regime switch back to $\nu_t = \nu_H$ and restart the firm. Each time the state switches from “low” to “high,” the firm optimally raises cash to $\bar{c}$ immediately: recapitalization is costless but valuable, as it avoids the risk of entering a recession with less than the optimal cash buffer. The firm continues to pay out dividends in every subsequent “high” realization, until the regime switches to “low” again. The tradeoff that determines the optimal cash level $\bar{c}$ is between the costs of cash in the firm in terms of foregone interest, and the risk of hitting a liquidity constraint that forces costly liquidation of human capital.

Model Predictions

I will show that a shareholder value maximizing firm (i) does not fire workers unless the liquidity constraint binds, (ii) optimal precautionary cash $\bar{c}$ is equity-financed and (iii) increases with labor adjustment costs $f$ and $\gamma$, while actual cash levels decrease with $f$ and $\gamma$ for financially constrained firms. The strategy is to assume (i) and (ii) to solve for (iii), and verify (i) and (ii) ex post.

\begin{prop}
The optimal precautionary cash level $\bar{c}$ increases with $\gamma$ and $f$.
\\[ \frac{d\bar{c}}{d(\gamma + f)} > 0 \]
\end{prop}

\begin{prop}
Precautionary cash $\bar{c}$ is equity-financed.
\end{prop}

\begin{prop}
Financially constrained firms’ cash level $c$ and equity decreases with $\gamma$ and $f$.
\end{prop}
CHAPTER 1. MANAGING HUMAN CAPITAL RISK

Derivation of Optimal Cash Holdings

The following subsections outline direct proofs for the above propositions. Details are in the appendix. The Hamilton-Jacobi-Bellman (henceforth, HJB) equation reflects the shareholders’ tradeoff between the value of cash in and outside the firm.

\[-rV(c_t, \nu_t)dt + E[dV(c_t, \nu_t)] + \pi dt = 0 \quad (1.1)\]

Note that the “regime” \( \nu_t \) is a state variable in this model. We therefore need to separately derive value functions for the “high” and “low” regimes. We get a first-order linear differential equation for the value of the firm in the “low” regime.

\[\frac{dV_L(c)}{dc} (\nu_L - w) = V_L(c)(r + \lambda_{L,H}) - \lambda_{L,H} \cdot (V_H(\bar{c}) + c - \bar{c}) \quad (1.2)\]

We need one condition for its solution, as well as one condition each to determine the two unknowns \( V_H(\bar{c}) \) and \( \bar{c} \). Hence, we need three equations. The HJB for the “high” regime provides one.

\[-rV_H(\bar{c}) + \lambda_{H,L} \cdot (V_L(\bar{c}) - V_H(\bar{c})) + (\nu_H - w) = 0 \quad (1.3)\]

The second equation comes from the tradeoff between risk and return at the upper bound. The firm trades off the additional benefit of a marginal unit of cash – a reduction of risk – with the marginal cost – a higher interest expense.

\[r = \lambda_{H,L} \left( \frac{dV_L(\bar{c})}{dc} \right) - 1 \quad (1.4)\]

The continuation value that results from temporarily shutting down the firm provides the third and last equation.
\[ \bar{V} = V_L(c = f) = V_H(0) \cdot \int e^{-(r+\lambda_{L,H})t} \lambda_{L,H} dt = (V_H(\bar{c}) - \bar{c} - \gamma) \cdot \frac{\lambda_{L,H}}{r + \lambda_{L,H}} \] (1.5)

Some algebra, outlined in the appendix, reveals that we can solve analytically for the optimal cash level \( \bar{c} \).

\[ \bar{c} = f + E \cdot \ln \left( \frac{B(f + \gamma + E)}{D} \right) \] (1.6)

\( E, B, \) and \( D \) are positive constants that only depend on the given parameters, see the appendix.\(^\text{15} \) Thus, we have that the optimal cash level \( \bar{c} \) increases in \( f + \gamma \).

The next step is to check under which conditions the assumption holds that the firm never voluntarily fires workers. Intuitively, note that the only reason of the model firm to keep cash is to prevent a forced and inefficient “liquidation” of the human capital in the firm. Formally, if it is ever beneficial to fire workers, it must be beneficial when \( \nu_t = \nu_L \). The firm retains its employees if the value of following the above described trajectory is more valuable than firing workers, cashing out, while keeping the option to start a firm in the future. Formally, the no-firing condition is

\[ V_L(c) > -f + c + \bar{V} \] (1.7)

As \( \bar{V} = V_L(f) \), the condition amounts to comparing \( V_L(c) - c > V_L(f) - f \). Cash inside the firm is more valuable than cash outside the firm for all cash levels smaller than the optimum, \( \frac{dV_L(c)}{dc} > \frac{1}{B} > 1 \) for all \( c < \bar{c} \). Therefore, condition (1.7)\(^\text{15} \) \( B \) has a natural interpretation: it discounts payoffs that happen when the regime switches from “low” to “high” from the perspective of any point of time in the “low” regime. The effect of firing costs is stronger than that of hiring costs, because hiring costs are incurred only in the future, and are therefore discounted.
holds for all feasible cash levels \( c : f < c < \bar{c} \). A shareholder value maximizing firm never fires a worker, unless forced by a liquidity constraint.

**Optimal Financing of Precautionary Cash with Equity**

There are several ways to show that financing any portion of the optimal cash level \( \bar{c} \) with debt is suboptimal.\(^6\) The simplest sufficient reason is that in a recession, cash inside the firm is more valuable than cash outside the firm, see equation (1.4). Interest payments on outstanding debt would increase the cash outflow in precisely that regime. That cannot be optimal. The following alternative proof highlights why the optimal managing of human capital risk has macro-economic implications.

**Employment Effects**

If the funds for wage payments are debt-financed, the firm fires employees inefficiently early in recessions.\(^7\) Interest payments on debt increase the cash outflows in the “low” regime. As a consequence, the firm’s cash hits \( c = f \) earlier than it would have under equity financing. Hitting \( c = f \) necessitates “liquidating human capital” at a cost of \( f + B\gamma \). Suffering this loss earlier than necessary cannot be efficient. If the debt-financed firm wants to retain workers for the same amount of time as under equity financing, it has to begin with a higher cash level. As cash is costly, and there are no benefits to debt either in the high state or the low state, that cannot be efficient either. In sum, a firm that debt-finances wage payments

---

\(^6\)I seem to implicitly assume that the contract space is limited to equity or debt. A simple mechanism design exercise shows that this assumption is not restrictive. The optimal contract has to be “equity-like” – it needs to leave to investors a positive marginal benefit of retaining employees. Any contract that involves unconditional, fixed promises disqualifies.

\(^7\)Sharpe (1994) and Benmelech, Bergman, and Seru (2011) report evidence that higher leverage is associated with higher labor turnover.
fires workers too often. In fact, shareholders of firms that use risky debt will fire workers even before they are forced to do so by the liquidity constraint under certain conditions in order to protect cash, and thus delay bankruptcy.\textsuperscript{18} For similar reasons, a credit line may lead to firing before a hard constraint binds if there are bankruptcy costs.

Note that firing in downturns protects short-term profits, but destroys shareholder value. Sraer and Thesmar (2007) provide empirical evidence by showing that firms with supposedly lower agency conflicts have lower labor turnover and are more profitable.

\textbf{Derivation of the Constrained Firm’s Response}

As a financially constrained firm cannot actively influence its financial policy, the proof of proposition 1.3 amounts to describing the passive response to an increase of adjustment costs $\gamma + f$ in regime $L$. The only situation in which increased adjustment costs impact the firm’s cash and equity is when it is forced to fire employees. When firing is more costly, the cash outflow is higher for those firms that hit the lower bound $c = f$, which proves the first part. The firm’s cash outflow is tantamount to its losses, i.e., the reduction in book equity. The second part of the proposition is therefore a mechanical consequence of lower cash in response to higher adjustment costs.

\textsuperscript{18}Bankruptcy must be costly, adjustment costs must be sufficiently low, the wage level sufficiently high, and “recessions” sufficiently unlikely to end soon, among other things. As employment effects are not the focus of this paper, I omit the derivation and precise statement of this result.
CHAPTER 1. MANAGING HUMAN CAPITAL RISK

Empirical Predictions

The empirical treatment is unionization. I argue that unionization increases $\gamma + f$. First, unions make it harder to fire workers, i.e., unionization increases the administrative burden and explicit cost captured by $f$. Indeed, severance pay to employees after a plant closure is mentioned multiple times in the Alcoa annual report cited in the introduction. Second, in addition to impediments to firing, any standard text on the impact of unions on firms, such as Freeman and Medoff (1984), describes the work rules that are imposed with the establishment of a union. Among other things, certain jobs have to be filled with specially trained workers. The associated training costs are captured by $\gamma$. In sum, as both components $\gamma$ and $f$ are at least non-decreasing, it is safe to attribute higher total adjustment costs $\gamma + f$ to unionization.\(^{19}\)

Note that it is immaterial for the model predictions whether unionization increases the firm-specific human capital in the firm, or only labor adjustment costs. Both increase the implicit liability associated with the human capital in the firm, and therefore have the same effect on risk management.

As real-world firms can improve their liquidity position not only by raising cash, but also by liquidating assets, a more robust empirical prediction than effects of labor adjustment costs on cash levels are effects on cash-to-asset ratios. Similarly, equity levels should be scaled by the size of the firm. Fixing the firm’s non-cash physical assets at $A$, and assuming that a share $L < A$ of them is collateralized and financed with debt, the following empirical predictions immediately follow from the

\(^{19}\)Not only is the assumption of higher adjustment costs due to unionization established in the labor literature, the direct predictions are backed as well: Blanchflower and Bryson (2004) and others document that unionized workers better sustain recessions and that unionized workers and shops exhibit lower labor turnover.
model predictions for cash holdings and their financing, as derived above.

**Prediction 1:** For financially unconstrained firms, the causal effect of unionization (a positive shock to $f + \gamma$) on cash-to-asset ratios is positive.

\[
\frac{d}{d(\gamma + f)} \left( \frac{\bar{c}}{A + \bar{c}} \right) > 0
\]

The causal effect of unionization on net leverage is negative.

\[
\frac{d}{d(\gamma + f)} \left( \frac{L - \bar{c}}{A + \bar{c}} \right) < 0
\]

Note that there is no mechanical link between higher cash-to-asset ratios and lower net leverage in general. For example, debt-financed cash increases cash-to-assets but leaves net leverage constant. In contrast, a decrease of net leverage concurring with an increase of cash-to-assets means the cash is equity financed.

The mechanical reaction of financially constrained firm tends in the opposite direction.

**Prediction 2:** For financially constrained firms, the causal effect of unionization (a positive shock to $f + \gamma$) on cash-to-asset ratios is non-positive.

\[
\frac{d}{d(\gamma + f)} \left( \frac{c}{A + c} \right) \leq 0
\]

The causal effect of unionization on net leverage is non-negative.

\[
\frac{d}{d(\gamma + f)} \left( \frac{L - c}{A + c} \right) \geq 0
\]
I write weak inequalities here, because the firm is forced to fire employees only in that subset of all recessions in which it runs out of cash. Thus, we should expect this effect to be weaker empirically than the predicted effect for unconstrained firms, which applies to all possible situations in the “high” regime. The “Hidden Balance Sheet” (Figure 1.1) illustrates these predictions in an intuitive way.

**Limitations of the Model and Additional Predictions**

Unionization, of course, does other things to the firm than increasing firing and hiring costs. Existing results and the empirical results of this paper rule out potential drivers other than increased adjustment costs. I discuss a subset in this subsection.

Most importantly, unionization is thought of as increasing employees’ bargaining power. As discussed in the introduction, the empirical predictions of such an effect go in the opposite direction than mine. It is therefore an empirical question which effect matters more. My empirical results show that the causal effect go in favor of the human capital risk effect.

A typical conception is that unionization increases wages. While it is true that the wage level of an average unionized worker is higher than of an average non-unionized worker, neither I nor DiNardo and Lee (2004) (with a better measure of wages) find a causal effect of unionization on the wage bill and wages. That convinces me that the right way of modeling unionization is as an increase of labor adjustment costs, not as an increase of wages.

Aside from directly checking for a causal effect of unionization on wages, an implication of higher wages – higher operating leverage – can be used to check the validity of my modeling assumptions. As physical capital can be flexibly adjusted in my model, operating leverage is best defined as the wage bill divided by the sum
of the wage bill and interest expenses. An increase in in the wage bill would thus increase operating leverage. As I do not find a wage increase, I also do not expect that unionization has a causal effect on operating leverage. Indeed, there is no such effect. This distinction therefore constitutes a conceptual difference of my approach and Simintzi, Vig, and Volpin (2010)’s, who find that operating leverage as a result of increased labor protection crowds out financial leverage. I find lower financial leverage without an increase of operating leverage. Figure 1.1, as described in the introduction, illustrates this distinction.

Another concern is that unionization can change other off-balance sheet items, aside from the implicit insurance contract I have in mind. For example, the effects I observe could be driven by unions causing underfunding of pension plans (see Shivdasani and Stefanescu (2010)). However, if anything, one would expect the increased bargaining power through unionization to have the opposite effect. Furthermore, the stock market does not react to unionization per se (see Lee and Mas (2011)), putting further doubt on that explanation.

A related concern is that there may be a causal effect of unionization on non-wage benefits. Aside from the non-response of the stock market, DiNardo and Lee (2004)’s results are a very strong indication against that explanation. It seems unlikely that all of any additional bargaining power goes to non-wage benefits and none to the wage level.

Even absent these counterarguments, note that implicit liabilities do not necessarily increase with labor costs! While the per-period obligation would be higher, the duration during which the firm optimally bears the human capital risk declines with the wage level.
The lack of a stock market reaction also rules out that an increased probability of strikes drives the risk response.

Lastly, unionization may make it harder to renegotiate wages in bad times. As the wage level is constant in my model in the first place, it cannot reflect effects of a potential increase in the fixity of the wage level. The underlying economic effect of such a change is the same than the one proposed by an increase of the duration during which the firm is willing to insure the worker, however.

In sum, the potential effects of unionization (other than an increase in adjustment costs) that I thought of are either orthogonal to human capital risk management, unlikely to be drivers of the effect, or predict a corporate finance response in the opposite direction, making the answer an empirical concern.

Possible Model Extensions

The model is sufficient to produce first-order predictions on human capital risk management, but extensions in various dimensions can be a fruitful field for future research. In particular, the model omits two important dimensions: the worker’s decision and labor market competition between firms. As a result, human capital looks like firm-specific, non-collateralizable, and not tradable physical capital in my model; the only difference is that human capital does not feature on the observed balance sheet the corporate finance researcher observes, and is therefore hard to measure. It may be best compared to an off-balance sheet lease with adjustment costs.

The omission of the worker’s decision is deliberate in the basic model I present. One reason is that I want to show that human capital risk should indeed be managed similar to the risk stemming from physical asset that are visible on the balance sheet.
A second reason is that I do not want to obscure that the firm insures workers as a mechanical consequence of adjustment costs, and not because workers are risk averse, as is the case in earlier models of the interaction of corporate finance and labor.

Of course, human capital is quite different from highly specialized, illiquid machines in even slightly more general contexts. Informally speaking: employees have preferences – machines don’t. Specifically, competition between firms may endow workers with an outside option that is worth as much as the NPV of the implicitly promised future wage payments. Workers effectively become creditors to the firm. In such a case, their decision to stay or leave the firm depends on whether the implicit promise is credible. A necessary condition for the promise to be credible is that it be collateralized with liquid funds, such as cash. Thus, if the firm is at risk of becoming liquidity constrained (or is actively threatening to expose itself to that risk, as in the strategic security literature), workers may run on the firm. A firm with a less aggressive financial strategy may be able to attract these workers.

In addition, a conservative financial policy may give a risk-averse worker an incentive to invest in firm-specific human capital: if she does, she is more likely to be retained in bad times.20 According to practitioners, the motive to attract and keep employees is in fact much stronger than the motive to avoiding firing. An extension of the model that captures the effect of financial policy on relational contracts is thus the next logical step to advance this research agenda.

20Notice that the firm can never literally invest in human capital, firm-specific or not. It can only provide opportunities and incentives that make it optimal for the worker to learn. Both can be costly to the firm, but the expense does not yield an an explicit claim on the cash flows generated by the additional human capital.
1.3 Institutional Background, Research Design, and Data Set Assembly

1.3.1 Institutional Background and Previous RD Studies on the Unionization Response

According to the National Labor Relations Act, unionized workers are guaranteed the right to bargain collectively with their employer. Unionized workers are thought of as having greater bargaining power than they do as individuals. Union members may thus be able to extract greater benefits in terms of working conditions, grievance procedures, seniority rules, pensions, or wages (Freeman and Medoff (1984)). I give only the essential elements of how workers can establish a collective bargaining unit, as DL and LM have already provided a detailed description. Though there are other ways to form a union, the establishment of a new bargaining unit usually requires an election, in which workers vote for or against the establishment of union representation (see Farber and Western (2001)). In my data set, a bargaining unit is created if and only if a simple majority votes in favor of unionization. As described in the next subsection, the discontinuous change of the unionization “treatment” at the 50% vote share threshold allows a quasi-experimental research approach.

1.3.2 The Regression Discontinuity Method

Previous studies of the interaction between corporate financial policy and the presence of unions may be subject to endogeneity problems. Formally, financial policy $Y$ can be described by
$Y = \alpha + \beta D + \gamma X + \varepsilon$

where $D$ is a dummy for union presence, and $X$ represents other determinants of financial policy $Y$. The estimate of $\beta$ will generally be biased: it may be more desirable for a union to form in more or less profitable, growing or shrinking, large or small, etc. industries or firms. The same, possibly unmeasured, factors are likely to impact financial policy. In contrast, if the unionization “treatment” could be randomly assigned, we could simply compare “treated” and “untreated” groups to infer the causal effect. The regression discontinuity approach provides just such a quasi-random variation of the treatment variable, unionization.

The idea is to compare the corporate finance response to unionization in cases where the union barely won with those where the union barely lost an election. The difference between these two groups’ responses is the causal effect of unionization on changes of financial policy under two assumptions. First, the election cannot be rigged – whether an election is barely won or lost must be random. As the election is secret, and thus nobody knows the outcome until all ballots have been submitted, it seems reasonable to assume that the election is not precisely predictable. To reiterate, it is not required that the probability of a particular election outcome is not predictable at all. It must only not be predictable with absolute certainty, and only at the discontinuity. This assumption can be checked. If it is true, the ex ante distribution of vote shares in favor of unionization should be continuous. Figure 1.3 does not show a discontinuous jump at 50% in the full sample or either subsample. The second assumption is that factors driving changes in corporate finance decisions are orthogonal to whether the union barely wins or barely loses. For example, a change in the growth rate of the firm should not increase the
likelihood of a 49% election outcome over a 51% outcome. I eliminate such concerns as much as possible in the validity checks of the next section. Under these two assumptions, the regression discontinuity design is as good as a randomized experiment that assigns unionization to some firms but not to others. See Lee and Lemieux (2010) for more details. Differences in the two groups’ respective finance response to a unionization election can then be interpreted as the causal effect of unionization on financial policy.

1.3.3 Data Set Assembly

I combine annual CRSP/Compustat fundamentals from WRDS with the NLRB database of unionization elections. The latter was assembled and previously used by LM. It spans the years 1961-1999, with most elections in the earlier part of the sample.

The most important manipulation of the data concerns the vote share. As DL explain, small elections are less likely to provide a close outcome than are large ones. I thus follow their procedure, dropping elections with fewer than 100 votes cast, adapt the vote share for elections with an even number of votes cast to correct for a potentially remaining bias, and “bin” the remaining data in 20 equally spaced bins for both the graphical and quantitative analysis. The results are very similar if this manipulation is not made, or if a different number of bins is used.

