The Role of Investment Banks in IPOs and Incentives in Firms

Essays in Financial and Behavioral Economics

Inaugural-Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2004

vorgelegt von
Björn Bartling

Referent: Prof. Dr. Klaus M. Schmidt
Korreferent: Prof. Sven Rady, Ph.D.
Promotionsabschlussberatung: 21. Juli 2004

Acknowledgements

First and foremost I would like to thank my supervisor Klaus Schmidt. This thesis would not have been possible without his superb support, guidance, and encouragement. I am also much indebted to my co-authors Andreas Park and Ferdinand von Siemens. The joint work with both of them was and continues to be a great source of inspiration and motivation.

Just as much I would like to thank my colleagues Brigitte Gebhard, Georg Gebhardt, Florian Herold, and Susanne Kremhelmer. They all contributed in many ways to the completion of this thesis and provided an inspiring and very pleasurable environment at the Seminar für Wirtschaftstheorie. I received comments and suggestions from many other friends and colleagues. My thanks extend to all of them – even more so as they are too numerous to be all mentioned in name at this point.

Finally, I would like to thank the Faculty of Economics and Politics at Cambridge University and Wolfson College Cambridge for hosting me as a visiting Ph.D. student in the academic year 2000/01. Financial support from the European Commission, grant no. HPMT-CT-2000-00056, is gratefully acknowledged.

Contents

Preface

This dissertation is comprised of two parts. Chapters 1 and 2 address the role of investment banks in initial public offerings. Chapters 3 and 4 analyze incentive provision when agents are subject to a 'behavioral bias'.

In Chapter 1 we model the procedure of an *initial public offering* (IPO) as a signaling game and analyze how the possibility of potentially profitable trading in the aftermarket influences pricing decisions by investment banks. When maximizing the sum of both the gross spread of the offer revenue and profits from aftermarket trading, investment banks have an incentive to distort the offer price by employing aftermarket short covering and exercise of the overallotment option strategically. This results either in informational inefficiencies or, on average, exacerbated underpricing. Wealth is redistributed in favor of investment banks.

In Chapter 2 we address two puzzles of the IPO literature: (1) Why do investment banks earn positive profits in a competitive market? And (2) Why do banks receive lower gross spreads in VC backed IPOs? The IPO procedure is modeled as a two-stage signaling game. In the second stage banks set offer prices given their private information and the level of the spread. Issuers anticipate the bank's pricing decision and set in the first stage spreads to maximize expected revenue. Investors are aware of this process and subscribe only if their expected profits are non-negative. As a result, issuers offer high spreads to induce banks to set high prices, allowing them profits. Competition may take place in additional features of the IPO contract as, for example, the number of co-managers or analyst coverage. We show that in equilibrium superiorly informed VC backed issuers impose smaller spreads.

In Chapter 3 we examine self-control problems – modeled as time-inconsistent, present-biased preferences – in a multi-tasking environment. An agent must allocate effort between an incentivized and immediately rewarded activity (e.g. effort at the workplace) and a private activity that pays out only tomorrow (e.g. studying for a degree). Present-biased agents take decisions that do not maximize their long-run welfare, irrespective of the intensity of incentives. Sophisticated agents are never harmed by incentives relative to the case where incentives are absent as they always receive their reservation utility levels. However, naive agents are always harmed in the presence of incentives as they wrongly predict future behaviors. Furthermore, we show that the loss to a naive agent can exceed the principal's gain from providing incentives. In this case social welfare is reduced if the principal provides incentives.

In Chapter 4 we analyze how *inequity aversion* interacts with incentive provision in an otherwise standard moral hazard model with two risk averse agents. We identify the conditions under which inequity aversion increases agency costs of providing incentives. We show, first, that inequity aversion can render equitable flat wage contracts optimal even though incentive contracts are optimal with selfish agents. Second, to avoid social comparisons the principal may employ one agent only, thereby forgoing the efficient effort provision of the second agent. We finally discuss the implications of social preferences for the internal organization and the boundary of the firm.