As in LM, I retain only union certification elections and drop decertification events. An alternative would be to see whether decertification triggers a corporate finance response opposite to the one that certification triggers. Because that would lead to more “overlapping” elections (see next paragraph), I stay clear of such attempts.
Multiple elections can happen in the same firm. In some cases, they even overlap, i.e., the year “before” one can be the year of another election. As the elections are likely to take place in different establishments, I regard such observations as independent, as do LM. For the same reason, I measure the variables with error, so my estimates should be underestimated. Excluding elections that are less than three years apart does not fundamentally change the results.

I use S&P and Fitch long-term ratings, as well as S&P short-term credit ratings, to construct a one-dimensional ratings variable. This way, I can use more ratings observations than I could by relying on long-term ratings alone. I assign ”1” to AAA, ”2” to AAA-, and so forth for long-term ratings, and corresponding values for A-1+, A-1, etc. for short-term ratings. The mean of the three numbers constitutes my ratings variable.

Nine observations don’t pass an “accounting sanity check.” Their dividends are negative or larger than shareholder’s equity, or long-term debt exceeds total liabilities.

The dependent variable is the change of financial policy in response to unionization. I use log differences $\ln(y_{after}) - \ln(y_{before})$ to measure that change, which makes my results less susceptible to outliers. I try multiple definitions of “before” and “after.” In the results I present, $y_{before}$ is the average of the three previous years’ values, and $y_{after}$ is the average of the current and next years’ value. Alternatively, I tried all other combinations of the previous three years’ values for $y_{before}$, and either only the current year’s value, or only the next year’s value for $y_{after}$. The results are similar in terms of magnitude. As one would expect, there is more random variation for definitions using less data. Thus my choice to use the three-year window before and after the election.
Lastly, I investigate whether I can hinder outliers from affecting the results and reduce noise by conditioning or imputing the data. Winsorizing at 1% or 5% levels gives similar results in terms of magnitude as does non-imputed data, but, as expected, the winsorized versions yield greater significance. Restricting the data to a range of \( p_{50} \pm 5 \cdot (p_{75} - p_{25}) \), i.e., median plus/minus five times the interquartile range, gives very similar results. Those are the results I report. I end up with 3,572 observations for which I can construct all Compustat variables of interest. For the main regressions of cash-to-asset ratios and net leverage, however, I can match about 4,800 elections with the outcome variable.

The variable definitions are as follows. i. cash-to-asset ratio: cash and short-term investments (Compustat: CHE) divided by total assets (AT). ii. net leverage: total liabilities (LT) minus cash and short-term investments, divided by total assets. iii. operating leverage: cost of goods sold (COGS) plus selling, general, and administrative expense (XSGA), divided by total assets. The results are robust to alternative definitions of operating leverage. iv. market leverage: total liabilities divided by the market value of equity.

I present summary statistics for the main variables in Table 1.1 for the full sample and in Table 1.2 for the “unconstrained” and “constrained” subsamples.

**A Note on Sample Selection**

To test the effects of variations in labor adjustment costs and specific human capital on financial policy, one would hope to access data from an industry that mainly uses labor as an input and at the same time is subject to significant labor market frictions. Specialized law or consulting firms in remote areas, in recent years, would be examples of choice. In contrast, the data I have span the years 1961-1999,
and contain mainly manufacturing firms. In addition, I am implicitly conditioning on firms that have unionization elections. Those are probably the ones in which workers carry insufficient specific human capital to have strong bargaining power individually. Another reason to expect attenuated and noisy estimates is that elections are on an establishment level, while the effects I observe on the balance sheet are at the firm level. As a result of these challenges, I am likely to starkly underestimate the effects, and a quantitative assessment based on the existing results is challenging. A simple way would be to strip the sample of elections that happen in firms with many subsidiaries. In order to be transparent, I do not do that in this paper. Instead, I will have all the more faith in my results if I find significant effects even for those firms for which I least expect to find them a priori.

1.4 Empirical Results

The model calls for a positive causal effect of unionization on cash-to-asset ratios and a negative causal effect on net leverage in unconstrained firms, and a tendency for lower cash-to-assets and higher net leverage in constrained firms. The idea of the regression discontinuity design is to compare observations just below and just above the threshold that determines unionization. The difference between those firms’ responses is the causal effect of unionization.

A simple comparison of the average responses just to the right and left of the threshold would not give significant results due to the small number of observations falling into an arbitrarily small window. Therefore, a regression model needs to be fitted. There are two decisions to take: whether and how to restrict the sample to close elections, and which regression model to fit. Of course, the results should be
robust to that choice, and also to the measure of financial constraints I use. I first present a benchmark that follows LM and discuss robustness in its own subsection.

Before presenting the results, the following subsection discusses validity checks of the RD approach. The second subsection gives the main results, followed by a subsection of robustness checks, and finally a discussion of limitations of the empirical approach.

1.4.1 Preliminary Steps

The principal validity of the election I use has already been tested by DiNardo and Lee (2004) (DL) and Lee and Mas (2011) (LM), which allows me to omit some steps. DL show that the presumed discontinuity is indeed associated with a discontinuous assignment of unionization. I showed the continuity of the vote share distribution in the previous section. Moreover, there are no significant jumps at “fake thresholds” and at fake election dates. Lastly, the key variables of firms that end up getting unionized (closely) are not significantly different from those of firms that end up escaping unionization.\textsuperscript{21} These statements are all true for the full sample, as well as for the unconstrained and constrained subsample according to the Whited-Wu index of financial constraints risk (“WI,” Whited and Wu (2006)).

\textsuperscript{21}The size (by assets) of constrained firms is smaller than the size of unconstrained firms by construction – smaller firms are more likely to be financially constrained. However, the distributions of the dependent variables, i.e., changes in cash-to-asset ratios and net leverage, are similar, both for the constrained and unconstrained subsets.
1.4.2 Causal Effects

Figure 1.4 shows the regression discontinuity plot for unconstrained firms’ cash-to-asset ratios, net leverage, operating leverage, and market leverage.\footnote{One would guess from looking at the dots alone that a locally linear regression that puts most weight on close outcomes would provide the strongest results. To preserve comparability with LM and transparency, however, I fit a sixth-order polynomial. The results are robust to that choice, as discussed in the robustness checks.} Cash-to-assets clearly jump at the discontinuity, and net leverage drops. Operating leverage, as defined by Novy-Marx (2011), is stable and therefore not the driver of the cash and leverage response. Market leverage is stable as well. There are two messages to take away: first, we observe a risk management response over and above a response of capital structure. Second, operating leverage is not the risk factor to which financial policy responds.

Figure 1.5 shows that constrained firms’ response is more noisy, but qualitatively reversed. If there is a significant causal effect of unionization on constrained firms’ cash-to-asset ratios, it is negative, and positive for net leverage. Remarkably, operating leverage seems to increase, if not in a statistically significant way. Certainly, variation in operating leverage does not positively explain variation in cash-to-assets and net leverage in this subset either. Aside, the wage bill is roughly stable for both subsamples as well, which further undermines operating leverage-based explanations for the observed corporate finance response.\footnote{The wage bill response is noisy in my data, so my non-finding of an effect could be due to bad data, and I do not show it here. It is better to rely on DL’s finding (no causal effect), because they combine the unionization election data with better wage data.} Market leverage sees a small, yet not significant, increase. As LM have already shown that there is no significant effect on the market value of equity, I omit reporting these estimates.
For the full sample of firms, there is no discontinuity, as the opposing causal effects of the constrained and unconstrained subsample cancel out, as Figure 1.6 shows. The reason is that the causal effects of the unconstrained and constrained subsamples cancel out when aggregated.

Table 1.3 confirms the qualitative results quantitatively. I report the union “win” coefficient of the same polynomial regression that featured in the plots as the discontinuity. Unconstrained firms significantly raise their cash-to-asset ratios by about 5%, while net leverage decreases by close to 2%. There is no significant effect for operating leverage. Constrained firms’ cash-to-assets increase and net leverage falls, but insignificantly. Confounded by the noisy response of constrained firms, there are no significant causal effects for the average firm.\(^{24}\)

In sum, the causal effects are precisely as the model predicts, and there are no causal effects that the theory did not predict. Specifically, financially unconstrained firms significantly increase equity-financed cash, while constrained firms show an ambiguous response, with a tendency to eat through their cash cushion, thus pushing up net leverage. Operating leverage was predicted not to move, and it does not move in either case. The economic significance of the results is best judged from the market response to sub-optimal risk management. Constrained firms get punished with an (insignificant) 5% decline in market value, supposedly for not responding to the same shock that leaves unaffected the market-to-book ratio of unconstrained

\(^{24}\)Not shown are results for the components of the ratios, as well as other variables. As discussed in the model section, unconstrained firms can respond to higher risk by reducing non-cash-assets and debt, and thus raise cash-to-asset ratios and reduce financial leverage. Indeed, a 6% increase in cash is insignificant, while assets drop 4.6% and total liabilities drop 5.4%; long-term debt drops 13%. The latter three are highly significant. A 5% reduction in firm value can be almost entirely ascribed to the reduced firm size. In contrast, constrained firms, by definition, have exhausted their collateralizable assets, so they cannot raise new cash by issuing a risk-free security. Also, selling assets does not help these firms to improve their cash and capital ratios. Indeed, they do not shrink. Assets increase 1%, cash falls 11%, and debt increases by 5%, all insignificantly.
firms, who optimally manage the increased human capital risk. These numbers may appear to be small, given that the average cash-to-asset ratio of the sample firm is only about 7%. However, these estimates capture the firm’s response to an establishment-level unionization election. While it is hard to quantify the attenuation, it is clear that the estimates must be regarded as a lower bound of the true economic effects.

1.4.3 Average Effects

Table 1.4 reports differences in means of the average unionized and non-unionized firm for the two subsamples, as well as the full sample. An OLS regression of the outcome variable on unionization with the slope forced to zero yields these non-causal correlations. Unionized firms on average have 2% lower cash-to-assets, 1% higher net leverage, and 3% higher market leverage. These correlations are driven by constrained firms; there is no significant correlation for unconstrained firms.

Let me summarize. First, I find a causal response towards a more conservative financial policy in response to unionization, but only in firms that can freely set their financial policy. Second, I can reconstruct existing results of the literature by not using the identification strategy at hand, but rather by estimating non-causal correlations, to exclude the possibility that different things are true for different data sets. I conclude that the difference in results is due to the different estimation technique, and not different data sets.

1.4.4 Robustness Checks

The choice of the regression function can drive RD results. Further, the results could be driven by outliers far away from the threshold that affect the estimated
functional form. To check robustness in this dimension, I run kernel regressions (as a special case of locally linear regressions) and polynomial regressions of orders one to six, and I restrict the data in windows varying from 5% to 35% distance from the threshold. Of course, I lose significance with some variables in some specifications due to the reduced number of observations, but the general result is robust.

Table 1.5 shows that an OLS regression on elections within a 20% vote share window to the threshold produces virtually the same results as a polynomial regression on the full sample from Table 1.3. Although there are particular reasons with this data set to construct the bandwidth as it is (discussed in DL), I also vary this dimension, and get similar results (not shown).

Another concern is that splitting the sample at the median somehow drives the result. To investigate this possibility in more detail, I split the sample into six non-overlapping subsamples according to the Whited-Wu index, and estimate the causal effect of unionization on cash-to-assets and net leverage with the same linear regression technique on the observations in the 20% window as above. Graph 1.7 reports the estimates. The effects seem quite monotonic with the severity of financial constraint risk.

Next, it might be the particular measure of financial constraints I employ that drives the results. Table 1.6 gives evidence to the contrary. I run a linear regression on the same set of close elections as above, but without splitting the sample. Instead, I construct a dummy for each type of financial constraint. If an observation belongs to a firm that is in the upper half of the data with regard to the respective measure financial constraints, the dummy takes the value 1. Apart from the Whited index (WI), I use Kaplan and Zingales (1995)’s index of financial constraints (KZ), credit ratings, total assets (AT), and a dividend payer dummy. As before, uncon-
strained firms increase cash and reduce net leverage in response to unionization, while constrained firms do the opposite, although I do not get significant results for all specifications. One could perform a principle components analysis to construct one single measure of financial constraints risk. But as the different measures already work individually, there is no need to sacrifice transparency. In sum, the results are robust to the choice of financial constraint proxy.

I do not present other standard RD tests that I performed. We saw already that not all variables jump at the threshold, e.g., operating leverage. Moreover, variables that do jump at the threshold do not exhibit significant jumps at other, “fake” thresholds. Baseline variables that are determined prior to the assignment, e.g., differences in cash-to-assets before the election, do not exhibit significant discontinuities either. These tests conclude the standard RD diagnostics.

1.4.5 Remaining Limitations

Despite the robustness of the results, some limitations of course remain. I briefly discuss potential caveats of the identification strategy, then data limitations, then external validity, and finally possible further tests of the model.

One can think of reasons why we observe the above-described corporate finance response without unionization and financial constraints being the cause. I discuss one. Just-so unionized firms may not be “really” unionized. Union support may have been weak enough to trigger a successful decertification election shortly after the certification, and firms may anticipate this. The observed cash response may then be a sign of relief by the firm that the union “almost lost” the election. Although my data set contains decertification elections, I cannot directly test this hypothesis. I know that the elections are on an establishment level, but I do not
see an establishment identifier for a given election. So a decertification at date \( t' \)
may – or may not – undo a certification election in the same firm at date \( t < t' \).
But as excluding observations for which an election closely follows another does not change the basic results, I am confident this is not the driver.

Second, even if unionization and financial constraints are the ultimate causes of the observed effects, the mechanism the model proposes may not be correct. The intuition of the model rests on predictions about labor turnover (risk). I can neither reject nor verify them with my data set. I have only a crude measure of employment. Moreover, calculating differences between year-on-year employment yields only net changes of the firm’s workforce. But the model predictions are about gross turnover, which might be quite different between two firms whose employees have a different skill distribution. Even if I had gross turnover figures, telling firing apart from quitting would be a non-trivial challenge. An employer-employee database would be needed to dig into this question.

External validity is often discussed as a challenge to the regression discontinuity methodology. As argued in section 1.3, selection for manufacturing firms prone to unionization in their early years should work against my case, as the human capital contributed by workers in my sample should be comparatively low. That I find effects nevertheless should strengthen confidence in my results. Yet, a similar study with data from different sectors and countries would of course be helpful. Second, RD estimates are always weighted average treatment effects, where the weights are the ex ante probability that the value of an individual’s assignment variable will be in the neighborhood of the threshold. I avoid that problem as much as possible by excluding small elections and replicating the vote share manipulation in DL and LM, but I can, of course, not completely reject that criticism.
I discussed limitations of the model at the end of section 3.2. Taking the model as it is, checking the validity of other obvious predictions outside the unionization context would help increase confidence in the intuition it conveys. Using measures of levels of human capital in firms as in Abowd, Haltiwanger, Jarmin, Lane, Lengermann, McCue, McKinney, and Sandusky (2005) may help investigate whether human capital risk is an omitted variable in cross-sectional and time series regressions of cash-to-asset ratios (Opler, Pinkowitz, Stulz, and Williamson (1999), Bates, Kahle, and Stulz (2009)). (This proposition assumes a positive correlation between general and firm-specific human capital on the worker- or firm-level.)

The field for further consistency checks widens if the model is extended in several dimensions, some of which I have discussed before. Asset pricing implications can be cross-checked with Gourio (2007). Labor economists may be interested in the prediction that employees have an incentive to invest in specific human capital, not (only) because it may allow them to earn higher wages, but because it buys them an insurance policy against fluctuations in their general human capital. Differences in the degree of alignment of managers’ and shareholders’ interests, e.g., the time horizon, as well as differences in the liquidity position of the firm can make the workers’ strategy more or less effective. A possible test of this hypothesis would be a comparison of public and private firms’ employment decisions. Lastly, if the firm is sufficiently human capital-intensive and the adjustment costs high enough, the conflict between workers and shareholders may be best resolved by letting employees own the firm, as Hansmann (2000) proposes in more general terms. Consulting firms are a good example.
1.5 Conclusion

To gain basic intuitions on managing human capital risk, employees are best thought of as valuable but illiquid assets of the firm. The firm optimally retains these “assets” when the value of their output is temporarily low. The firm thus insures employees against fluctuations in the value of their human capital. Adjustment costs for replacing the implicit asset “human capital” are the underlying reason, and equity-financed cash is the efficient tool to manage that risk. Corporate finance decisions based on the view that employees are an outside supplier of services rather than a part of the firm lead to suboptimal risk management with too much debt and too little cash. The ensuing inefficiency through excessive firing in recessions destroys shareholder value. The view of employees as external suppliers of services is only valid if they are perfectly exchangeable and replaceable at no cost. My empirical results indicate that this is not the case even for shop floor workers in manufacturing firms. Even in that sample, the benefits of managing human capital risk outweigh possible benefits of debt for strategic reasons in response to the change of the firm-worker relationship represented by unionization. A strategic element in the interaction between corporate finance and labor may be that low debt and high cash signal credibly that the firm is willing and able to insure workers commensurate to their firm-specific skills. Conservative financial policies may thus help attract and retain valuable employees, and give them incentives to invest in firm-specific human capital.

The paper leaves for future research (i) investigating theoretically the firm-worker relationship with a focus on labor adjustment costs and the resulting risk transfer in a more comprehensive way than the present model does, (ii) the identification of other implicit risk factors associated with the firm’s relationship to
stakeholders, such as customer relationships, and (iii), based on the risk factors identified, revisiting the literature on cash holdings and capital structure both in the cross-section and time-series using measures of replacement costs as a proxy for implicit liabilities.
CHAPTER 1. MANAGING HUMAN CAPITAL RISK

Appendix

Derivation of Equation (1.1)

The value of the firm is

\[ V_t = E_t[\pi \Delta t + \frac{V_{t+1}}{1 + r \Delta t}] \]

For small \( \Delta t \)

\[ V_t = \pi \Delta t + E_t[V_{t+\Delta t}](1 - r \Delta t) \]

Rearranging yields

\[ 0 = \pi \Delta t + E_t[V_{t+\Delta t} - V_t] - r E_t[V_{t+\Delta t}] \Delta t \]

For small \( \Delta t \),

\[ 0 = \pi dt + E_t[dV_t] - r V_t \]

Solving the ODE

It is helpful to first derive an accounting identity. As recapitalization is costless, the firm must be worth the same before and after raising cash, for any starting level \( c \).

\[ V_H(c) = V_H(\bar{c}) + (c - \bar{c}) \quad (1.8) \]

Using (1.8) in (1.1) yields the ODE (1.2). Its solution is the sum of a particular and a general part, \( V_L(c) =: V_{L,p} + V_{L,g} \). I choose to write the particular solution as
\[ V_{L,p}(c) = A + B \cdot (c - \bar{c}) \]

where \( A \) and \( B \) are constants that can be determined from plugging the particular solution into equation (1.2). They and all following constants are defined as to be positive. They are written out in terms of exogenous parameters at the end of this appendix. The general solution is

\[ V_{L,g}(c) = -D \cdot \exp\left(-E^{-1} \cdot (c - \bar{c})\right) \]

where \( E := -\frac{\nu - w}{r + \lambda_{H,L}} \), and \( D \) is a constant to be determined. We are now left with three unknowns \( D, \bar{c}, V_H(\bar{c}) \), so we need three equations to solve the problem: equations (1.3), (1.4), and (1.5), whose derivation is given below. It turns out that \( D \) and \( V_H(\bar{c}) \), and thus \( V_L(\bar{c}) \), are constants that can be written entirely as a function of the parameters. We can thus derive an analytical solution, as given in the main text.