The decision whether or not to conduct an *initial public offering* is an important decision in the life cycle of a firm. The advantages of having shares in a firm quoted on a stock exchange are manifold. The owner of a firm can realize part of her investments, it includes the ability to raise additional equity finance, or even the opportunity to set up share option plans as incentive device for employees. However, there are also costs of going public. In this context, *initial underpricing* is most extensively discussed. Ritter and Welch (2002) report for 6,249 IPOs in the U.S. between 1980 and 2001 an average first-day return of 18.8 percent. It is usually argued that initial underpricing constitutes a wealth transfer from the owner of the firm to the new shareholders, and as such can be regarded as a cost of going public.

A number of explanations have been advanced for the 'underpricing anomaly' which seems to violate the fundamental tenet of 'no arbitrage'. The most prominent ones assume informational asymmetries between (or among) some of the main parties involved: the issuing firm, the investment bank, and the investors.

Rock (1986) proposes a variant of Akerlof's (1970) 'lemons problem'. He assumes asymmetric information between different types of investors. Some are perfectly informed about the intrinsic value of the shares on offer whereas others are uninformed. Given the presence of informed investors, uninformed investors face a 'winner's curse'. Informed investors subscribe only to 'hot' IPOs. Assuming that shares are rationed, uninformed investors stand a greater chance of being allocated shares in 'cold' IPOs from which informed investors abstain. To however attract uninformed investors to subscribe to IPOs, shares have to be underpriced on average.

Another strand of the literature assumes asymmetric information between the issuing firm and the investors. The signaling models by Allen and Faulhaber (1989), Grinblatt and Hwang (1989), and Welch (1989) argue that underpricing can – in analogy to Spence's (1973) job market signaling – be a signal for a high 'quality' of the firm. A single crossing property is established by assuming that subsequent to the IPO a secondary offering is conducted. In between these two offerings new information may arise and reveal a low quality firm's true value. This firm will then be unable to recoup the loss from underpricing its shares by way of a secondary offering. A separating equilibrium can thus be established in which only high quality firms underprice because they can reap the gain from doing so in the secondary offering.

Apart from missing empirical support for signaling theories of underpricing[1] there is the question why firms would not opt for a different, less costly signal? Booth and Smith (1986), for example, put forth a theory of investment bank choice. Investment banks as repeated players have reputational capital at stake and can thus certificate the value of a firm. Other theories stressing the role of investment banks include Benveniste and Spindt (1989). They assume asymmetric information between investment

[1]Helwege and Liang (1996) report for a U.S. sample of IPOs in 1983 that only 4 percent of firms conducted a secondary offering in the subsequent 10 years.

bank and investors. The latter hold superior information about the value of the shares, and they are assumed to subscribe repeatedly to IPOs. Benveniste and Spindt design a mechanism in which banks use underpricing and rationing to elicit investors' information prior to an IPO. If shares are underpriced, investors can be punished by small allocations in subsequent offerings of other firms if the post-IPO phase reveals that material information was withheld.

The existing theoretical literature however almost completely neglects that the role of investment banks does not end with the distribution of shares at the day of the offering. In fact investment banks pursue supposedly price stabilizing activities in the aftermarket of IPOs that provide potentially profitable trading opportunities. This is where the model in Chapter 1 adds to the literature. We explicitly account for stabilizing activities by investment banks in the aftermarket of an IPO and analyze how this influences the offer price decision in the first place.

The regulating authorities allow investment banks to establish a short position in an IPO by selling more shares than initially announced. *Aftermarket short covering* refers to the practice of filling these positions in the aftermarket of an IPO. This is done if the market price falls below the offer price. The idea is that filling short positions stabilizes prices by increasing demand. The difference between market price and offer price is – along the way – pure profit for the investment bank. If the price instead rises, the bank is hedged by an *overallotment option* which grants the right to obtain additional shares from the issuer at the offer price. The U.S. Securities and Exchange Commission (SEC) and the Committee of European Securities Regulators (CESR) put forward the argument that stabilizing activities ensure an 'orderly market' as sudden selling pressure can be countered. In their latest respective release the SEC (1997, p. 81) opines that aftermarket price stabilization "promotes the interests of shareholders, underwriters, and issuers."

In Chapter 1 we challenge this view by showing that – in the context of our model – stabilizing activities result in either informational inefficiencies or, on average, exacerbated underpricing. Furthermore, wealth is redistributed in favor of investment banks.