**Derivation of Equation (1.4)**

Shareholders trade off the respective values inside and outside the firm of a marginally higher optimal cash level \( \Delta \bar{c} \), given that the firm is in the “high” regime. Almost surely, there will be no recession next period (the switching probability is \( \lambda_{H,L} dt \)). The marginal cost of holding cash is \( \Delta \bar{c} \). The marginal benefit is \( \lambda_{H,L} (V_L(\bar{c} + \Delta \bar{c}) - (V_L(\bar{c}) + \Delta \bar{c})) \). Define

\[ \Phi(\Delta c) = [-r \Delta \bar{c} + \lambda_{H,L} (V_L(\bar{c} + \Delta \bar{c}) - (V_L(\bar{c}) + \Delta \bar{c}))] dt \]

Per definition of the optimal cash level, \( \Phi'(\Delta c)|_{\Delta c=0} \), which yields equation (1.4).
Derivation of Equation (1.5)

Restarting the firm when the state switches back to “high” is possible at cost \( \gamma \geq 0 \). The value then, by the accounting identity from the beginning, is \( V_H(0) = V_H(\bar{c}) - \bar{c} - \gamma \). At any point during the “low” state, this future prospect must be discounted with the expected waiting time. The constant \( B \) turns out to capture that discount.

**Derived constants, and constants defined for notational convenience, in terms of model parameters**

\[
A = B \left( -E + \frac{E + E'}{1 - B \cdot B'} \right)
\]

\[
B = \frac{\lambda_{L,H}}{r + \lambda_{L,H}}
\]

\[
B' := \frac{\lambda_{H,L}}{r + \lambda_{H,L}}
\]

\[
E := \frac{-\nu_L - w}{r + \lambda_{L,H}}
\]

\[
E' := \frac{\nu_H - w}{r + \lambda_{H,L}}
\]

\[
D = (B'-1 - B) \cdot E
\]

\[
V_H(\bar{c}) = \frac{E' - E}{1 - B \cdot B'}
\]
\[ V_L(\bar{c}) = \frac{B \cdot E' - B'^{-1} \cdot E}{1 - B \cdot B'} \]
Figure 1.1: The firm’s economic balance sheet. The white part is the observed balance sheet. The grey part is the unobserved “Hidden Balance Sheet.” “A” is non-cash assets. “CH” is cash. “L” is total liabilities. “E” is book equity. “IA” is implicit off-balance sheet assets. “IL” is implicit off-balance sheet liabilities. “IE” is the residual of “IA” minus “IL”. An increase in off-balance sheet implicit liabilities triggers an on-balance sheet risk-management response: the cash-to-asset ratio increases and net leverage drops.
Figure 1.2: The evolution of cash. The firm pays dividends as to keep the optimal cash level $c$ while financially unconstrained in the “high” regime, $H$. The firm depletes its cash buffer in the “low” regime, $L$. Endogenously, the firm fires employees only if cash is depleted, i.e., at $c = f$. When the state switches to $H$, the firm recapitalizes immediately as to reach the optimal cash level. Firing and/or hiring employees is costly.
Figure 1.3: Vote share distributions for the full sample (top), unconstrained firms (center), and constrained firms (bottom). The 20 vote share bins are equally spaced and constructed as in Lee and Mas (2011). The unconstrained/constrained split is made at the sample median of the WI index. The identifying assumption of the RD design is that no party can precisely control the election outcome right and left of the threshold. In other words, there should not be systematic sorting of firms, or subsets of firms, to the right or left of the threshold. Such sorting would be visible from a discontinuity in the vote share distribution at the 50% vote share threshold. The distributions are quasi-continuous at the 50% vote share, also if the sample is split according to industry or other variables.
Figure 1.4: Election response of financially unconstrained firms’ cash-to-asset ratios (c2a), net leverage (bln), operating leverage (ol), and market leverage (mlev). The dots are the average value of the same bins as in Figure 1.3. Winning margin is vote share minus 50%. Net leverage is debt minus cash divided by total assets. Operating leverage is defined as in Novy-Marx (2011). Market leverage is total liabilities divided by the sum of market value of equity and total liabilities. “dev” indicates log differences. Unconstrained means being below the sample median according to the WI index. The right hemisphere shows the response of firms in which a union gets established. Left is the response of union-loss firms. The solid line represents predicted values based on a sixth-order polynomial regression; the dotted lines represent 95% confidence intervals. As multiple elections can happen in the same firm, standard errors are clustered at the firm level. A discontinuity of the outcome variable at the 50% threshold represents the causal effect of unionization.
Figure 1.5: Election response of financially constrained firms’ cash-to-asset ratios (c2a), net leverage (bln), operating leverage (ol), and market leverage (mlev). All definitions are parallel to Figure 1.4.
Figure 1.6: Election response of the average sample firm’s cash-to-asset ratios (c2a), net leverage (bln), operating leverage (ol), and market leverage (mlev). All definitions are parallel to Figure 1.4.
Figure 1.7: Estimated causal effects of unionization on cash-to-assets (c2a; blue) and net leverage (bln; red) from linear regressions of the dependent variables on a union “win” dummy, winning margin, and their interaction, for six non-overlapping subsamples of the Whited-Wu index of financial constraints risk, for elections within a 20% vote share window of the threshold. Higher groups are more financially constrained.
Table 1.1: Summary statistics for the full sample.

<table>
<thead>
<tr>
<th>Variable</th>
<th>mean</th>
<th>p50</th>
<th>sd</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Assets</td>
<td>3476.993</td>
<td>767.2685</td>
<td>10161.46</td>
<td>1.341</td>
<td>217123.4</td>
</tr>
<tr>
<td>Cash</td>
<td>210.6329</td>
<td>33.083</td>
<td>648.4461</td>
<td>0</td>
<td>11044.3</td>
</tr>
<tr>
<td>Total Liabilities</td>
<td>2040.225</td>
<td>353.4785</td>
<td>7335.525</td>
<td>.473</td>
<td>193777.9</td>
</tr>
<tr>
<td>Net Liabilities</td>
<td>1829.592</td>
<td>299.4805</td>
<td>6967.059</td>
<td>-206.184</td>
<td>182733.6</td>
</tr>
<tr>
<td>Book Equity</td>
<td>1409.587</td>
<td>387.4665</td>
<td>3342.523</td>
<td>.868</td>
<td>43542</td>
</tr>
<tr>
<td>Market Cap</td>
<td>2685.934</td>
<td>525.6165</td>
<td>7272.341</td>
<td>.3036879</td>
<td>130562.3</td>
</tr>
<tr>
<td>Employees</td>
<td>60.5321</td>
<td>22.18</td>
<td>115.9008</td>
<td>.003</td>
<td>853</td>
</tr>
<tr>
<td>Cash-to-Assets</td>
<td>.0667017</td>
<td>.0457799</td>
<td>.0629211</td>
<td>0</td>
<td>.5488421</td>
</tr>
<tr>
<td>Net Leverage</td>
<td>.4237899</td>
<td>.4373283</td>
<td>.1607228</td>
<td>-.2349006</td>
<td>.9144532</td>
</tr>
<tr>
<td>Operating Leverage</td>
<td>2.886831</td>
<td>2.472512</td>
<td>1.671889</td>
<td>.2783285</td>
<td>11.22916</td>
</tr>
<tr>
<td>Market Leverage</td>
<td>1.146872</td>
<td>.7384894</td>
<td>1.426595</td>
<td>.0240333</td>
<td>25.02227</td>
</tr>
<tr>
<td>Delta-c2a</td>
<td>-.0366708</td>
<td>-.0372289</td>
<td>.2213158</td>
<td>-.944753</td>
<td>1.044239</td>
</tr>
<tr>
<td>Delta-bln</td>
<td>.0261387</td>
<td>.0243548</td>
<td>.0875587</td>
<td>-.4146557</td>
<td>.4324271</td>
</tr>
<tr>
<td>Delta-ol</td>
<td>.0412287</td>
<td>.0320761</td>
<td>.1921128</td>
<td>-.949031</td>
<td>1.026049</td>
</tr>
<tr>
<td>Delta-mlev</td>
<td>.3261518</td>
<td>.1371515</td>
<td>.7657016</td>
<td>-.9513371</td>
<td>3.960541</td>
</tr>
<tr>
<td>Observations</td>
<td>3572</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1.2: Summary statistics for key variables for the “unconstrained” subsample (Panel A) and the “constrained” subsample (Panel B). The number of observations indicates the number of election observations with a complete set of Compustat variables.

<table>
<thead>
<tr>
<th>Panel A: Unconstrained firms</th>
<th>mean</th>
<th>p50</th>
<th>sd</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Assets</td>
<td>5879.01</td>
<td>2012.4</td>
<td>13054.23</td>
<td>198.174</td>
<td>217123.4</td>
</tr>
<tr>
<td>Cash</td>
<td>354.8466</td>
<td>95.705</td>
<td>819.2358</td>
<td>0</td>
<td>11044.3</td>
</tr>
<tr>
<td>Total Liabilities</td>
<td>3441.704</td>
<td>991.7</td>
<td>9573.407</td>
<td>37.807</td>
<td>193777.9</td>
</tr>
<tr>
<td>Net Liabilities</td>
<td>3086.857</td>
<td>859.809</td>
<td>9131.027</td>
<td>-206.184</td>
<td>182733.6</td>
</tr>
<tr>
<td>Book Equity</td>
<td>2389.427</td>
<td>962.441</td>
<td>4181.281</td>
<td>134.798</td>
<td>43542</td>
</tr>
<tr>
<td>Market Cap</td>
<td>4656.162</td>
<td>1615.959</td>
<td>9266.849</td>
<td>104.247</td>
<td>130562.3</td>
</tr>
<tr>
<td>Employees</td>
<td>99.52757</td>
<td>47.664</td>
<td>140.6703</td>
<td>0</td>
<td>853</td>
</tr>
<tr>
<td>Cash-to-Assets</td>
<td>.0678315</td>
<td>.048895</td>
<td>.0608757</td>
<td>0</td>
<td>.4696362</td>
</tr>
<tr>
<td>Net Leverage</td>
<td>.4279353</td>
<td>.4404667</td>
<td>.514072</td>
<td>-.175484</td>
<td>.8450245</td>
</tr>
<tr>
<td>Operating Leverage</td>
<td>2.652092</td>
<td>2.298625</td>
<td>1.508323</td>
<td>.302783</td>
<td>11.04185</td>
</tr>
<tr>
<td>Market Leverage</td>
<td>.864508</td>
<td>.6432391</td>
<td>.8525593</td>
<td>.0243033</td>
<td>12.48054</td>
</tr>
<tr>
<td>Delta-c2a</td>
<td>-.042581</td>
<td>-.0427978</td>
<td>.214954</td>
<td>-0.9403229</td>
<td>1.044239</td>
</tr>
<tr>
<td>Delta-bln</td>
<td>.0267235</td>
<td>.0248315</td>
<td>.0802867</td>
<td>.0802867</td>
<td>.4010054</td>
</tr>
<tr>
<td>Delta-ol</td>
<td>.0281471</td>
<td>.02956</td>
<td>.1676328</td>
<td>-.8634019</td>
<td>.9647384</td>
</tr>
<tr>
<td>Delta-mlev</td>
<td>.2445613</td>
<td>.1183753</td>
<td>.6293721</td>
<td>-.8831711</td>
<td>3.783424</td>
</tr>
<tr>
<td>Observations</td>
<td>1959</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Constrained firms</th>
<th>mean</th>
<th>p50</th>
<th>sd</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Assets</td>
<td>559.7255</td>
<td>150.793</td>
<td>2493.959</td>
<td>1.341</td>
<td>38024</td>
</tr>
<tr>
<td>Cash</td>
<td>35.48431</td>
<td>6.5</td>
<td>245.5708</td>
<td>0</td>
<td>5149</td>
</tr>
<tr>
<td>Total Liabilities</td>
<td>338.1192</td>
<td>66.606</td>
<td>1610.95</td>
<td>.473</td>
<td>25071</td>
</tr>
<tr>
<td>Net Liabilities</td>
<td>302.6349</td>
<td>56.384</td>
<td>1415.316</td>
<td>-23.715</td>
<td>19922</td>
</tr>
<tr>
<td>Book Equity</td>
<td>219.5638</td>
<td>77.289</td>
<td>964.59</td>
<td>.868</td>
<td>17814.6</td>
</tr>
<tr>
<td>Market Cap</td>
<td>293.0778</td>
<td>84.1743</td>
<td>1550.818</td>
<td>.3036879</td>
<td>47639.19</td>
</tr>
<tr>
<td>Employees</td>
<td>13.17183</td>
<td>5.757</td>
<td>40.36991</td>
<td>0.003</td>
<td>746</td>
</tr>
<tr>
<td>Cash-to-Assets</td>
<td>.0653295</td>
<td>.0428291</td>
<td>.0653119</td>
<td>0</td>
<td>.5488421</td>
</tr>
<tr>
<td>Net Leverage</td>
<td>.4187552</td>
<td>.4345657</td>
<td>.1712699</td>
<td>-.2349006</td>
<td>.9144532</td>
</tr>
<tr>
<td>Operating Leverage</td>
<td>3.171924</td>
<td>2.693193</td>
<td>1.81121</td>
<td>.2783285</td>
<td>11.22916</td>
</tr>
<tr>
<td>Market Leverage</td>
<td>1.489805</td>
<td>.8992655</td>
<td>1.84689</td>
<td>.0405119</td>
<td>25.02227</td>
</tr>
<tr>
<td>Delta-c2a</td>
<td>-.0294929</td>
<td>-.029604</td>
<td>.2286663</td>
<td>-.944753</td>
<td>.9374832</td>
</tr>
<tr>
<td>Delta-bln</td>
<td>.0254284</td>
<td>.0236576</td>
<td>.095671</td>
<td>-.4146557</td>
<td>.4324271</td>
</tr>
<tr>
<td>Delta-ol</td>
<td>.0571163</td>
<td>.035959</td>
<td>.2171786</td>
<td>-.949031</td>
<td>1.026049</td>
</tr>
<tr>
<td>Delta-mlev</td>
<td>.4252441</td>
<td>.18131</td>
<td>.8942926</td>
<td>-.9513371</td>
<td>3.960541</td>
</tr>
<tr>
<td>Observations</td>
<td>1613</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1.3: Causal effect of unionization on changes of cash-to-asset ratio (c2a), net leverage (net lev), operating leverage (ol), and market leverage (mlev). Coefficients on the union-win dummy from sixth-order polynomial regressions on the winning margin and win dummy. “Unconstrained” means being in the lower half of the sample according to the Whited-Wu index of financial constraints; the definition of “constrained” is symmetric. Standard errors are clustered at the firm level.

<table>
<thead>
<tr>
<th>Panel A: Unconstrained firms</th>
<th>Δ(c2a)</th>
<th>Δ(net lev)</th>
<th>Δ(ol)</th>
<th>Δ(mlev)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Union-win</td>
<td>0.0532**</td>
<td>-0.0180**</td>
<td>0.0175</td>
<td>-0.0130</td>
</tr>
<tr>
<td></td>
<td>(0.0209)</td>
<td>(0.00809)</td>
<td>(0.0195)</td>
<td>(0.0751)</td>
</tr>
<tr>
<td>Observations</td>
<td>2,441</td>
<td>2,435</td>
<td>2,146</td>
<td>2,313</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Constrained firms</th>
<th>Δ(c2a)</th>
<th>Δ(net lev)</th>
<th>Δ(ol)</th>
<th>Δ(mlev)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Union-win</td>
<td>-0.0346</td>
<td>0.0150</td>
<td>0.0352</td>
<td>0.145</td>
</tr>
<tr>
<td></td>
<td>(0.0263)</td>
<td>(0.0110)</td>
<td>(0.0261)</td>
<td>(0.105)</td>
</tr>
<tr>
<td>Observations</td>
<td>2,437</td>
<td>2,362</td>
<td>2,056</td>
<td>2,069</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: Full sample</th>
<th>Δ(c2a)</th>
<th>Δ(net lev)</th>
<th>Δ(ol)</th>
<th>Δ(mlev)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Union-win</td>
<td>0.00541</td>
<td>-0.00135</td>
<td>0.0236</td>
<td>0.0528</td>
</tr>
<tr>
<td></td>
<td>(0.0173)</td>
<td>(0.00694)</td>
<td>(0.0162)</td>
<td>(0.0626)</td>
</tr>
<tr>
<td>Observations</td>
<td>4,886</td>
<td>4,800</td>
<td>4,207</td>
<td>4,384</td>
</tr>
</tbody>
</table>

*** p<0.01, ** p<0.05, * p<0.1
Table 1.4: Non-causal correlations between unionization and changes of cash-to-asset ratio (c2a), net leverage (net lev), operating leverage (ol), and market leverage (mlev). OLS regression of the outcome variable on a union-win dummy, with the slope constrained to zero. The definition of “(un-)constrained” is according to Whited-Wu, as before. Standard errors are clustered at the firm level.

<table>
<thead>
<tr>
<th>Panel A: Unconstrained firms</th>
<th>Δ(c2a)</th>
<th>Δ(net lev)</th>
<th>Δ(ol)</th>
<th>Δ(mlev)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Union-win</td>
<td>-0.00879</td>
<td>0.00505</td>
<td>-0.00948</td>
<td>0.0285</td>
</tr>
<tr>
<td></td>
<td>(0.0110)</td>
<td>(0.00399)</td>
<td>(0.0111)</td>
<td>(0.0315)</td>
</tr>
<tr>
<td>Observations</td>
<td>2,441</td>
<td>2,435</td>
<td>2,146</td>
<td>2,313</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Constrained firms</th>
<th>Δ(c2a)</th>
<th>Δ(net lev)</th>
<th>Δ(ol)</th>
<th>Δ(mlev)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Union-win</td>
<td>-0.0276***</td>
<td>0.0104**</td>
<td>0.0195</td>
<td>0.0270</td>
</tr>
<tr>
<td></td>
<td>(0.0105)</td>
<td>(0.00478)</td>
<td>(0.0123)</td>
<td>(0.0435)</td>
</tr>
<tr>
<td>Observations</td>
<td>2,437</td>
<td>2,362</td>
<td>2,051</td>
<td>2,069</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: Full sample</th>
<th>Δ(c2a)</th>
<th>Δ(net lev)</th>
<th>Δ(ol)</th>
<th>Δ(mlev)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Union-win</td>
<td>-0.0188**</td>
<td>0.00812**</td>
<td>0.00638</td>
<td>0.0322</td>
</tr>
<tr>
<td></td>
<td>(0.00738)</td>
<td>(0.00315)</td>
<td>(0.00824)</td>
<td>(0.0262)</td>
</tr>
<tr>
<td>Observations</td>
<td>4,886</td>
<td>4,800</td>
<td>4,207</td>
<td>4,384</td>
</tr>
</tbody>
</table>

*** p<0.01, ** p<0.05, * p<0.1
Table 1.5: OLS regression of the outcome variable on a union “win” dummy (reported), the vote share, and their interaction. The sample is restricted to elections within 20% of the unionization threshold. Standard errors are clustered at the firm level.