We presume that these 'side effects' will not be intended by the regulating authorities. Even without trading off potential beneficial effects of stabilization against our findings, a policy implication arising from the analysis might be the alert that current, well meant regulation can be gamed to the disadvantage of issuers and investors.

We propose a signaling model of the IPO procedure in which both the investment bank and investors hold private information about the intrinsic value of the shares. The bank moves first and sets the offer price. Besides possible trading profits in the aftermarket, banks are directly remunerated for their services by a fraction of the offering revenue, the *gross spread*. In our model banks choose the offer price strategically to maximize their profits form both the gross spread and trading profits in the aftermarket. A higher offer price promises a higher revenue, it however reduces the probability that the IPO is successful. An IPO gets called off if there are not enough investors subscribing to it, and a higher offer prices reduces the number of investors subscribing.[2]

As benchmark, in a setting without aftermarket activities we identify the conditions for the price equilibrium to be separating. In a separating equilibrium banks with different information set different offer prices. A bank with favorable information about the value of the firm deems it more likely that enough investors will hold alike information. It will thus set a higher price than a bank with less favorable information. We call a separating equilibrium *informationally efficient* since the bank's information is fully revealed by the offer price. In the aftermarket prices adjust according to market demand. In equilibrium the security can turn out to be either under- or overpriced, but on average there is underpricing.

We then introduce stabilizing activities to the model. This augments the incentive to set high offer prices because the potential profit from aftermarket activities is higher at higher prices. We find that – relative to the benchmark – either the offer price falls on average or there is a pooling offer-price equilibrium. In the first case, to uphold a separating equilibrium, an investment bank with favorable information distorts the

[2]Busaba, Benveniste, and Guo (2001) report that about 14 percent of cases in their U.S. sample of more than 2,500 IPOs between 1984-1994 get called off.

price downwards. This increases, on average, underpricing, i.e. the cost of going public. In the second case, a separating equilibrium cannot be upheld and investors are thus unable to infer the investment bank's signal from the offer price. This equilibrium is informationally inefficient since investors' decisions are based on private information only and not, in addition, on the information of the bank. A major objective of financial market regulation is market transparency. It is thus beneficial if prices contain more rather than less information and, consequently, pooling equilibria are undesirable.

The debate among financial economists on the costs of going public has mainly focused on initial underpricing. Other, direct costs of going public like legal expenses, audit fees, management time, accountancy, and the gross spread as investment bank compensation, have received relatively little attention. Among practitioners matters are different. When Chen and Ritter (2000) published that in more than 90 percent of IPOs the gross spread is exactly 7 percent, numerous lawsuits against investment banks for price collusion and a U.S. Department of Justice investigation of "alleged conspiracy among securities underwriters to fix underwriting fees" were initiated.[3] Chapter 2 proposes a different, subtle explanation of why gross spreads are so high that investment banks are left with profits despite market competition. Furthermore, we address the related puzzle of why venture capital (VC) backed IPOs are associated with lower gross spreads than non-VC backed IPOs. To the best of our knowledge, Chapter 2 offers the first theoretical model to explain the level of gross spreads.

In Chapter 2 we model the IPO procedure as a two-stage signaling game. As in Chapter 1 we assume that both investment banks and investors hold private information about the intrinsic value of the shares. While Chapter 1 is silent about the role of issuers, in Chapter 2 two different types of issuers are explicitly modeled. VC backed issuers are assumed to hold private information about the value of the firm. In contrast, non-VC backed issuers are taken to be uninformed. In the second stage of the signaling game the investment bank decides on the offer price, given both its private information and the level of the spread. Issuers anticipate the bank's pricing

[3]See Hansen (2001) for an overview of reactions to the Chen and Ritter (2000) article.

decision. Hence, in the first stage they set spreads strategically to maximize expected revenue. On both stages we can have either pooling or separating equilibria in spreads and prices, respectively. Investors are aware of this process and subscribe only if their expected profits are non-negative.

We find that it can be in the best interest of the issuer to offer high spreads. It is the investment bank's discretion to set the final offer price. At high offer prices there is the danger that the IPO gets called off, because at high prices there may not be enough investors to subscribe to the offering. We assume that the bank then suffers a reputation loss. To nevertheless induce the bank to set a high offer price the issuer must offer a high level of the gross spread. The investment bank then earns a rent: Given any level of the gross spread, it could always deviate to a low price (at which the IPO will never fail) and receive its share in the offer revenue with certainty. Hence, to prevent banks from deviating they are offered a rent at high prices.

Our second main result addresses the differences in spreads between VC and non-VC backed issuers. We show that in equilibrium superiorly informed VC backed issuers impose smaller spreads. A VC backed issuer with 'good news' about the value of the shares regards it as likely that the investment bank also holds favorable information. The issuer then wants the bank to transform this information to the investors via a separating offer price equilibrium. An issuer with 'bad news' will however always mimic the issuer with favorable information: Issuers receive more than 90 percent of the offer revenue and thus have a strong interest in high prices – and signaling 'bad news' via the level of the spread reduces investors' rating of the shares. In equilibrium we thus observe a pooling spread level and separation offer prices. In contrast, uninformed non-VC backed issuers prefer the bank to hide its information and set a pooling price (with uninformed issuers the level of the spread itself cannot carry information). We show that the according spread offered by non-VC backed issuers is smaller than the separating price equilibrium inducing spread set by VC backed issuers.

The economic agents in Chapters 1 and 2 – the investors, investment banks, and issuers – are modeled in accordance with the standard paradigm of neoclassical economics: Individuals have stable and coherent preferences, are purely self-interested, and fully rational. The investors have no trouble in solving a rather complicated Perfect Bayesian Equilibrium in order to decide whether to order or to abstain. The investment bank is interested in its own material payoff only and does not care that some actions taken may harm others. And the issuer does not waive his decision to go public just because the day of the offering arrives.

Over the last decade, however, empirical and experimental evidence mounted against the paradigm of *homo economicus*. Numerous studies have shown that even in economically relevant environments people systematically deviate from the predictions of the standard theory. The resulting pursuit for greater psychological realism led to the field of *behavioral economics*, which extends the scope of economics by incorporating findings from experimental economics, psychology, and sociology into economic theory.[4] The models in Chapters 3 and 4 pay tribute to this development. We apply recent behavioral insights, especially *present-biased preferences* and *inequity aversion*, to contract theory in order to analyze how optimal incentive provision within firms changes if the 'idea of agent' is broadened.

One strand of the literature in behavioral economics analyzes the consequences of *time-inconsistent preferences*. In a parsimonious way, the standard model of *exponential discounting* captures the fact that people have a preference for immediate gratification. Exponential discounting however implies, in addition, that intertemporal trade-offs remain unaffected no matter when a decision is taken. The evidence, by contrast, shows that people exhibit *present-biased preferences*.[5] They show very sharp impatience for short horizons but are much more patient at long horizons. By way of example: When being asked to decided whether to work 8 hours on, say, Monday four weeks from now and relax on Tuesday or, alternatively, to relax on Monday and work 9 hours on Tuesday, most people will opt for the first choice. However, when being asked again Sunday four weeks from now, present-biased preferences may come into

[4]For an overview see, for example, Camerer, Loewenstein, and Rabin (2003).

[5]In the wording of Rabin (2002): "As absurd as it sounds, it is probably true to say that exactly zero papers in all social and behavioral sciences have proposed a test of the basic exponential versus hyperbolic discounting [...] and claimed exponential explains the generated data better." (p. 19)

play. Many people then reverse their decisions and put off the work for another day – even though this implies to work an extra hour. This kind of behavior is often referred to as *hyperbolic discounting*.

The theory of hyperbolic discounting has very fruitfully been applied to savings decisions. Laibson (1997) shows that present-biased preferences can explain why consumption tracks income more tightly than predicted by the standard life-cycle model of savings, especially in the absence of commitment devices. O'Donoghue and Rabin (1999c) apply hyperbolic discounting to retirement planning. In the U.S. many employees are eligible for a so called 401(k) retirement savings plan. Participation rates in these savings plans are relatively low, which is surprising because they are subsidized by the government and sometimes by the employer as well. Even though present-biased agents will want to eventually participate (in the long run people are patient), there is always something that hinders them to join because it promises a greater immediate reward (in the short run people are impatient). Consequently, these agents may procrastinate indefinitely. O'Donoghue and Rabin's theory is well supported by recent evidence. Madrian and Shea (2001) find that automatic enrollment of employees in 401(k) plans (employees must choose to opt out of rather than opt into the plan) exerts a strong influence over their saving choices.