<table>
<thead>
<tr>
<th>Panel A: Unconstrained firms</th>
<th>∆(c2a)</th>
<th>∆(net lev)</th>
<th>∆(ol)</th>
<th>∆(mlev)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Union-win</td>
<td>0.0494***</td>
<td>-0.0147**</td>
<td>0.0131</td>
<td>-0.00505</td>
</tr>
<tr>
<td></td>
<td>(0.0187)</td>
<td>(0.00725)</td>
<td>(0.0179)</td>
<td>(0.0690)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,732</td>
<td>1,730</td>
<td>1,516</td>
<td>1,639</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Constrained firms</th>
<th>∆(c2a)</th>
<th>∆(net lev)</th>
<th>∆(ol)</th>
<th>∆(mlev)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Union-win</td>
<td>-0.0316</td>
<td>0.00804</td>
<td>0.0330</td>
<td>0.0908</td>
</tr>
<tr>
<td></td>
<td>(0.0233)</td>
<td>(0.0100)</td>
<td>(0.0245)</td>
<td>(0.0923)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,761</td>
<td>1,701</td>
<td>1,480</td>
<td>1,482</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: Full sample</th>
<th>∆(c2a)</th>
<th>∆(net lev)</th>
<th>∆(ol)</th>
<th>∆(mlev)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Union-win</td>
<td>0.00581</td>
<td>-0.00263</td>
<td>0.0241</td>
<td>0.0428</td>
</tr>
<tr>
<td></td>
<td>(0.0156)</td>
<td>(0.00630)</td>
<td>(0.0152)</td>
<td>(0.0561)</td>
</tr>
<tr>
<td>Observations</td>
<td>3,498</td>
<td>3,432</td>
<td>3,000</td>
<td>3,122</td>
</tr>
</tbody>
</table>

*** p<0.01, ** p<0.05, * p<0.1
Table 1.6: Causal effect of unionization on cash-to-assets ($c2a$) and net leverage (net lev) of unconstrained and constrained firms for different measures of financial constraints. The dummy $D$ takes value one if the firm falls in the upper half of the respective measure. OLS regressions on a union-“win” dummy, union-“win” dummy interacted with constraint-dummy, winning margin, and all interactions. Sample restricted to elections within 20% of the threshold. Standard errors clustered at the firm level. KZ-D, ratings R-D, and WI-D dummies are 1 if more likely to be constrained; total assets AT-D and Dividend Div-D dummies are 1 if less likely to be constrained according to the respective measure.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>$\Delta(c2a)$</th>
<th>$\Delta($net lev$)$</th>
<th>$\Delta(c2a)$</th>
<th>$\Delta($net lev$)$</th>
<th>$\Delta(c2a)$</th>
<th>$\Delta($net lev$)$</th>
<th>$\Delta(c2a)$</th>
<th>$\Delta($net lev$)$</th>
<th>$\Delta(c2a)$</th>
<th>$\Delta($net lev$)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>win</td>
<td>0.0434***</td>
<td>-0.0101</td>
<td>0.0811</td>
<td>-0.0493*</td>
<td>0.0494***</td>
<td>-0.0241</td>
<td>0.00609</td>
<td>-0.129*</td>
<td>0.0598*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0207)</td>
<td>(0.00739)</td>
<td>(0.0597)</td>
<td>(0.0253)</td>
<td>(0.0187)</td>
<td>(0.0242)</td>
<td>(0.0104)</td>
<td>(0.0669)</td>
<td>(0.0331)</td>
<td></td>
</tr>
<tr>
<td>KZ-D-win</td>
<td>-0.0677**</td>
<td>0.0140</td>
<td>0.0140</td>
<td></td>
<td>-0.0786</td>
<td>0.0495*</td>
<td>0.0241**</td>
<td></td>
<td>-0.0684**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0305)</td>
<td>(0.0124)</td>
<td>(0.0612)</td>
<td>(0.0262)</td>
<td>(0.0612)</td>
<td>(0.0262)</td>
<td>(0.0125)</td>
<td>(0.0303)</td>
<td>(0.0337)</td>
<td></td>
</tr>
<tr>
<td>R-D-win</td>
<td>-0.0852***</td>
<td>0.0241*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0297)</td>
<td>(0.0125)</td>
<td>(0.0297)</td>
<td>(0.0125)</td>
<td></td>
<td></td>
<td></td>
<td>(0.0303)</td>
<td>(0.0125)</td>
<td></td>
</tr>
<tr>
<td>WI-D-win</td>
<td></td>
<td></td>
<td>-0.0852***</td>
<td>0.0241*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0297)</td>
<td>(0.0125)</td>
<td></td>
<td></td>
<td></td>
<td>(0.0303)</td>
<td>(0.0125)</td>
<td></td>
</tr>
<tr>
<td>AT-D-win</td>
<td></td>
<td></td>
<td>0.0605**</td>
<td>-0.0174</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.147**</td>
<td>-0.0684**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0303)</td>
<td>(0.0125)</td>
<td></td>
<td></td>
<td></td>
<td>(0.0689)</td>
<td>(0.0337)</td>
<td></td>
</tr>
<tr>
<td>Div-D-win</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.147**</td>
<td>-0.0684**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0303)</td>
<td>(0.0125)</td>
<td>(0.0689)</td>
<td>(0.0337)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*** p<0.01, ** p<0.05, * p<0.1
Chapter 2

Anxiety and Overconfidence in the Face of Risk
Abstract

We model an anxious agent as one who is more risk averse for imminent than for distant risk. Such preferences can lead to dynamic inconsistencies with respect to risk trade-offs. We derive implications for financial markets such as a term structure of risk premia, as well as overtrading and price anomalies around announcement dates, which are found empirically. We show that strategies to cope with anxiety can explain costly delegation of investment decisions. Finally, we model how an anxiety-prone agent may endogenously become overconfident and take excessive risks.\footnote{This chapter is coauthored with Thomas M. Eisenbach, Federal Reserve Bank of New York, thomas.eisenbach@ny.frb.org. The views expressed in the paper are those of the authors and are not necessarily reflective of views at the Federal Reserve Bank of New York or the Federal Reserve System. All errors are our own. For helpful comments and discussion we would like to thank Roland Bénabou, Markus Brunnermeier, Sylvain Chassang, Daniel Gottlieb, Edoardo Grillo, Daniel Kahneman, Stephen Morris, Paulo Natenzon, Wolfgang Pesendorfer, Hyun Shin, Marciano Siniscalchi, Satoru Takahashi, Paul Tetlock, and Wei Xiong, participants and our discussant Andrew Hertzberg at the 2nd Miami Behavioral Finance Conference, the 2011 Whitebox Student Conference at Yale, as well as seminar participants at Princeton.}
CHAPTER 2. ANXIETY AND OVERCONFIDENCE

2.1 Introduction

Economists have extensively investigated dynamically inconsistent preferences. The literature has, however, focused on inconsistency of time preferences, while neglecting implications for risk preferences. We study a particular case of dynamically inconsistent risk preferences.

We define an anxiety-prone decision maker as more risk averse for imminent than for distant risk. As the resolution of uncertainty draws close, such an agent wants to pull back from gambles he previously decided to take, although there is no new information, and despite his beliefs not having changed for any other reason.

This basic preference is illustrated in Figure 2.1. It displays the choices between intertemporal consumption streams consisting of sure payoffs of 1 and a coin toss with payoffs \{0, 3\}. When the coin toss comes in the first period and is therefore imminent, the agent prefers the deterministic consumption stream but when the coin toss comes in the second period, the agent prefers the uncertain consumption stream.

Such behavior is the result of dynamically inconsistent preferences with respect to risk trade-offs. This is markedly different from an agent having time-changing risk preferences. For an agent who simply values risks differently at different points
in time, there is no intrapersonal disagreement about risk preferences (and the price of risky assets, for that matter). It is also distinct from a preference for the timing of the resolution of uncertainty, as an anxiety prone decision maker violates the axiom in Kreps and Porteus (1978) that assumes temporal consistency.

For comparison, Figure 2.2 displays a preference for late resolution using the consumption streams analogous to Figure 2.1.

If an anxiety-prone agent trades in a financial market, he will require a higher risk premium for uncertainty resolved in the near future than for uncertainty resolved in the distant future. If there are sufficiently many anxious agents in the population, this implies a down-ward sloping term structure for risk premia. In fact, recent empirical work by van Binsbergen, Brandt, and Koijen (2011) finds just that, with claims to short-term dividends having almost twice the risk adjusted return of claims on long-term dividends.

In addition, anxiety causes an agent to trade excessively around dates of resolution of uncertainty such as earnings announcements. More specifically, he will sell risky securities just before information about these securities’ payoffs is revealed, and buy back his position after the resolution of the risks. Such trading causes a predictable price dip before announcement dates, and price increases in the period the risk gets resolved. The empirical literature has found such an anomaly, and discussed it as the ‘earnings announcement premium’ (Bernard and Thomas, 1989). Lamont and Frazzini (2007) confirm that the selling pressure before the announce-
ment as well as the buy pressure after the event stems from small investors, with large and presumably sophisticated investors taking the other side of the trades. Our theory predicts both of these features.

We also predict investor returns associated with such behavior. Overtrading due to anxiety is costly for two reasons. First, trading costs eat up returns even if trading per se does not lead to losses. Odean (1999) famously documents this. Second, anxious investors sell before announcements when prices tend low, and buy back at higher prices after the resolution of uncertainty, thus losing with each round of trading in expectation, even absent trading costs. The sum of transaction costs and systematic trading losses may explain why retail investors shun equity exposure at prices neoclassical theory would predict. This gives rise to the equity premium puzzle. Our theory thus views (i) overtrading (ii) price anomalies around announcements and (iii) the equity risk premium as stemming from a single behavioral distortion – anxiety.

It is natural to expect sophisticated agents to come up with strategies to cope with anxiety. Such strategies involve the delegation of investment decisions, which is otherwise puzzling in light of sub-par performance of money managers (Gruber, 1996). Paying an agent to carry out future decisions according to present preferences is a simple but effective way to solve the dynamic inconsistency with respect to risks. Our theory also suggest a demand for particular fee schedules featured in investment funds and brokerage accounts. For example, an anxiety prone decision maker will prefer to have to pay for – or better yet be denied – immediate information about fund performance, because such information may prompt his future self to trade out of a position deemed reasonable presently. This is particularly true for information about increased risks, as we will explain in the section on overconfidence. The
timing of investment decisions will be affected as well. Agents will invest in recent winners and pull out funds from recent losers, as Sirri and Tufano (1998) observe. No learning about fund managers’ ability is required to explain this pattern.

As another strategy to cope with anxiety, we present a model of endogenous overconfidence. The desire to confine future behavior to present preferences gives rise to a demand for overconfidence. If exposed to a risky environment, the agent finds it beneficial to have overconfident beliefs in the future, as overconfidence helps counterbalance the anxiety he expects his future self to exhibit. Underestimating the risks, his future self will be more likely to take gambles that are favorable according to the manipulating self’s preferences, but not according to the anxious self’s preferences. We show that the agent can deceive himself to generate such biased beliefs in an intrapersonal strategic communication game between his present and future self, despite the future self being a rational Bayesian updater and being aware of being deceived by its previous self. A comparative statics analysis confirms the intuition that agents more prone to anxiety are more likely to be overconfident, and that they tend to be overconfident to a greater degree. We thus provide a first micro-foundation of a systematic bias of beliefs that has helped explain many puzzles in financial economics that neoclassical theory has left open, such as seemingly excessive amounts of trade.

Moreover, as a result of overconfidence, an anxiety-prone agent may appear to take excessive risks. In our model, overconfidence arises only in high-risk environments. Therefore, we suggest that excessive risk-taking should feature most prominently in inherently risky domains such as securities trading. We conjecture that features of intra-organizational communication patterns can be explained with our theory. Occupational choices and associated cognitive dissonance are other ar-
CHAPTER 2. ANXIETY AND OVERCONFIDENCE

As we see fit. Ben-David, Graham, and Harvey (2010) confirm that financial top executives are systematically overconfident (realized market returns are within their 80% confidence intervals only 33% of the time). Ben-David, Graham, and Harvey (2007) show that this overconfidence translates into riskier corporate policy.

The rest of this paper is organized as follows. In Section 2.1.1 we relate our work to previous research on anxiety, both in psychology and economics. We present experimental evidence to support our assumptions, as well as a short overview on the literature on overconfidence, a prediction of our model. Section 3.2 presents our formal setup. Section 2.3 investigates how an anxiety-prone agent behaves in a stylized financial market. We also discuss implied institutional effects in that section. Section 2.4 presents our model of endogenous overconfidence. We conclude and lay out ideas for future research in Section 2.5.

2.1.1 Related Literature

Anxiety

People become ‘anxious’ as they approach risky situations. To measure ‘anxiety’ (in a popular sense of the word), psychologists have investigated physiological, emotional, and cognitive responses to anxiety provoking situations. All of them involve being exposed to risks, and immediacy of the risk is found to be a leading determinant for physiological and behavioral reactions to the risk. To illustrate, Roth, Breivik, Jørgensen, and Hofmann (1996) continues a whole series of psychological studies on anxiety of parachutists as the moment of the jump approaches, as well as during the fall (Fenz and Epstein, 1967; Fenz and Jones, 1972). Self-reported anxiety, heart rate and other measures peak right before the jump in novices.2

2Fear of flying seems a more commonly experienced situation. Accident statistics rarely change significantly between the time of ticket purchase and the actual
Experienced jumpers learn to inhibit or control their fear, which helps them to perform better in their risky endeavor. Paterson and Neufeld (1987) also find imminence to be a major determinant of the appraisal of a threat in the laboratory. Objectively observable physiological responses besides heart beat and self-reported anxiety include sweating (Monat and Lazarus, 1991).

Lo and Repin (2002) measure the same physical responses of day traders to anxiety provoking situations. In a follow-up paper, Lo, Repin, and Steenbarger (2005) confirm that traders with stronger emotional response generate lower returns. We will argue in this paper that the response to anxiety in the face of risk includes changes of risk preferences, which cause trading losses. Indeed, Loewenstein, Weber, and Hsee (2001) list changes of risk preferences as emotional reactions to the immediacy of risk, despite cognitive evaluations of the risks remaining unchanged.

Economists have used the term anxiety before only in very specific circumstances. Maybe most notably, Epstein and Kopylov (2007) have a model of ‘cold feet’, in which a decision maker becomes more pessimistic as risks approach. Besides the prediction that people may pull back from risks previously decided to take, their axiomatization has little in common with our approach.

**Experimental Evidence**

There is a significant body of experimental evidence documenting agents who are more risk averse if the resolution of uncertainty is temporally close than when it is distant. We want to highlight three studies which are particularly close to the phenomenon we address in this paper.\(^3\)

\(^3\)For other classic studies see Shelley (1994), Keren and Roelofsma (1995), and Sagristano, Trope, and Liberman (2002). Very recent work documenting the effect is
Jones and Johnson (1973) have subjects participate in a simulated medical trial for a new drug where they have to decide on a dose of the drug to be administered. The subjects are told that the probability of experiencing unpleasant side-effects is increasing in the dose administered, as is the monetary compensation. More risk averse subjects should then choose lower doses than less risk averse subjects. In line with the predictions of our theory of anxiety, the study finds that subjects choose higher doses if they are to be administered the next day than when they are to be administered immediately.

In a second, more recent study by Onculer (2000), subjects are asked to state their certainty equivalent for a lottery to be resolved immediately, as well as for the same lottery to be resolved in the future. A lower certainty equivalent corresponds to higher risk aversion. The study finds that subjects state significantly lower certainty equivalents for the immediate lottery than for the future lottery.

The third study is by Noussair and Wu (2006). The study presents subjects with a list of choices between two binary lotteries as in Holt and Laury (2002). The first lottery always has prizes ($10.00, $8.00) while the second lottery always has prizes ($19.25, $0.50). Going down the list, only the respective probabilities of the two prizes change, varying from (0.1, 0.9) to (0.9, 0.1). As probability mass shifts from the second prize to the first prize, the second lottery becomes increasingly attractive compared to the first lottery. Subjects are asked to pick one of two lotteries for each of the probability distributions. The probability distribution at which a subject switches from the “safe” lottery to the “risky” lottery is a proxy for the subject’s risk aversion. One of the chosen lotteries is actually played out, either on the same day or three months later. The study finds that 38.5% of subjects are

\[ \text{in Baucells and Heukamp (2010), Coble and Lusk (2010), and Abdellaoui, Diecidue, and Onculer (2011).} \]
more risk averse for the present than for the future.\footnote{7.7\% are more risk averse for the future than the present and the risk aversion of the remaining subjects does not change.} Note that this study finds a within-subject effect!

In sum, people react differently to risks as a function of the time to resolution of the uncertainty \textbf{without believing the situation to get more risky}.\footnote{We emphasize the distinction between overconfidence, which refers to holding beliefs with excessively high precision, and over-optimism, which refers to over-estimating the mean of a distribution. Neither is implied by the other, as Hvide (2002) clearly illustrates. Over-optimism is also documented in the psychology literature, albeit less prominently than overconfidence (see Langer, 1975; Weinstein, 1980; Moore and Healy, 2008). Applications of overoptimism to economics, such as Van den Steen (2004), are almost exclusively outside of the finance domain.}

\textbf{Overconfidence and its Relation to Forgetting}

The previous subsection provided evidence to support the \textit{assumption} of our model – higher risk aversion if the resolution of uncertainty is more imminent. In this section, we review psychological evidence of one of the model’s \textit{predictions}, namely that anxiety-prone agents exhibit overconfidence.\footnote{Beginning with Adams and Adams (1961), countless studies in cognitive psychology on the calibration of subjective probabilities have reported that people overestimate the precision of their knowledge (Kahneman and Tversky, 1973; Alpert and Raiffa, 1982). Subjects often answer general knowledge questions incorrectly, yet with high reported confidence or even certainty. Indeed, they are so confident that they are willing to bet on their answers’ correctness (Fischhoff, Slovic, and Lichtenstein, 1977). The effect abates, but does not disappear, when subjects are informed about other subjects’ overconfidence in the task at hand. Psychologist as subjects are no exception (Oskamp, 1965). More particularly, overconfidence is greatest for difficult tasks, for forecasts with low predictability, and for undertakings}
lacking fast and clear feedback (Fischhoff, Slovic, and Lichtenstein, 1977; Hoffrage, 2004). Financial markets are a prime example of such an environment.

As for the mechanism how overconfidence is generated, in his essay “On the psychological mechanism of forgetting,” Freud (1901) suggests that anxiety triggering information is prevented from entering memory and gets suppressed (see also Guenther, 1988). An implication is that forgetting probabilities in anxiety triggering environments should be higher than in subjectively safe situations. Zeller (1950) shows that more anxious people are more forgetful as a result of repression. Holmes (1995) gives a review of other experiments validating the memory manipulation implications of anxiety. Deliberate memory manipulation is also implied in Pearlin and Radabaugh (1976), who find that people “who experienced increased anxiety [...], showed stronger tendencies to endorse drinking as a way of controlling distress” (see also Morris and Reilly, 1987).

While overconfidence is a prediction of our model, existing models use overconfidence as an ingredient for finance applications. Agents in those models usually overestimate the precision of signals. Quite naturally, it leads to overreaction to the news associated with the overweighted signal (Daniel, Hirshleifer, and Subrahmanyan, 1998). Other uses of overconfidence are in explaining possibly excessive trade volume (Scheinkman and Xiong, 2003), and pricing of consumer products (Grubb, 2009). We are not aware of prior work that is concerned with overconfidence as a commitment device to take risks.

2.2 Model

Denote a possibly random intertemporal payoff stream from period $t$ to period $T$ by $X_t^T = (x_t, x_{t+1}, \ldots, x_T)$. Our anxiety-prone agent evaluates the consumption
stream $X_t^T$ according to the utility function

$$U_t(X_t) = E_t \left[ v(x_t) + \delta u(x_{t+1}) + \cdots + \delta^{T-t} u(x_T) \right],$$

where $v$ and $u$ are von Neumann-Morgenstern utility indices, $\delta \leq 1$ is a discount factor and $E_t$ is the expectations operator conditional on the information available at the beginning of period $t$.\(^6\)

The only difference between our agent and a standard agent is that uncertainty in the current period is evaluated according to the utility function $v$ while uncertainty in all future periods is evaluated according to the utility function $u$. To capture the effect of anxiety affecting imminent uncertainty, we assume that $v$ is more risk averse than $u$.\(^7\) The key effect of this assumption is that it introduces a time inconsistency in the agent’s preferences which implies that he may choose differently from a given set of alternatives depending on the period of choice. The following example illustrates this point.

**Example** Let $v(x) = \sqrt{x}$ and $u(x) = x$ and let $\delta = 1$. Then the decision maker is risk averse with respect to current uncertainty and risk neutral with respect to

---

\(^6\)Assuming time separable utility inevitably implies marginal rates of inter-temporal substitution. Several recent writings find this a desirable trait, and explore the joint effects of non-exponential discounting and implied non-constant risk aversion. See, for example Fudenberg and Levine (2010), and Halevy (2008). The inter-temporal effect is, however, not the focus of our paper. Therefore, for simplicity of exposition, we choose to relegate a treatment with Epstein-Zin preferences to a technical version of this paper, and choose examples in which the inter-temporal implications do not affect the results.

\(^7\)Our notion of “more risk averse than” is the standard one going back to Pratt (1964).
future uncertainty. Now consider the following two lotteries:

\[ \tilde{x} = \begin{cases} 
4 & \text{with prob. } \alpha \\
0 & \text{with prob. } 1 - \alpha 
\end{cases} \quad \text{and} \quad \tilde{y} = 1 \]

Then \( v \) prefers the risky \( \tilde{x} \) to the safe \( \tilde{y} \) if \( \alpha > \frac{1}{2} \) while \( u \) prefers \( \tilde{x} \) to \( \tilde{y} \) if \( \alpha > \frac{1}{4} \) and there is disagreement between the two utility functions for all \( \alpha \in \left( \frac{1}{4}, \frac{1}{2} \right) \). In particular, suppose that \( \alpha = \frac{1}{3} \) and that the lotteries are resolved and paid out in period \( t \). Then the agent will choose the safe option \( \tilde{y} \) in period \( t \) but would prefer to commit to the risky option \( \tilde{x} \) in all prior periods \( t' < t \). He is willing to pay up to \( \frac{1}{3} \) to commit to the risky option before period \( t \) and is willing to pay up to \( \frac{5}{9} \) to avoid the risky option in period \( t \).

### 2.3 Finance Applications

#### 2.3.1 Term Structure of Risk Premia

Recent empirical work by van Binsbergen, Brandt, and Koijen (2011) finds a downward-sloping term structure of risk premia in the stock market. Based on the S&P 500, the paper prices a claim on the dividends in the near future in contrast to the value of the S&P 500 itself which is a claim on all future dividends.\(^8\) The striking result is that the returns from holding the claim to only the short-term dividends is much higher than the return to holding the claim to all future dividends as displayed in Table 2.1 adapted from Table 1 in van Binsbergen, Brandt, and Koijen (2011).

\(^8\)Since these dividend strips are not actually traded, the prices are derived from the prices of options on the S&P 500 using only a no-arbitrage condition (put call parity).
Table 2.1: Monthly returns of short-term dividend strip and of the S&P 500 itself. Adapted from Table 1 in van Binsbergen, Brandt, and Koijen (2011).

\[ \begin{array}{|c|c|c|} 
\hline
 & ST claim & S&P 500 \\
\hline
\text{Mean} & 1.16\% & 0.56\% \\
\text{Std. dev.} & 7.80\% & 4.69\% \\
\text{Sharpe ratio} & 0.1124 & 0.0586 \\
\hline
\end{array} \]

but the Sharpe ratios show that also the risk adjusted excess return is almost twice as high for the short-term claim. These results strongly suggest that the risk premium for uncertainty resolved in the near future is significantly higher than the risk premium for uncertainty resolved in the distant future. We now show how our model of anxiety can easily account for this effect.

We consider a standard asset pricing setup in discrete time with two periods \( t = 0, 1 \). There are two assets, asset 0 pays a random dividend \( d_0 \) at the end of period 0 while asset 1 pays a random dividend \( d_1 \) at the end of period 1. Each asset is in net supply of 1 and the dividends \( d_t \) are i.i.d. At the beginning of period \( t = 0 \), the agent has to form a portfolio \( (\phi_0, \phi_1, \xi_0, \xi_1) \) of the two assets as well as borrowing/lending for \( t = 0, 1 \), given some initial wealth \( w \) to solve the following problem

\[
\max_{\{\phi_0, \phi_1, \xi_0, \xi_1\}} \mathbb{E}[v(c_0) + \delta u(c_1)]
\]

\[
\text{s.t.} \quad c_t = d_t \phi_t + \xi_t \quad \text{for} \quad t = 0, 1
\]

\[
p_0 \phi_0 + \xi_0 + p_1 \phi_1 + \frac{\xi_1}{1 + r} \leq w
\]
For simplicity we assume that the risk-free rate \( r = 0 \) and that the agent’s discount factor \( \delta = 1 \). Then the first-order conditions for an interior solution are

\[
E [v'(c_0) (d_0 - p_0)] = 0
\]

and

\[
E [u'(c_1) (d_1 - p_1)] = 0.
\]

For an anxiety-prone representative agent we have \( c_0 = d_0 \) and \( c_1 = d_1 \) which gives us the following result on risk premia. (All proofs are relegated to the appendix.)

**Proposition 2.1** If \( v \) is more risk averse than \( u \), the return on the short-term claim is higher than the return on the long-term claim:

\[
\frac{E[d_0]}{p_0} > \frac{E[d_1]}{p_1}
\]

This result shows that the anxiety model can directly account for the downward-sloping term structure of risk premia documented in van Binsbergen, Brandt, and Koijen (2011), in contrast to the leading asset pricing models currently in use.\(^9\)

### 2.3.2 Announcement Effects

We now turn to the effects anxiety has in the context of announcements. We continue to use the standard setup of the previous section with two periods \( t = 0, 1 \). However, we now consider only a single asset with net supply of 1 and a random payoff \( d \) which is realized at the end of period 1. No uncertainty is resolved between period 0 and period 1. The uncertainty about the asset’s payoff is meant

\(^9\)van Binsbergen, Brandt, and Koijen (2011) show that the term structure of risk premia is **upward-sloping** in both the habit formation model of Campbell and Cochrane (1999) as well as the long-run risk model of Bansal and Yaron (2004) which uses the recursive preferences of Epstein and Zin (1989).
to represent a scheduled earnings announcement which provides information about
the stock’s dividend. It can also be interpreted more generally as the resolution of
payoff-relevant information for holders of the stock – the key element is that the
timing of the resolution is fixed and known in advance.

The price of the stock in period $t$ is denoted by $p_t$ and borrowing and lending is
possible at a risk-free rate of zero. At the beginning of each period $t$, the agent has to
form a portfolio $(\phi_t, \xi_t)$ of stock holdings and borrowing/lending, given beginning-
of-period wealth $w_t$.

We solve backwards. In period 1, the uncertainty of the stock’s payoff is immi-


tent so the anxious agent chooses a portfolio $(\phi_1, \xi_1)$ to solve

$$
\max_{(\phi_1, \xi_1)} E [v(c_1)]
\text{ s.t. } c_1 = \phi_1 d + \xi_1 \\
\phi_1 p_1 + \xi_1 \leq w_1
$$

The first-order condition for an interior solution is

$$
E [v' (\phi_1 d + w_1 - \phi_1 p_1) (d - p_1)] = 0. \quad (2.1)
$$

If the agent already makes the portfolio decision in period 0, the non-anxious pref-

erences $u$ apply, and the first-order condition is

$$
E [u' (\phi_0 d + w_0 - \phi_0 p_0) (d - p_0)] = 0. \quad (2.2)
$$
Overtrading

Consider our anxiety-prone agent in an asset market dominated by standard agents with dynamically consistent risk aversion. Since there is no additional information revealed between period 0 and period 1, there is no reason for the price to change between the periods and we have \( p_1 = p_0 =: p \). In addition, assume that the agent’s wealth does not change so we have \( w_1 = w_0 =: w \). Then, the first order conditions (2.1) and (2.2) simplify to

\[
E [v'(\phi_1 (d - p) + w) (d - p)] = 0
\]

and

\[
E [u'(\phi_0 (d - p) + w) (d - p)] = 0.
\]

This gives us the following result adapted from Wang and Werner (1994).

**Proposition 2.2** If \( v \) is more risk averse than \( u \), we have \( \phi_0 > \phi_1 \).

This result shows that our agent wants to hold more of the risky asset in period 0, with some distance to the risk, than in period 1, when the resolution of uncertainty is imminent. The implications of this result depend on the degree of sophistication of the agent. A sophisticated agent anticipates in period 0 that he will want to change his portfolio in period 1. If the agent has no way of preventing his future self from rebalancing, he may already choose the anticipated portfolio \( \phi_1 \) in period 0 to avoid trading costs.

The more interesting case is that of a naive agent. In period 0, he will choose a portfolio \( \phi_0 \) but once the resolution of uncertainty is imminent in period 1, he sells some of the risky asset to attain the portfolio \( \phi_1 < \phi_0 \). When we view the asset market as a sequence of periods with and without news about the asset, the agent overtrades, selling some of the stock before announcements and buying it back.
afterwards. Lamont and Frazzini (2007) find evidence that selling pressure before announcements indeed stems from small and supposedly unsophisticated traders, as does the buy pressure after announcements. Large and supposedly sophisticated traders take the other side of these trades.

Notably, in the presence of transaction costs, an anxious investor will earn lower returns than a buy-and-hold investor due to overtrading, as in Odean (1999). We examine other factors affecting individual investors’ returns in the following sections.

**Price Dip**

To derive pricing implications, we now model an economy with an anxiety-prone representative agent. This implies that he has to hold the entire net supply of the stock, $\phi_t = 1$, consumes the entire payoff, $x_1 = d$, and cannot borrow or lend, $\xi_t = 0$. Substituting these values into the first order conditions (2.1) and (2.2), they simplify to

$$E[v'(d)(d - p_1)] = 0 \quad (2.3)$$

$$E[u'(d)(d - p_0)] = 0 \quad (2.4)$$

and we have the following result.

**Proposition 2.3** If $v$ is more risk averse than $u$, we have $p_0 > p_1$.

This result shows that the price at which the agent is willing to hold the stock is lower when the resolution of uncertainty is imminent than when it is still distant. If the agent is naive about his anxiety, he will be happy to hold the stock at a price of $p_0$ in period 0, irrationally expecting the price not to change in period 1. Once
the earnings announcement is imminent, the agent becomes anxious and the price drops to \( p_1 \). Note that the price jumps after the announcement (albeit not as much) also in a model with a standard risk averse agent. However, the price dip before the announcement is uniquely produced by anxiety.

Rewriting the expectations in conditions (2.3) and (2.4) allows us to write the prices explicitly:

\[
p_0 = E[d] + \frac{Cov(u'(d), d)}{E[u'(d)]} \quad \text{and} \quad p_1 = E[d] + \frac{Cov(v'(d), d)}{E[v'(d)]}
\]

The second term in the two price equations is the risk premium. It discounts expected dividends more strongly at \( t = 1 \) than at \( t = 0 \), as shown in Proposition 2.3. In particular, the covariances are negative and the expectations positive, as both \( u' \) and \( v' \) are positive but decreasing. As one should expect, increasing but risk averse utility functions imply a price discount of the risky asset, relative to expected value. More risk aversion makes for heavier discounting, and vice versa.

In the case of risk neutrality, \( u'(d) = c \), the covariance term is zero as \( u''(d) = 0 \). Then, the asset trades at expected dividends.\(^{10}\)

In a market populated by both anxious and standard agents, there will be a price drop before any scheduled announcement but not as large as in a market with only anxious agents. Accompanying the price drop we should expect to see anxious agents selling part of their stocks to standard agents. Right after the announcement, prices should on average appreciate as anxiety-prone agents buy back their positions.

Our theory thus combines predictions about both asset price movements and trade volume around announcement dates, which is a crucial feature of announcement anomalies, as Lamont and Frazzini (2007) explain. These authors also confirm

\(^{10}\)The same pricing equations result if the representative agent maximizes \( u(x_0) + v(x_1) \) in period \( t = 0 \) and consumes out of wealth.
that institutional investors lean against the individual investors’ trades. A strategy of buying before announcement dates and selling thereafter yields excess returns of 7% to 18%, which they call the announcement premium. While their paper focuses on explaining the price and volume patterns with the ‘attention grabbing hypothesis’ (see also Lee, 1992; Hou, Peng, and Xiong, 2009; Barber and Odean, 2008), their empirical results provide equal support for our theory. Our theory shifts the focus to the other side of the same medal that Lamont and Frazzini (2007) examine: we ask why prices tend relatively lower before the announcement, which is depicted by Bernard and Thomas (1989). We call this the ‘pre-earnings announcement dip’. We thereby offer a possible “common underlying cause for both volume and the premium” that Lamont and Frazzini (2007) have called for, as an alternative to the ‘attention-grabbing hypothesis’.

**Realized Returns**

The stylized model above is not yet suited to be calibrated with data. However, the analysis in Bernard and Thomas (1989) suggests a pre-earnings announcement dip on the order of $-0.5\%$ (smaller for large firms than for small-caps). With four scheduled earnings announcements per year, a naive agent as depicted above stands to lose about 2% per year by overtrading in the face of scheduled quarterly earnings announcements alone. This loss comes on top of the transaction costs of overtrading. This squares nicely with the empirical result by Lo, Repin, and Steenbarger (2005), who confirm that more anxious agents generate lower returns. Our model predicts a similar price effect for scheduled news events relevant to the equity market as a whole, such as the publication of unemployment figures.
A naive anxiety-prone investor’s actual equity returns, i.e. the returns he enjoys from investing in equity after accounting for the losses imposed by anxious behavior, are lower than buy-and-hold returns derived from market data may suggest. This helps explain the equity premium puzzle. The Equity Premium Puzzle (EPP) states that equity returns are too high relative to bond returns than can be explained by reasonable levels of risk aversion and discount rates. If risk aversion were as high as implied by the difference between equity and bond returns, bond returns would have to be much higher than they actually are. The latter part of the problem is known as the “risk-free rate puzzle.” Hence, models attempting to explain the EPP with agents, who are, effectively, very risk averse, can only explain the difference in returns between bonds and equity, but fail to explain the ensuing risk-free rate puzzle. For example, models assuming ambiguity aversion typically run into that problem. In contrast, anxiety explains part of the EPP without running into the risk-free rate puzzle by showing that effective equity returns to an anxiety-prone investor are not as high as they appear in the data, while bond returns are unaffected by anxiety. Dynamic inconsistency with respect to risks only distorts the price of the locally risky equity, but not the price of locally risk-free bonds.

Our predictions stem from the analysis of a naive anxiety-prone agent. As we will discuss in the next section, a sophisticated anxiety-prone agent may find ways to behave in a dynamically consistent way and thus suffer to a lesser extent from the costs of overtrading. Yet, the disutility implied by the use of a commitment device needs to be subtracted from the utility from equity returns of such an agent. For example, the following section shows how overconfidence can let an anxiety-prone decision maker make more dynamically consistent decisions. But then, the disutility from overconfidence, stemming from ‘excessive risk-taking’, needs to be subtracted
from the now higher utility from holding equity without overtrading. Consequently, even a sophisticated anxiety-prone agent will find equity a worthwhile investment only at returns that are higher than the ones a standard consumption-based asset pricing model yields.

Note that most firms' equity prices may also be depressed, since institutional counterparties may find it more profitable to use their capital to exploit the behavioral distortions of retail investors trading in stocks that have immanent earnings announcements, instead of pushing up equity prices across the board. Moreover, anxiety-prone agents' counterparties may anticipate the selling pressure by anxious agents before earnings announcements. If (they know that) they can not absorb the sales at the same price level, they will demand higher premia already ahead of the announcement date.

In sum, our theory of anxiety in the face of risk thus links the equity premium, price reactions to earnings announcements, and overtrading, and square nicely with the results of Lo, Repin, and Steenbarger (2005) on the relation of anxiety and trading performance.

2.3.3 Institutional Effects

An agent who plans according to preferences $u$, but is afraid his future self will disagree with these plans (because of having preferences $v$), may try to find ways to commit his future actions to his presently chosen plan of action. While Schelling (1984) and others have discussed the ethical aspects such a possibility brings about, the present discussion is only concerned with that, and how, the agent can restrict his future self’s behavior – simply by virtue of having a first-mover advantage. Indeed, dynamic inconsistency with respect to risks gives a strong economic rationale
for doing so. As sketched out above, an anxiety-prone agent faces losses that are not compensated by higher consumption at any time (as is the case for a hyperbolic discounter).

**Delegation**

Hiring an agent to carry out risk-taking decisions in the future according to the current self’s preferences is one way to prevent future selves’ preferences from conflicting with the current self’s plans. In an investment setting, it may be the case that the anxious self is too risk averse to invest in equity, although the agent realizes this has long-run benefits. In this situation it makes sense for the agent to delegate investment decisions to a portfolio manager. The manager can still react to news about particular assets, but has to stick to a predetermined split of asset classes.

As is the case for commitment devices for hyperbolic discounters, it is clear that having them is desirable, but it is less clear when an agent would start using them. The delegation of investment decisions provides a nice exemption to that rule. An agent prone to anxiety differs from a standard agent only in his evaluation of immediate risks. Thus, we expect to see greater inflows to money managers when immediate risks seem to be low, relative to the associated returns, even if such a temporary calm does not carry information about future performance. This may help to reinterpret respective evidence from the mutual fund industry. As falling prices increase risk estimates, low returns should be associated with low inflows to money managers. Indeed, Sirri and Tufano (1998) find that high returns trigger fund inflows, and vice versa.
Of course, effort costs of managing one’s portfolio may also lead to delegation of investment management. However, effort costs can not justify hiring an agent that underperforms the index on average, as buying index funds is virtually costless and free of effort. Yet, the mutual funds industry is huge, and actual fund managers still tend to underperform the market Gruber (1996). While buying the index is free of effort, it is not free of anxiety. Self 0 may thus correctly anticipate that the anxious self 1 will underperform the market even more than a random portfolio manager by failing to invest in equity at all. Self 0 will therefore be willing to pay an investment manager, even if he expects him to underperform the market. The obvious solution would be to hire an agent to simply buy the index, but that may be infeasible in a model of career concerns.

**Fees**

A redemption fee is another feature of investment funds that sophisticated anxiety-prone decision-makers will demand. This may be one explanation why management and other fees are being competed away in the mutual funds industry, while lock-in fees continue to feature prominently. Variations of punishments for pulling out of risks an investor previously decided to take include fees for changing the equity/bonds ratio of one’s investment in mutual funds, as well as fees imposed if the total exposure to a certain asset class falls below a threshold.

**Timing of Orders**

The widespread practice of retail investors to submit overnight limit orders can be viewed as another costly way of coping with anxiety. Submitting overnight limit orders deprives the investor from the possibility to react to news in the time be-
between submission of the order and execution, and furthermore represents a positive externality to other market participants: it represents an option to buy/sell at the quoted price. See Harris (2003) for a discussion. Writing such an option to trade, as well as foregoing the option to react to overnight news, would never be optimal for a standard agent. However, it helps overcome commitment problems imposed by anxiety. Instead of waiting to see his future self pull out from the decision to invest in the stock, the current self preempts the decision before going to bed, when the uncertainty is not yet imminent.

**Demand for Delayed Resolution and Costs of Information**

Self 1’s risk preferences about future gambles are identical to self 0’s preferences about the same gambles if there is no immediate resolution of uncertainty at \( t = 1 \). This implies a disutility for resolution of uncertainty, i.e. a disutility for information, in period \( t = 1 \). Self 0 will therefore be willing to pay for delaying the resolution of uncertainty from \( t = 1 \) to a later date in order to harmonize self 1’s behavior with self 0’s preferences. To be sure, this is not driven by a preference for the timing of the resolution of uncertainty, which requires temporal consistency Kreps and Porteus (1978). Hedge funds impose pull-out restrictions and publish performance reports at low frequencies, although the information is available continuously and creating a report is a largely automatable task. Note that the cost of having to provide liquidity does not explain such clauses. Imposing costs on deposits with short maturities will compensate the fund for the cost of liquidity provision, but putting a temporal distance between the investor’s decision to pull out and the payout of the funds does neither protect the fund from withdrawals nor compensate
for the implied costs. Concealing present risks, however, prevents anxious investors from pulling out.

2.4 Overconfidence

If commitment devices are not available, an anxiety-prone agent has an incentive to distort his future self’s beliefs. In particular, the present self would like to convince his future self that risks are lower than they actually are. This would lead the future self to take riskier decisions which are more in line with the current self’s preferences. However, if the future self has access to additional information, the distorted beliefs may lead to decisions that are excessively risky, even from the current self’s point of view. In this section we analyze such a situation similar to the model of Bénabou and Tirole (2002).\(^{11}\)

For the sake of simplicity, we again restrict ourselves to two time periods, \(t = 0, 1\), and set the discount factor to \(\delta = 1\). In period 1 the agent has to choose between a risky or a safe alternative. The risky alternative is given by a lottery with random payoff \(x\). The lottery is characterized by its distribution function \(G_\theta\) where \(\theta \in \{H, L\}\) denotes a state of the world that determines how risky the lottery is. We assume that \(G_H\) is a mean-preserving spread of \(G_L\) so the risky alternative is unambiguously riskier in state \(H\) than in state \(L\). The prior probability of the high-risk state \(H\) is given by \(\pi\). The safe alternative, on the other hand, is given by a constant payoff \(a\).

\(^{11}\)For earlier work studying belief manipulation in a setting with \(\beta-\delta\) time inconsistency see Carrillo and Mariotti (2000).
CHAPTER 2. ANXIETY AND OVERCONFIDENCE

The anxious agent in period 1 wants to take the risky alternative whenever

\[ E_\theta[v(x)] > v(a), \]

where \( E_\theta \) denotes the expectation with respect to \( G_\theta \). Denoting the certainty equivalent of \( G_\theta \) given the utility function \( v \) by \( c_\theta^v \), this condition can be rewritten as

\[ c_\theta^v > a. \]

The agent wants to take the risky alternative, whenever its certainty equivalent \( c_\theta^v \) is greater than the safe alternative \( a \).

The agent in period 0, when the risk is not imminent, wants to take the risky alternative whenever

\[ E_\theta[u(x)] > u(a) \iff c_\theta^u > a. \]

Since \( v \) is more risk averse, we have \( c_\theta^u > c_\theta^v \) for both \( \theta \in \{H, L\} \) so the agent in period 0 (self 0) and the agent in period 1 (self 1) will disagree about the course of action if \( a \in [c_\theta^v, c_\theta^u] \).

To make this problem interesting, we assume that the payoff of the safe alternative \( a \) is not known to the agent until period 1. Self 0 only knows the prior distribution \( F \) on \([a, \bar{a}]\) but self 1 observes the realized value of \( a \). The state of the world \( \theta \), on the other hand, is revealed to the agent at the beginning of period 0 in form of a perfectly informative “red flag” warning signal \( s \) if the state is high-risk

\[ s = \begin{cases} R & \text{if } \theta = H \\ \emptyset & \text{if } \theta = L \end{cases} \]
If he receives a red flag, self 0 can choose the probability \( \lambda \in [0, 1] \) with which he will remember the signal, i.e.,

\[
\lambda = \Pr [\hat{s} = R | s = R],
\]

where \( \hat{s} \) is self 1’s recollection of the signal. We assume that self 1 is fully aware of his prior incentive to forget warning signals, so if he expects a memory probability \( \lambda^e \) and doesn’t remember seeing a red flag he uses a Bayesian posterior

\[
\pi(\lambda^e) = \frac{\pi(1 - \lambda^e)}{\pi(1 - \lambda^e) + 1 - \pi}.
\]

Given this setup, self 0 and self 1 are playing a kind of Stackelberg game. First self 0 chooses the memory probability \( \lambda \) taking into account self 1’s behavior and then self 1 decides between the risky and the safe alternative taking into account self 0’s behavior. We are interested in the perfect Bayesian equilibria of this intrapersonal game.

First, we derive self 1’s best response in \( t = 1 \), taking as given an expected memory probability \( \lambda^e \). If self 1 remembers seeing a red flag, \( \hat{s} = R \), he knows that the state of the world is high-risk and chooses the risky alternative if \( c_v^H > a \). If self 1 doesn’t remember seeing a red flag, \( \hat{s} = \emptyset \), he uses the Bayesian posterior \( \pi(\lambda^e) \) and chooses the risky alternative if \( c_v(\lambda^e) > a \) where \( c_v(\lambda^e) \) is the certainty equivalent of the risky alternative given \( \lambda^e \) defined by

\[
E[v(x) | \pi(\lambda^e)] = v(c_v(\lambda^e)).
\]
Second, we derive self 0’s best response in $t = 0$, taking as given self 1’s behavior for an expected $\lambda^e$. If self 0 receives a warning signal and chooses a memory probability $\lambda$, his expected utility is

$$
\lambda \left[ \int_{c_H^H}^{c_H^v} E_H[u(x)] dF(a) + \int_{c_H^v}^{c_H^e} u(a) dF(a) \right] + (1 - \lambda) \left[ \int_{a}^{c_v(\lambda^e)} E_H[u(x)] dF(a) + \int_{c_v(\lambda^e)}^{c_H^e} u(a) dF(a) \right].
$$

With probability $\lambda$ the agent remembers the warning signal in period 1 and uses the certainty equivalent $c_H^v$ as the threshold, choosing the risky alternative for payoffs of the safe alternative below the threshold and choosing the safe alternative for payoffs above the threshold. With probability $1 - \lambda$ the agent forgets the warning signal and uses the certainty equivalent $c_v(\lambda^e)$ as the threshold.

We denote the derivative of self 0’s expected utility with respect to $\lambda$ by

$$
D(\lambda^e|v) := \int_{c_H^v}^{c_v(\lambda^e)} \left( u(a) - E_H[u(x)] \right) dF(a).
$$

This expression has a very natural interpretation. The warning signal changes self 1’s decision only for values of $a \in [c_H^v, c_v(\lambda^e)]$. In this interval, self 1 chooses the risky alternative whenever he remembers seeing a red flag and the safe alternative otherwise. The effect on self 0’s expected utility of remembering the warning signal more often is exactly the difference in utility from the safe action compared to the risky action for the values of $a$ where the decision is affected.

There are three possibilities for perfect Bayesian equilibria in this setting:
• Honesty Equilibrium: If $D(1|v) \geq 0$, there is an equilibrium with $\lambda^* = 1$. In this equilibrium the agent never ignores red flags and doesn’t influence his future self’s beliefs.

• Overconfidence Equilibrium: If $D(0|v) \leq 0$, there is an equilibrium with $\lambda^* = 0$. In this equilibrium the agent always ignores red flags and makes his future self maximally overconfident.

• Mixed Equilibrium: If $D(\bar{\lambda}|v) = 0$ for some $\bar{\lambda} \in (0,1)$, there is an equilibrium with $\lambda^* = \bar{\lambda}$. In this equilibrium the agent plays a mixed strategy, ignoring the red flag with probability $1 - \bar{\lambda}$, and makes his future self partially overconfident.

Proposition 2.4 One of the extreme equilibria always exists, either the honesty equilibrium or the overconfidence equilibrium or both. If both extreme equilibria exist, a mixed equilibrium also exists.

The existence of each kind of equilibrium depends on the degree of anxiety of the agent, i.e., how big the difference in risk aversion is for risks that are imminent compared to risks that are distant. In particular, we can say that an agent $i$ is more prone to anxiety than an agent $j$, if $u_i$ and $u_j$ are equally risk averse but $v_i$ is more risk averse than $v_j$. This enables us to state the following result.

Proposition 2.5 For an agent that is more prone to anxiety, (i) the honesty equilibrium is less likely to exist, (ii) the overconfidence equilibrium is more likely to exist, and (iii) if the mixed equilibrium exists, then it is associated with more overconfidence.

Somewhat counterintuitively, people who are most prone to anxiety in the face of risk are the same ones that are most likely to exhibit overconfidence. Note further
that a risky environment is necessary for overconfidence to arise, and to show effects in decision making. Financial markets are a prime example of such an environment. Ben-David, Graham, and Harvey (2010) confirm that financial top executives are systematically overconfident: realized market returns are within their 80% confidence intervals only 33% of the time. A manifestation of overconfidence that is important in finance, and possibly important to understand individual agents’ behavior during the recent financial crisis, is excessive risk-taking.

**Excessive Risk-Taking**

Equilibria with partial or maximal overconfidence can display excessive risk taking. In these equilibria it can be the case that the future self ends up taking risks which even the less risk averse current self would have avoided. To an observer who is unaware of the agent’s intrapersonal conflict, the agent seems to take risks that are greater than can be explained with ‘reasonable’ preferences, i.e. $u$. This can happen if the true state of riskiness is high and the agent forgets the warning signal. In this case, whenever the payoff of the safe alternative is below the cutoff $c_v(\lambda^*)$ self 1 uses but above the cutoff $c^H_u$ self 0 would like him to use, i.e. $a \in (c^H_u, c_v(\lambda^*))$, the agent takes risks in period 1 that that self 0 considers excessive. Analytically, this can arise since the condition for an equilibrium with overconfidence, $D(\lambda^*|v) \geq 0$, does not necessarily imply $E_H[u(x)] > u(a)$ for all $a < c_v(\lambda^*)$, where self 1 chooses the risky alternative. Such a situation arises in all equilibria $\lambda^*$ with $c^H_u < c_v(\lambda^*)$, i.e. the equilibrium cutoff used by self 1 is greater than the cutoff self 0 would use. To an outside observer who knows that the state is $H$, the anxious agent using the cutoff $c_v(\lambda^*)$ appears as if he were less risk averse than the non-anxious preference $u$. 
**Proposition 2.6** In an equilibrium with $\lambda^* < 1$ and $c_u^H < c_v(\lambda^*)$, the agent will be observed to take excessive risks, i.e. he will appear less risk averse than $u$.

Ben-David, Graham, and Harvey (2007) confirm empirically that overconfidence, observed in Ben-David, Graham, and Harvey (2010), translates into riskier corporate policy.

### 2.5 Conclusion and Plans for Future Research

In this paper, we define an anxiety-prone decision maker as an agent, whose risk aversion is higher the closer in time the resolution of uncertainty is. We discuss experimental evidence that is predicted by our model and show in examples and in a financial market model how this leads to dynamically inconsistent behavior. Linking such behavior to established puzzles about price and volume around earnings announcements, we suggest a clean, and arguably more credible way to think about these patterns than existing theories propose. Evidence from the trading floor also confirms our prediction that more anxiety-prone traders perform worse. We explain how sophistication about dynamic inconsistency and the associated costs will trigger institutional responses such as delegation of investment decisions, and the distinct design of brokerage and investment fund fees. We further suggest a connection to optimal patterns of information provision in financial markets. Finally, we show why it may be beneficial to a sophisticated anxiety-prone agent to hold overconfident beliefs, and how this can be accomplished.

Combining the above model of endogenous overconfidence with problems in financial economics seems a fruitful field of future research. We conjecture four possible areas of applications.
CHAPTER 2. ANXIETY AND OVERCONFIDENCE

First, there should be an equilibrium level of overconfidence in financial markets. The costs of overtrading due to anxiety around news announcements can be mitigated by overconfidence. On the other hand, overconfidence may cause overtrading independent of news announcements according to Scheinkman and Xiong (2003). Such trading, while not directly causing expected losses, still bears transaction costs. But in addition, an overconfident agent also suffers from excessive risk taking, implying a disutility for the planning-self at $t = 0$. Trading off these costs should yield an optimal amount of overconfidence according to self 0’s preferences. The equilibrium level of overconfidence should be increasing in transactions costs and bid-ask spreads, and thus be more pronounced in more illiquid securities. It should be negatively related to the earnings announcement premium, i.e. the predictable price fluctuations between announcement periods and periods without earnings announcements, and positively to the frequency of scheduled announcements.

Second, recent influential works by Akerlof and Shiller (2010) and Reinhart and Rogoff (2009) have strongly suggested that time-changing confidence needs to be part of realistic models of market dynamics and the business cycle. Empirically, confidence is high when leverage is high and maturities are short, and vice versa. This is consistent with our notion that overconfidence arises when risks are high, and (not shown in the above model) under-confidence may arise when risks are low. As overconfident traders have a greater demand for risk than rational types do, overconfidence sustains excessive risk levels. Conversely, under-confidence helps sustain price levels below fundamentals in the crisis. Both outcomes may be possible under the same parameters in a model with multiple equilibria. Extending this static argument to a dynamic model will be more challenging.
Third, the above model of self-delusion is not necessarily to be taken literally, but can be seen as a metaphor for the choice of information systems and communication structures in organizations. Given a preference for a biased posterior, an anxiety-prone leader will implement information and communication systems that have him misinformed about risks. The scarcity of critical upward feedback, which is often said to be mandated by the head of the organization (“killing the messenger”), may be explained in this way. The more anxiety-prone the leader, the less upward feedback will be provided.\(^{12}\) In the investment domain, the “Ostrich Effect” may serve as an example. Karlsson, Loewenstein, and Seppi (2009) find that investors look up their portfolio performance less often after receiving a signal about increased risks.\(^{13}\)

Fourth, occupational choices and associated cognitive dissonance may be a fruitful domain for applications of the overconfidence model. Nothing in the model prevents that the agent, rather than nature, choose the riskiness of the environment (and the thus implied perfectly informative signal). Parallel to the mechanism in the present model, the agent may choose to forget the information he based his prior decision upon, i.e. that he chose a risky job over a safe one, and thus render himself overconfident (Akerlof and Dickens, 1982). This will be beneficial if the agent’s job involves risk-taking. Professions such as securities trading should then be particularly likely to feature overconfident agents.

\(^{12}\)Management publications view the lack of upward feedback as the source of countless corporate disasters and a widespread phenomenon. There are also examples in history, where leaders that were certainly not known for pronounced propensity to anxiety, demanded critique by any means. Queen Elizabeth I is said to have rebuked a jester “for being insufficiently severe with her.”

\(^{13}\)The original finding is that investors tend to not look up their portfolio’s performance after market-wide declines, about which they are likely to become informed via generic news reports. Note that (i) price drops may be caused by increases in risk levels, but also (ii) falling prices increase volatility estimates. Thus, in any case, falling prices are a signal for increased risk.
Appendix

Proof of Proposition 2.1. Since $v$ is more risk averse than $u$ we have

$$-\frac{v''(x)}{v'(x)} > -\frac{u''(x)}{u'(x)}$$

$$\Rightarrow -\frac{d}{dx} \log v'(x) > -\frac{d}{dx} \log u'(x)$$

Integrating both sides yields

$$\frac{v'(d)}{v'(p)} < \frac{u'(d)}{u'(p)}$$

for $d > p$ and the reverse inequality for $d < p$. For general $p, d$ we then have

$$\left( \frac{u'(d)}{u'(p)} - \frac{v'(d)}{v'(p)} \right) (d - p) > 0$$

Taking expectations we get

$$\frac{E [u' (d) (d - p)]}{u' (p)} > \frac{E [v' (d) (d - p)]}{v' (p)}$$  \hspace{1cm} (2.5)

Substituting in $p_0$ the RHS is zero and we get

$$E [u' (d) (d - p_0)] > 0,$$

which implies that $p_0 < p_1$ and therefore

$$\frac{E [d]}{p_0} > \frac{E [d]}{p_1}.$$

\hfill □

Proof of Proposition 2.2. See Wang and Werner (1994).  \hfill □
Proof of Proposition 2.3. The proof is analogous to that of Proposition 2.1. Substituting $p_1$ in equation (2.5) the RHS is zero and we get

$$E[u'(d)(d - p_1)] > 0,$$

which implies that $p_0 > p_1$. \hfill $\square$

Proof of Proposition 2.4. The belief $\pi(\lambda^e)$ is continuous and decreasing in $\lambda^e$. Therefore the certainty equivalent $c_v(\lambda^e)$ is continuous and increasing in $\lambda^e$. Finally, this implies that $D(\lambda^e|v)$ is continuous and increasing in $\lambda^e$. We then have either $D(1|v) \geq 0$ or $D(0|v) \leq 0$ or both so one of the extreme equilibria $\lambda^* \in \{0, 1\}$ always exists. For the case where $D(1|v) \geq 0$ and $D(0|v) \leq 0$, there exists a $\bar{\lambda} \in (0, 1)$ such that $D(\bar{\lambda}|v) = 0$ so the mixed equilibrium $\lambda^* = \bar{\lambda}$ also exists. \hfill $\square$

Lemma 2.1 Consider two von Neumann-Morgenstern utility functions $v_1$ and $v_2$. If $v_2$ is more risk averse than $v_1$, then $D(\lambda^e|v_2) < D(\lambda^e|v_1)$ for all $\lambda^e$.

Proof. If $v_2$ is more risk averse than $v_1$, then $c_{v_2}^H < c_{v_1}^H$ and $c_{v_2}(\lambda^e) < c_{v_1}(\lambda^e)$ for all $\lambda^e$. This implies that for all $\lambda^e$:

$$D(\lambda^e|v_2) = - \int_{c_{v_2}^H}^{c_{v_2}(\lambda^e)} (u(a) - E_H[u(x)])dF(a)$$

$$< - \int_{c_{v_1}^H}^{c_{v_1}(\lambda^e)} (u(a) - E_H[u(x)])dF(a)$$

$$= D(\lambda^e|v_1)$$

$\square$
Lemma 2.2 Consider two von Neumann-Morgenstern utility functions $v_1$ and $v_2$. If $v_2$ is more risk averse than $v_1$ and if there are $\bar{\lambda}_1$ and $\bar{\lambda}_2$ such that $D(\bar{\lambda}_1|v_1) = 0$ and $D(\bar{\lambda}_2|v_2) = 0$, then $\bar{\lambda}_1 < \bar{\lambda}_2$.

Proof. If $v_2$ is more risk averse than $v_1$, then $c_{v_2}^H < c_{v_1}^H$ so the integral in $D(\bar{\lambda}_2|v_2)$ has a smaller lower bound. Since $(u(a) - E_H[u(x)])$ is a strictly increasing function of $a$, for $D(\bar{\lambda}_1|v_1) = D(\bar{\lambda}_2|v_2) = 0$ it is necessary that $c_{v_2}(\bar{\lambda}_2) > c_{v_1}(\bar{\lambda}_1)$, i.e. that the integral in $D(\bar{\lambda}_2|v_2)$ must have a greater upper bound. Since $c_{v_2}(\lambda) < c_{v_1}(\lambda)$ for a given $\lambda$, and $c_v(\lambda)$ is increasing in $\lambda$ for $v_1$ and $v_2$, this implies $\bar{\lambda}_2 > \bar{\lambda}_1$. □

Proof of Proposition 2.5. From Lemma 2.1 we know that $D(1|v_2) < D(1|v_1)$ for $v_2$ more risk averse than $v_1$. Therefore an honesty equilibrium exists for $v_1$ if it exists for $v_2$. Again using Lemma 2.1 we know that $D(0|v_2) < D(0|v_1)$ for $v_2$ more risk averse than $v_1$. Therefore an overconfidence equilibrium exists for $v_2$ if it exists for $v_1$. Finally, if a mixed equilibrium exists for $v_1$ and $v_2$, characterized by $\bar{\lambda}_1$ and $\bar{\lambda}_2$ respectively, then by Lemma 2.2 we have $\bar{\lambda}_1 < \bar{\lambda}_2$. □

Proof of Proposition 2.6. Follows directly from the derivation in the main text. □
Chapter 3

Revealing Downturns
Abstract

Parameter learning can explain asymmetries between upturns and downturns both in asset pricing and in corporate finance. Good performance in good times can be due to either desirable good idiosyncratic performance, or due to undesirable positive correlation with a market-wide factor. In contrast, good performance in bad times can come either from desirable good idiosyncratic performance, or due to desirable negative correlation with a market-wide factor. Therefore, good performance in good times is a less strongly positive signal about firm value than good performance in bad times. The same holds for bad performance: the signal is weaker in upturns than in downturns. As a result, in downturns, (ia) stocks react more sensitively to cash flow news, (ib) the cross-sectional dispersion of returns spikes, (ii) stock picking is more profitable, and (iii) boards fire CEOs with higher frequency.¹

¹This paper is coauthored with Sergey Zhuk. For helpful comments and discussions we would like to thank Markus Brunnermeier, David Hirshleifer, Burt Malkiel, Stephen Morris, Monika Piazzesi, David Sraer, Fabio Trojani, Luis Viceira, Wei Xiong, as well as seminar participants at Princeton. Schmalz gratefully acknowledges support through a Fellowship of Woodrow Wilson Scholars.
CHAPTER 3. REVEALING DOWNTURNS

3.1 Introduction

Asymmetries between upturns and downturns puzzle researchers both in asset pricing and corporate finance. For example, there is evidence that fund managers are more effective at picking stocks in downturns than in upturns. In corporate finance, boards tend to fire CEOs more often in downturns than in upturns, which contradicts relative performance evaluation.\(^2\)

The asymmetries in asset pricing have been addressed with rather intricate partial equilibrium models of effort choice and attention allocation over the business cycle (Glode (2011), Kacperczyk, Nieuwerburgh, and Veldkamp (2011)), and the combination of negatively skewed consumption and the convexity of the flow-performance relationship of funds (Kaniel and Kondor (2011)). A possible reason for higher forced CEO turnover in downturns is that boards, for some reason, fail to fulfill their monitoring duties in upturns. Jenter and Kanaan (2006) discuss and provide support for an alternative hypothesis: performance in recessions may be more informative about the quality of the CEO, or the CEO-firm match.\(^3\)

This paper proposes a simple theory according to which news in downturns indeed carry more relevant information about the utility that risk-averse investors derive from holding the asset than news in upturns. It thereby reduces the number and strength of the “behavioral” assumptions needed to explain the data. Our model proposes that risk-averse investors like good idiosyncratic performance (“cash-flow alpha” – henceforth, \(a\)), but dislike high correlation with a


\(^3\)Other papers addressing asymmetries between upturns and downturns with learning models include Veldkamp (2005) and Van Nieuwerburgh and Veldkamp (2006), who are interested in the speed of booms and busts. Their agents learn about real economic activity, which fluctuates over the business cycle.
market-wide factor (“cash-flow beta” – henceforth, $b$). They are uncertain about both parameters. High cash-flows in good times are both a signal for high $a$ and high $b$. One is a good signal, one is a bad signal for the value of holding the asset to the investor. Exceptionally high cash flows in good times are therefore a somewhat ambiguous signal about firm value. Investors may sense that the reported performance is likely to have been achieved with exceptionally high risk exposure. Therefore, investors will not attach high confidence to good news in good times. Similarly, low cash flows in an upturn can be due to low $a$, or due to a negative $b$. This, again, is a somewhat mixed signal: the sub-par performance might be due to negative market exposure, which will turn out to be a valuable hedge for bad times. Therefore, prices will not adjust strongly in response to bad news in good times either. The model thus predicts that the reaction of market prices to news and the associated volatility is less strong in upturns.

In contrast, relatively good performance in bad times can be either due to high cash-flow alpha or due to low cash-flow beta, both of which are a positive signal about firm value. Similarly, bad performance in bad times is clearly a bad signal about firm value: it can either be due to bad idiosyncratic performance or high market correlation – both of which are undesirable attributes. In sum, cash flow news in downturns are unambiguous signals about firm value, in one direction or another. Therefore, investors place higher “weights” on information pertaining to firm performance in downturns than to performance in upturns. Table 3.1 illustrates the conclusions investors draw from observing good and bad news in good and bad times.

A more technical way of describing the mechanics is as follows. Estimates of $a$ and $b$ are positively correlated in upturns, and negatively in downturns. As a
CHAPTER 3. REVEALING DOWNTURNS

<table>
<thead>
<tr>
<th></th>
<th>high payoff</th>
<th>low payoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>market upturn</td>
<td>high $a$, high $b$</td>
<td>low $a$, low $b$</td>
</tr>
<tr>
<td>market downturn</td>
<td>high $a$, low $b$</td>
<td>low $a$, high $b$</td>
</tr>
</tbody>
</table>

Table 3.1: The table shows inferences about $a$ and $b$ from observing stocks’ cash flows in market upturns and downturns.

result, investors are better able to distinguish in bad times between “good firms” that have high $a$ and low $b$ and “bad firms” that have low $a$ and high $b$.

Yet another way of understanding the idea is to realize that to value the stock, investors put high weights on observations made when the stochastic discount factor is high, i.e., in downturns. The reason is that risk-averse investors care more about the performance of an asset in bad times, which is when they operate at a very steep part of their utility function. No effort choice or rational inattention (as used in the above-discussed extant theories) is needed to derive this result.

The theory explains the asymmetries in corporate finance and asset pricing as follows. Since (ia) prices react more strongly to news received in downturns, (ib) the cross-sectional dispersion of returns is higher at that time. As news will thus change valuations more significantly in downturns, the return to analyzing them correctly and trading on superior information is higher then. As a result of the larger spread between before-news and after-news asset prices, (ii) active funds will perform better in market downturns. Lastly, if replacing a CEO comes at a cost (see Taylor (2010)), boards will only move when they are reasonably sure that the manager in place performs worse than what they can expect to get by randomly drawing a new CEO from the pool of candidates. As the left tails of market values becomes fatter in downturns, more firms’ valuations fall below the threshold at which boards rationally sense that replacing the CEO is thus sufficiently likely to
add shareholder value. The model therefore predicts that (iii) boards are more likely to fire CEOs after observing the performance of the CEO-firm match in downturns.

Predictions (ia) and (ib) are in this form new to the literature. Predictions generating (ii) and (iii) have been made with more complicated models with stronger assumptions before, as described above. Providing empirical evidence for the claims (i) is subject of current research. We do not replicate the existing empirical evidence for (ii) and (iii), as ample evidence exists, which serves as the motivation of this paper.\(^4\)

The paper contributes to a growing literature on Bayesian learning in financial markets.\(^5\) In contrast to much of the literature which focuses on learning about the mean of a single asset’s payoff, we investigate the implications of learning about the covariance parameters of the assets’ fundamentals as well. The trick to keep the model tractable is to consider an economy with many assets, whose payoffs depend on several unknown parameters. For a large number of assets, there is no leaning about the aggregate payoff from cross-sectional observations, and as a result no learning about the stochastic discount factor.

The closest papers may be Veronesi (1999) and Veronesi and Ribeiro (2002). The respective purpose of their and our paper is different, however. Veronesi (1999) explains the asymmetric reaction to good and bad news at any time. We explain the asymmetric reaction to any type of news in good and bad times. Veronesi and Ribeiro (2002) explain higher co-movement of stock returns in recessions. We

---

\(^4\)Maybe because the cross-sectional dispersion is irrelevant for relative pricing in a neo-classical model, it has not received much attention in the literature (one exception is Duffee (2001)). However, it is relevant if stock pickers can exist, which is the presumption in our framework.

\(^5\)Pastor and Stambaugh (2012) study the seeming unconditional underperformance of active mutual funds. See Pastor and Veronesi (2009) for an overview over less recent research.
explain higher cross-sectional dispersion in recessions. The two are not mutually exclusive, see Figure 3.1 for an illustration.

Figure 3.1: Two stocks can be uncorrelated or negatively correlated in upturns and perfectly correlated in downturns, while relative valuations do not diverge in upturns, but diverge strongly in downturns.

In Veronesi and Ribeiro (2002), investors update their beliefs about the state of the world by observing potentially multiple securities’ cash flows, whose covariance matrix is known. In our model, investors can perfectly infer the state of the world at all times, but the covariance matrix is uncertain.

We do not discuss in detail our model’s implications for asymmetric patterns of time-series volatility between upturns and downturns, as in Mele (2007), and of investment decisions as in Dangl and Wu (2011).
The paper is organized as follows. Section 3.2 has the basic model. Section 3.3 discusses applications. Section 3.4 concludes.

3.2 Model

3.2.1 General Case

The Setup

There is a large number of assets \( i = 1, 2, \ldots, N \) in the economy. Each asset \( i \) pays dividends \( Y^i_t \) at time \( t \). The random vector \( Y_t = (Y^1_t, \ldots, Y^N_t) \) is iid. The distribution of \( Y^i_t \) depends on parameters \( \psi^i \), which are drawn from a known distribution. There is an overlapping generations representative agent with vNM utility \( u \).

As \( N \) is large, there is no learning about the aggregate dividend \( Y^i_t = \sum Y^i_t = N \cdot E[\psi^i] \). As a result, there is no learning about the stochastic discount factor (sdf) \( m_t \), which prices the uncertain dividend stream of each asset.

Lemma 3.1 The sdf in an OLG model is iid and given by

\[
m_{t+1} = \frac{u'(Y^i_{t+1})}{E_t[u'(Y^i_{t+1})]} \]

As there is no learning about \( m_t \), we can solve recursively for the price of asset \( i \):

\[
p^i_t = \sum_{k=1}^{\infty} \frac{1}{R^k} E_t[(m_{t+1} \cdots m_{t+k}) \cdot Y^i_{t+k}] = \sum_{k=1}^{\infty} \frac{1}{R^k} E_t(m_{t+1}) \cdots E_t(m_{t+k-1}) \cdot E_t(m_{t+k} \cdot Y^i_{t+k})
\]
\[
\sum_{k=1}^{\infty} \frac{1}{R^k} E_t[m_{t+k} \cdot Y_{t+k}^i] = \sum_{k=1}^{\infty} \frac{1}{R^k} E_t[m_{t+1} \cdot Y_{t+1}^i]
\]

\[
= \frac{1}{R-1} E_t[m_{t+1} \cdot Y_{t+1}^i]
\]

**Intuition**

The intuition for the main result is as follows. As per lemma 3.1, since agents are risk averse, the sdf is higher in downturns. Thus, for valuing the asset it is more important to know how the asset performs in downturns. As a result, the agent will put more weight on observations in downturns, compared to observations in the upturns.

To illustrate the intuition, we can approximate

\[
E_t[m_{t+1} \cdot Y_{t+1}^i] \approx \frac{1}{t} \sum_{\tau=1}^{t} m_{\tau} Y_{\tau}^i
\]

a consistent estimate of \( p_t \cdot (R-1) \). (Note that there is no asymmetric information in this model. The price therefore reflects the best estimate of value.)

As a result,

\[
p_t^i \approx \frac{1}{R-1} \frac{1}{t} \sum_{\tau=1}^{t} m_{\tau} Y_{\tau}^i
\]

Since the sdf is higher in the downturns, the price will react more to the observations received in downturns. There is not more information in prices in downturns, but the information there is is more important to gauge the utility the investor derives from holding the asset. We show this result more formally in a special case that assumes normal parameter distributions.
3.2.2 The Normal Case

The Setup

To obtain analytical solutions, we specialize the model. Let firm-level dividends depend on two parameters \( a^i \) and \( b^i \), \( Y_t^i = a^i + b^i \xi_t + \varepsilon_t^i \). \( a^i \) is firm \( i \)'s cash-flow alpha, or idiosyncratic performance parameter. \( b^i \) is the firm’s cash-flow beta, i.e., the correlation of the firm’s cash flows with a market-wide shock \( \xi_t \), which is distributed normally \( \xi_t \sim \mathcal{N}(0, \sigma_\xi^2) \). For each security, \( a^i \) and \( b^i \) are drawn from

\[
\begin{pmatrix}
a^i \\
b^i
\end{pmatrix}
\sim \mathcal{N}
\begin{pmatrix}
\bar{a} \\
\bar{b}
\end{pmatrix},
\begin{pmatrix}
\sigma_a^2 & \sigma_{ab} \\
\sigma_{ab} & \sigma_b^2
\end{pmatrix}
\]

Moreover, assume the OLG agents have utilities over cash flows \( u(Y) = -exp(-\gamma Y) \).

**Lemma 3.2** The price of asset \( i \) is

\[
p_t = \frac{1}{R-1} E_t[m_t Y_{t+1}^i] = \frac{1}{R-1} E_t[a^i - \phi b^i]
\]

where \( \phi = \gamma \bar{N} \bar{b} \cdot \sigma_\xi^2 \).

The parameter \( \phi \) captures the degree of risk aversion and the riskiness of the economy. Intuitively, risk-averse investors’ willingness to pay for the asset increases in the asset’s cash-flow alpha, and decreases in its cash-flow beta.

**Intuition**

Our main result is best illustrated by comparing in which state \( \xi_t \) of the economy the alignment is greatest between what the investor cares about and what she observes. The investor is interested in learning about \( \pi_t^i = (R-1) \cdot p_t^i = E_t[a^i - \phi b^i] \), and
observes cash flows $Y_t^i = a^i + b^i \cdot \xi_t + \varepsilon_t^i$, from which she can also infer the state of the economy $\xi_t$. That is, she calculates

$$
\pi_t^i = E_t \left[ a^i - \phi b^i \mid a^i + b^i \cdot \xi_t + \varepsilon_t^i \right]
$$

$$
= E_t \left[ a^i - \phi b^i \mid a^i - \phi b^i + b^i(\xi_t + \phi) + \varepsilon_t^i \right]
$$

What the investor observes and what she is interested in learning is most aligned when $\xi_t + \phi$ is close to zero, or when $\xi_t$ is close to $-\phi$, which is smaller than 0. In other words, one component of the noise is “shut off” when the market-wide shock is (moderately) negative.

**Formal Results**

We now show the results more formally. Conditional on the realization of $\xi_t$, beliefs about the distributions of $Y_t^i$ are mutually normally distributed. Thus, the conditional belief $\Omega_t$ at time $t$ about parameters $\psi^i = \begin{pmatrix} a^i \\ b^i \end{pmatrix}$ will always remain normal

$$
\Omega_t^i = \psi^i | I_t \sim \mathcal{N}(\mu_t, \Sigma_t)
$$

The standard equations for Bayesian updating of beliefs apply

$$
\mu_t = \mu_{t-1} + \frac{\text{cov}[\psi^i, Y_t^i]}{\text{var}[Y_t^i]} (Y_t^i - E[Y_t^i])
$$

$$
\Sigma_t = \Sigma_{t-1} - \frac{\text{cov}[\psi^i, Y_t^i] \text{cov}[\psi^i, Y_t^i]^T}{\text{var}[Y_t^i]}
$$
Here all expectations are conditional on information available at $t-1$ and $\xi_t$. For example, $\text{cov}[\psi^i, Y^i_t] = \text{cov}[\psi^i, Y^i_t|I_{t-1}, \xi_t]$.

If we denote

$$\Sigma_{t-1} = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}$$

Then

$$\text{var}[Y^i_t] = \sigma_a^2 + 2\sigma_{ab}\xi_t + \sigma_b^2\xi_t^2 + \sigma_\epsilon^2$$

$$\text{cov}[\psi^i, Y^i_t] = \begin{pmatrix} \sigma_a^2 + \sigma_{ab}\xi_t \\ \sigma_{ab} + \sigma_b^2\xi_t \end{pmatrix}$$

We can rewrite the asset price in vector notation

$$p^i_t = \frac{1}{R-1} E_t[a^i - \phi b^i] = \frac{1}{R-1} (1 - \phi) \cdot \mu_t$$

With that notation in place, we can derive the price change in response to a piece of news.

**Lemma 3.3** The price change of asset $i$ from time $t-1$ to time $t$, when the realization of the macro-shock is $\xi_t$, is

$$p^i_t - p^i_{t-1} = \frac{\lambda(\xi_t)}{R-1} \cdot (Y^i_t - E[Y^i_t])$$

where

$$\lambda(\xi_t) = \frac{\sigma_a^2 - \phi \sigma_{ab} - (\phi \sigma_b^2 - \sigma_{ab})\xi_t}{\sigma_a^2 + 2\sigma_{ab}\xi_t + \sigma_b^2\xi_t^2 + \sigma_\epsilon^2}$$
We can characterize how strongly prices react to news by the variance of the price changes.

\[
Var_p = Var[p_t - p_{t-1}] = \frac{(\lambda(\xi_t))^2}{(R-1)^2} \cdot Var[Y_t^i]
\]

\[
= \frac{1}{(R-1)^2} \cdot \frac{(\sigma_a^2 - \phi \sigma_{ab} - (\phi \sigma_b^2 - \sigma_{ab})\xi_t)^2}{\sigma_a^2 + 2\sigma_{ab}\xi_t + \sigma_b^2\xi_t^2 + \sigma_e^2}
\]

For a moment, think of the prior belief of correlations of \(a\) and \(b\), \(\sigma_{ab}\) as close to zero. For \(\sigma_{ab}\), the expression is simplified to

\[
Var_p = \frac{1}{(R-1)^2} \cdot \frac{(\sigma_a^2 - \phi \sigma_b^2\xi_t)^2}{\sigma_a^2 + \sigma_b^2\xi_t^2 + \sigma_e^2}
\]

Clearly, prices react more strongly to news received when \(\xi_t\) is smaller, i.e., in downturns. A proposition that makes this claim formally follows.

**Proposition 3.1** There exist \(0 < \bar{\sigma}_{ab}^p < \sigma_a \sigma_b\) such that

- \(Var_p(\xi_t = -x) > Var_p(\xi_t = +x)\) for \(-\sigma_a \sigma_b \leq \sigma_{ab} < \bar{\sigma}_{ab}^p\)

- \(Var_p(\xi_t = -x) < Var_p(\xi_t = +x)\) for \(\bar{\sigma}_{ab}^p < \sigma_{ab} \leq \sigma_a \sigma_b\)

The proposition says that there is an asymmetry in price responses between market upturns and downturns (see figure 3.2).

Unless the correlation between \(a^i\) and \(b^i\) is high (which means that we already know a lot about how the asset behaves in the downturns) the prices will react more strongly to news received in market downturns. Figure 3.3 shows an example of how the variance of price changes depends on the realization of the market-wide factor.

With empirical applications in mind, we also derive implications for returns (rather than prices, as above). Define returns as
Figure 3.2: The figure illustrates that when $\sigma_{ab}$ is less than some positive cutoff $\bar{\sigma}_{ab}$, the variance of prices changes in market downturns is higher than in market upturns.

\[ r_t = (p_t + Y_t) - p_{t-1} \]

An innovation in returns is then given by

\[ \eta_t = r_t - E[r_t] = \]

\[ = \left( \frac{\lambda(\xi_t)}{R - 1} + 1 \right) (Y_t - E[Y_t]) \]
Figure 3.3: Plot of the variance of the price changes over the realization of the market-wide shock $\xi$, for different values of $\rho = \frac{\sigma_{ab}}{\sigma_a \sigma_b}$. Unless $\rho_{ab}$ is strongly positive, the part of the graph left of $\xi = 0$ tends to be higher than the part to the right of $\xi = 0$, i.e., variance is higher in downturns than in upturns. That means, unless $\sigma_{ab}$ is very high, the cross-sectional dispersion in response to an observation is higher in downturns than in upturns. Simulation results for $\phi = 0.1$, $\sigma_a^2 = 1$, $\sigma_b^2 = 1$, $\sigma_{ab} = 0$, $\sigma_\varepsilon = 1$, and $R = 1$.

The variance of return innovations is then

$$Var_r = Var[\eta_t] = \left( \frac{\lambda(\xi_t)}{R - 1} + 1 \right)^2 \cdot Var[Y_t^i]$$

Analogous to proposition 3.1, we can then derive

**Proposition 3.2** There exist $\bar{\sigma}_{ab}(\sigma_a^2, \sigma_b^2, x) > 0$ such that for any $\sigma_{ab} < \bar{\sigma}_{ab}$

$$Var_r(\xi_t = -x) > Var_r(\xi_t = +x)$$
It is useful to note that the cross-sectional variance of price responses and returns are equal to the time series variances that we calculated in propositions 3.1 and 3.2.

Conditional on the realization of $\xi_t$, dividends $Y^i_t$ are independent across different stocks. As a result, the price changes and returns are independent across different stocks. Moreover, they have the same distributions for different stocks. Therefore, for a large number of stocks, the cross-sectional variance of price changes across stocks is equal to the conditional time series variances for a given stock.

### 3.2.3 Model Limitations and Extensions

As is common in the learning literature, we explicitly solve our model only for the normal case. It should be clear from the general setup, however, that normality is not driving the direction of our results. However, relaxing the normality assumption would certainly be useful for numerical evaluations and calibrations of the model. Similar things apply for the independence assumption for $\xi_t$. This assumption should be relaxed in more applied follow-up papers.

In the model, the parameters $a^i$ and $b^i$ are fixed and do not change over time. As a result, investors would eventually learn the true values. In the real world, the parameters change over time in response to changes of leadership, the competitive landscape, and innovations. As a result, investors never perfectly learn the true values. A stationary model that reflects these features of reality with time-changing parameters leads to similar results. To simplify the exposition, we therefore restrict the model in this paper to the “static case” above.
CHAPTER 3. REVEALING DOWNTURNS

3.3 Applications

3.3.1 CEO Turnover

We continue with the normal case: firm $i$ pays dividend $Y^i = a^i + b^i \xi_t + \varepsilon^i_t$ at time $t$; $a_i$ and $b_i$ represent a match between the CEO and the firm. $a_i$ can be interpreted as firm $i$ CEOs ability to generate cash flows in firm $i$, e.g., by efficiently running the organization. $b_i$ can be interpreted as the CEOs strategy in firm $i$. Some strategies generate high cash flows when the macro-economy is doing well (e.g., developing big luxury cars), while others are less profitable in upturns, but fare better in economic downturns (e.g., developing small fuel-efficient cars). For the moment, let there be no strategic choice of the strategy by the CEO. When a new CEO is hired, $a_i$ and $b_i$ are drawn from

\[
\left( \begin{array}{c}
  a^i \\
  b^i
\end{array} \right) \sim \mathcal{N} \left( \left( \begin{array}{c}
  \bar{a} \\
  \bar{b}
\end{array} \right), \left( \begin{array}{cc}
  \sigma_a^2 & \sigma_{ab} \\
  \sigma_{ab} & \sigma_b^2
\end{array} \right) \right)
\]

A CEO can be in place for at most two periods. After observing the performance of the CEO after one term in office, shareholders have an option to fire and replace the CEO with a new one that is chosen from a pool of available CEOs. The timing is illustrated in Figure 3.4.

There is no general equilibrium effect of firing the CEO. Replacing the CEO, however, comes at a cost $C$, which includes severance pay to the outgoing CEO and search and hiring frictions for the new CEO. We assume this cost is relatively large. As a result, the sdf is not affected by the option to fire, the option is far out of the money, and the ex ante value of option is small.
The value of the firm with a new CEO (who is unlikely to get fired in the next period) is

\[ p^{\text{new}} = \frac{1}{R} E_t^{\text{new}} [a^i - \phi b^i] + \frac{1}{R^2} \left( E_t^{\text{new}} [a^i - \phi b^i] + p^{\text{new}} \right) \]

\[ p^{\text{new}} = \frac{1}{R - 1} (\bar{a} - \phi \bar{b}) \]

The value of the firm with the old CEO (who will leave the firm after the next period for sure due to mandatory retirement) is

\[ p^{\text{old}} = \frac{1}{R} E_t^{\text{old}} [a^i - \phi b^i] + \frac{1}{R} p^{\text{new}} = \frac{1}{R} E_t^{\text{old}} [a^i - \phi b^i] + \frac{1}{R(R - 1)} (\bar{a} - \phi \bar{b}) \]
Shareholders decide to replace the CEO if and only if replacing is a NPV-positive project:

\[ p^{\text{new}} > p^{\text{old}} + C \iff E^\text{old}_t[a^i - \phi b^i] < \bar{a} - \phi \bar{b} - R \cdot C \]

Given that shareholders observe \( Y^i_t \),

\[ E^\text{old}_t[a^i - \phi b^i] - (\bar{a} - \phi \bar{b}) \]

\[ = (Y^i_t - \bar{Y}_t) \cdot \frac{\text{cov}[Y^i_t, a^i - \phi b^i]}{\text{var}[Y^i_t]} \]

\[ = (Y^i_t - (\bar{a} + \bar{b} \xi_t)) \cdot \frac{\sigma_a^2 - \phi \sigma_{ab} - (\phi \sigma_b^2 - \sigma_{ab}) \xi_t}{\sigma_a^2 + 2 \sigma_{ab} \xi_t + \sigma_b^2 \xi_t^2 + \sigma_e^2} \]

The dispersion of market prices of firms with “old” CEOs is higher when \( \xi \) is negative.

\[ V(\xi_t) = \text{Var}[E^\text{old}_t[a^i - \phi b^i]] = \frac{(\sigma_a^2 - \phi \sigma_{ab} - (\phi \sigma_b^2 - \sigma_{ab}) \xi_t)^2}{\sigma_a^2 + 2 \sigma_{ab} \xi_t + \sigma_b^2 \xi_t^2 + \sigma_e^2} \]

Analogous to the proof of proposition 3.1, there exists \( 0 < \bar{\sigma}_{ab}^p(\sigma_a^2, \sigma_b^2, x) < \sigma_a \sigma_b \), such that

- \( V(\xi_t = -x) > V(\xi_t = +x) \) when \( \sigma_a \sigma_b < \sigma_{ab}^p < \bar{\sigma}_{ab} \),
- \( V(\xi_t = -x) < V(\xi_t = +x) \) when \( -\sigma_a \sigma_b < \sigma_{ab}^p < \bar{\sigma}_{ab} \).

As a result, more firms’ market value falls below the cutoff \( C \) when the market-wide shock is negative. Therefore, more boards – rationally – decide to replace their CEO. See Figure 3.5 for an illustration.
Figure 3.5: An increasing number of firms fall below the cutoff below which replacing the CEO is NPV-positive when cross-sectional dispersion increases.
In words: unless $\sigma_{ab}$ is very large, firms tend to fire their CEOs in downturns more often than in upturns.

### 3.3.2 Active Funds Puzzle

Our model also provides a simple explanation why active funds perform better in market downturns relative to upturns. We only provide an explanation for this asymmetry. We do not provide a rationale why active funds underperform the market unconditionally, which is a more controversial claim in the literature.

In a world with only active investors and noise/liquidity traders as counter parties, prices reflect fundamental information due to active funds’ trading of mispriced assets. Fundamental news lets funds detect such mispriced assets. Active investors buy or sell the mispriced asset until it adjusts to fair value. During the process of the price adjustment they can buy or sell the asset at better than a fair price. The counter-party that loses money is noise and liquidity traders. During market downturns, funds detect more severe mispricing, and make more money as a result.

We do not microfound the process of price adjustment as existing models (e.g., Kyle (1985)) have shown that active funds make is proportional to the variance in price changes – and that variance is higher in downturns.

### 3.4 Discussion

This paper provided a rationale for asymmetries between upturns and downturns both in asset pricing and in corporate finance. The theory assumes that (i) investors try to infer firm value from cash flow news, (ii) both firm value and cash flows depend on multiple parameters, one of which is a factor affecting all stocks. As a result, the extent to which the parameters governing cash flows and value are aligned
is different in upturns and downturns. Therefore, investors learn more about firm
value from any given observation sampled in a downturn rather than one sampled in
an upturn. As market prices thus react more strongly to news in downturns, active
investing based on superior information (“stock picking”) tends to produce greater
abnormal returns, and boards tend to fire CEOs in downturns more frequently than
in upturns.

The focus of future research should be on a thorough empirical validation of the
empirical predictions of the model. One is that a greater cross-sectional dispersion
after downturns is caused by a stronger stock market reaction to news in downturns
than in upturns. An additional and more direct consistency test is to track the
confidence intervals of stock analysts over the business cycle. These tests are subject
to ongoing research.

Appendix

Proof of Lemma 3.1. (Stochastic Discount Factor for the
OLG model)

Consider an agent that buys at time $t$ an asset at price $p_z$ that pays $Z_{t+1}$ at $t + 1.$
The agent consumes $Y_{t+1}$ at $t + 1.$ Then, the expected utility of the agent

$$U(x) = E_t[u(Y_{t+1} + x(Z_{t+1} - R p_z))]$$

The representative agent’s utility is maximized when $x = 0$

$$0 = U'(x)|_{x=0} = E_t[u'(Y_{t+1})(Z_{t+1} - R p_z)]$$
Therefore,
\[ p_z = \frac{1}{R} E_t \left[ \frac{u'(Y_{t+1})}{E_t[u'(Y_{t+1})]} Z_{t+1} \right] \]
and the stochastic discount factor is equal to
\[ m_{t+1} = \frac{u'(Y_{t+1})}{E_t[u'(Y_{t+1})]} \]

**Proof of Lemma 3.2. (Valuation for the Case of Normally Distributed Parameters)**

For normally distributed \( Y \)
\[ E[e^{\gamma Y}] = e^{\gamma E[Y] + \frac{\gamma^2 V}{2} Y} \]
\[ E[Y \cdot e^{\gamma Y}] = (E[Y] + \gamma V[Y]) \cdot e^{\gamma E[Y] + \frac{\gamma^2 V}{2} Y} = (E[Y] + \gamma V[Y]) \cdot E[e^{\gamma Y}] \]

Aggregate consumption is equal to \( Y_{t+1} = N(\bar{a} + \bar{b} \cdot \xi_{t+1}) \) and the stochastic discount factor is
\[ m_{t+1} = \frac{u'(Y_{t+1})}{E_t[u'(Y_{t+1})]} \]

Therefore, for exponential utility \( u'(Y_{t+1}) = \gamma e^{-\gamma Y_{t+1}} \)
\[ G(\psi^i) = E_t[m_{t+1} Y_{t+1} | \psi^i] = \frac{1}{E_t[e^{-\gamma Y_{t+1}}]} E_t \left[ e^{-\gamma Y_{t+1}} (a^i + b^i \xi_{t+1} + \varepsilon_{t+1}^i) | \psi^i \right] \]

Since (using above formulas)
\[ E_t \left[ e^{-\gamma Y_{t+1}} \xi_{t+1} \right] = E_t \left[ e^{-\gamma N(\bar{a} + \bar{b} \xi_{t+1})} \xi_{t+1} \right] = -\gamma N \bar{b} \sigma_\xi^2 \cdot E_t \left[ e^{-\gamma Y_{t+1}} \right] \]

then \( G(\psi^i) = a^i - \phi \cdot b^i \) where \( \phi = \gamma N \bar{b} \cdot \sigma_\xi^2 \).
Thus, the price of the asset follows

\[ p_t^i = \frac{1}{R-1} E_t[G(\psi^i)] = \frac{1}{R-1} E_t[a^i - \phi b] \]

**Proof of Lemma 3.3**

In the defined notation

\[ \text{var}[Y_t^i] = \sigma_a^2 + 2\sigma_{ab}\xi_t + \sigma_b^2\xi_t^2 + \sigma_\varepsilon^2 \]

\[ \text{cov} \left[ \begin{pmatrix} a^i \\ b^i \end{pmatrix}, Y_t^i \right] = \begin{pmatrix} \sigma_a^2 + \sigma_{ab}\xi_t \\ \sigma_{ab} + \sigma_b^2\xi_t \end{pmatrix} \]

\[ \mu_t = \mu_{t-1} + \frac{Y_t^i - E[Y_t^i]}{\text{var}[Y_t^i]} \begin{pmatrix} \sigma_a^2 + \sigma_{ab}\xi_t \\ \sigma_{ab} + \sigma_b^2\xi_t \end{pmatrix} \]

Therefore

\[ p_t - p_{t-1} = (1, -\phi)(\mu_t - \mu_{t-1}) = \frac{Y_t^i - E_t-1[Y_t^i]}{\text{var}[Y_t^i]}(\sigma_a^2 + \sigma_{ab}\xi_t - \phi(\sigma_{ab} + \sigma_b^2\xi_t)) \]

and we get the following expression for \( \lambda(\xi_t) \)

\[ \lambda(\xi_t) = \frac{\sigma_a^2 - \sigma_{ab}\phi - \xi_t(\phi\sigma_b^2 - \sigma_{ab})}{\sigma_a^2 + \sigma_b^2\xi_t^2 + \sigma_\varepsilon^2 + 2\sigma_{ab}\xi_t} \]

**Proof of Proposition 3.2**

Although we know that

\[ \text{Var}_p = \frac{(\sigma_a^2 - \sigma_{ab}\phi - \xi_t(\phi\sigma_b^2 - \sigma_{ab}))^2}{\sigma_a^2 + \sigma_b^2\xi_t^2 + \sigma_\varepsilon^2 + 2\sigma_{ab}\xi_t} \]
it is easier to get to the result indirectly. The updated conditional variance is

\[
\Sigma_t = \begin{pmatrix}
\sigma_a^2 & \sigma_{ab} \\
\sigma_{ab} & \sigma_b^2
\end{pmatrix}
- \frac{1}{\text{var}[Y_t^i]}
\begin{pmatrix}
\sigma_a^2 + \sigma_{ab} \xi_t \\
\sigma_{ab} + \sigma_{ab}^2 \xi_t
\end{pmatrix}
\begin{pmatrix}
\sigma_a^2 + \sigma_{ab} \xi_t \\
\sigma_{ab} + \sigma_{ab}^2 \xi_t
\end{pmatrix}^T
\]

\[
= \frac{1}{\text{var}[Y_t^i]}
\begin{pmatrix}
\sigma_a^2 \sigma_b^2 + (\sigma_a^2 \sigma_b^2 - \sigma_{ab}^2) \xi_t^2 & \sigma_{ab} \sigma_{ab}^2 - (\sigma_a^2 \sigma_{ab}^2 - \sigma_{ab}^2) \xi_t \\
\sigma_{ab} \sigma_{ab}^2 - (\sigma_a^2 \sigma_b^2 - \sigma_{ab}^2) \xi_t & \sigma_b^2 \sigma_{ab}^2 + (\sigma_a^2 \sigma_{ab}^2 - \sigma_{ab}^2)
\end{pmatrix}
\]

Then (here \(\Phi = (1, -\phi)\))

\[
\text{Var}[p_t - p_{t-1}] = \text{Var}[\Phi(\mu_t - \mu_{t-1})] = \Phi \text{Var}[\mu_t - \mu_{t-1}] \Phi^T = \\
= \Phi(\Sigma_{t-1} - \Sigma_t) \Phi^T
\]

Since \(\Sigma_{t-1}\) does not depend on \(\xi_t\) we can consider only the second term.

\[
H = \Phi \Sigma_t \Phi^T = \frac{\sigma_a^2 (\sigma_a^2 - 2\phi \sigma_{ab} + \phi^2 \sigma_b^2) + (\sigma_a^2 \sigma_b^2 - \sigma_{ab}^2)(\phi^2 + 2\phi \xi_t + \xi_t^2)}{\sigma_a^2 + 2\sigma_{ab} \xi_t + \sigma_{ab}^2 \xi_t^2 + \sigma_b^2 \xi_t^2}
\]

If we denote

\[
A = \sigma_a^2 (\sigma_a^2 - 2\phi \sigma_{ab} + \phi^2 \sigma_b^2) + (\sigma_a^2 \sigma_b^2 - \sigma_{ab}^2)(\phi^2 + x)
\]

\[
B = 2\phi (\sigma_a^2 \sigma_b^2 - \sigma_{ab}^2)
\]

\[
C = \sigma_a^2 + \sigma_b^2 x^2 + \sigma_{ab}^2
\]

\[
D = 2\sigma_{ab}
\]

then \(H(\xi_t) = \frac{A + B \xi_t}{C + D \xi_t}\) and
\[ H_{\xi=+x} > H_{\xi=-x} \Leftrightarrow A \cdot D < B \cdot C \]

The last equation is equivalent to

\[ \left( \sigma^2 \left( \sigma^2_a - 2\phi \sigma_{ab} + \phi^2 \sigma^2_b \right) + \left( \sigma^2_a \sigma^2_b - \sigma^2_{ab} \right) (\phi^2 + x^2) \right) \cdot 2\sigma_{ab} < \]

\[ 2\phi \left( \sigma^2_a \sigma^2_b - \sigma^2_{ab} \right) \cdot (\sigma^2_a + \sigma^2_b x^2 + \sigma^2_a x^2) \Leftrightarrow \]

\[ \sigma_{ab} \left( \sigma^2_a - 2\phi \sigma_{ab} + \phi^2 \sigma^2_b \right) \cdot \left( \sigma^2_a \sigma^2_b - \sigma^2_{ab} \right) + (\phi^2 + x^2) < 2\phi \left( \sigma^2_a + \sigma^2_b x^2 + \sigma^2_a x^2 \right) \]

We can rewrite the left hand side as

\[ \sigma_{ab} \left( \sigma^2_a - 2\phi \sigma_{ab} + \phi^2 \sigma^2_b \right) \cdot \left( \sigma^2_a \sigma^2_b - \sigma^2_{ab} \right) + (\phi^2 + x^2) + 2\phi \sigma_{ab} \cdot \left( \sigma^2_a \sigma^2_b - \sigma^2_{ab} \right) \]

The expression is negative for negative \( \sigma_{ab} \), is equal to zero when \( \sigma_{ab} = 0 \), and is an increasing function of \( \sigma_{ab} \) for positive \( \sigma_{ab} \). Moreover, \( H(\xi_t = -x) > H(\xi_t = x) \) for \( \sigma_{ab} = \sigma_a \sigma_b \). Thus, there exists a positive cutoff \( 0 < \sigma_{ab}^P < \sigma_a \sigma_b \) such that

- \( \text{Var}_p[\xi_t = -x] > \text{Var}_p[\xi_t = +x] \) for \( -\sigma_a \sigma_b \leq \sigma_{ab} < \sigma_{ab}^P \)

- \( \text{Var}_p[\xi_t = -x] < \text{Var}_p[\xi_t = +x] \) for \( \sigma_{ab}^P < \sigma_{ab} \leq \sigma_a \sigma_b \)

**Proof of Proposition 3.2**

We know that

\[ \text{Var}_r = \left( \frac{\lambda}{R-1} + 1 \right)^2 \cdot \text{Var}[Y_t] = \]
\[
\lambda^2 \text{Var}[Y^i_t] + \frac{2\lambda \text{Var}[Y^i_t]}{R - 1} + \text{Var}[Y^i_t] = \\
= \text{Var}_p + \frac{1}{R - 1} (2\lambda \text{Var}[Y^i_t] + \text{Var}[Y^i_t])
\]

From proposition 3.1 there exists \(\bar{\sigma}^p_{ab}\) such that for \(\sigma_{ab} < \bar{\sigma}^p_{ab}\)

\[\text{Var}_p[\xi_t = -x] > \text{Var}_p[\xi_t = +x]\]

Now

\[Q = 2\lambda \text{Var}[Y^i_t] + (R - 1)\text{Var}[Y^i_t] =
\]

\[2(\sigma_a^2 - \phi\sigma_{ab} - \xi_t(\phi\sigma_b^2 - \sigma_{ab})) + (R - 1)(\sigma_a^2 + \sigma_b^2 \xi_t^2 + \sigma_e^2 + 2\sigma_{ab}\xi_t) =
\]

\[(R + 1)\sigma_a^2 - 2\phi\sigma_{ab} + (R - 1)(\sigma_b^2 \xi_t^2 + \sigma_e^2) - 2\xi_t(\phi\sigma_b^2 - \sigma_{ab}R)
\]

Again there exists \(\bar{\sigma}^q_{ab}\) such that for \(\sigma_{ab} < \bar{\sigma}^q_{ab}\)

\[Q(\xi_t = -x) > Q(\xi_t = +x)\]

Taking \(\bar{\sigma}^r_{ab} = \min(\bar{\sigma}^p_{ab}, \bar{\sigma}^q_{ab})\) we get the result of the proposition.
Bibliography


Implications of Investor Attention for Price and Earnings Momentum,” mimeo.

paper, National Bureau of Economic Research.

HVIDE, H. K. (2002): “Pragmatic Beliefs and Overconfidence,” Journal of Eco-


KACPERCZYK, M., S. V. NIEUWERBURGH, AND L. VELDKAMP (2011): “Time-
Varying Fund Manager Skill,” National Bureau of Economic Research Working
Paper Series, No. 17615.


investment is correlated with cash flow?,” National Bureau of Economic Research
Cambridge, Mass., USA.