O'Donoghue and Rabin (1999b) examine the implications of hyperbolic discounting for incentive provision. They analyze a principal-agent setting in which the agent must accomplish a single task but has discretion when to do it. A procrastinating agent is assumed to face stochastic costs of completion. The principal has an interest in timely completion and thus offers an incentive contract to induce the agent to finish the task in time. With asymmetric information about costs timely completion and efficient delay must be traded-off. The optimal contract involves 'deadlines' and increasing punishment for delay.

The model in Chapter 3 adds to the literature by further exploring how present-biased preferences and incentive provision interact. Complementary to O'Donoghue and Rabin (1999b), we examine self-control problems in a *multi-tasking* environment.

An agent must allocate effort between an incentivized and immediately rewarded activity – effort at the workplace – and a private activity that pays out tomorrow only – say, caring for one's health. The seminal contribution on multi-tasking principal-agent theory is Holmström and Milgrom (1991). They analyze the implications on optimal incentive intensity, ownership structure, and job design if some activities are more difficult to measure than others. The focus of Chapter 3 is different. We explore how a present-biased agent's allocation of effort between different tasks is affected by incentives if effort invested in some tasks pays out immediately while effort devoted to other tasks pays out with some delay.

Following the literature we assume that present-biased agents can be either naive or sophisticated. Naive agents are ignorant to the fact that they exhibit time-inconsistent preferences. They think in the future they will act like time consistent agents. In contrast, sophisticated agents are aware of their time-inconsistent preferences. In the context of our model, the difference in crucial. We consider a three period setting. In the first period a principal offers an incentive contract and the agent decides whether or not to participate. In the second period the agent chooses effort levels whose cost accrue immediately. While effort devoted to the principal's purposes is remunerated in the second period, effort devoted to the agent's private benefit pays out in the final period only. Therefore, by the time a naive agent is offered a contract he holds – in contrast to a sophisticated agent – a too optimistic belief about second period behavior. The principal takes agents' beliefs into account when offering a contract, and we show that he can exploit naiveté.

The focus of the model is the comparison between a situation in which the agent can engage in the private activity only, and a situation in which the principal adopts an additional production opportunity (for engagement in which he offers incentives). We show that present-biased agents take decisions that do not maximize their long-run welfare, irrespective of the presence of incentives. Sophisticated agents are, however, never harmed additionally by the adoption of incentive contracts relative to the case without incentives. They always receive their reservation utility levels. In contrast, naive agents can be harmed if the principal offers an incentive contract. The incentive contract endows them with an additional 'occasion' to give in to their present bias. Due to the fact that naive agents erroneously belief to act like time-consistent agents in the future, they do not ask for – and consequently do not get – a compensating payment when accepting the contract. Furthermore, we show that the resulting loss to the naive agent can even exceed the principal's gain from providing incentives. In this case, we find that social welfare is reduced by the adoption of incentive contracts.

Another strand of the literature in behavioral economics is concerned with the notions of *fairness* and *reciprocity*. The standard hypothesis on human motivation assumes that all people are exclusively interested in their material self-interest. In recent years results from experimental economics have challenged this hypothesis by showing that many people are, in addition, strongly motivated by concerns for fairness and reciprocity. The most prominent experiment in this context is the *ultimatum game*. A pair of subjects has to agree on a division of fixed amount of money. One subject (the proposer) must propose a division of the amount. The other subject (the responder) can accept or reject the division. In case of rejection both subjects receive nothing. In case of acceptance the proposed division is implemented. Given standard assumptions on human motivation, the subgame perfect equilibrium prescribes that the proposer offers the smallest monetary unit (or even nothing) to the responder, and the latter accepts because 'little is better than nothing'. The experimental outcomes across hundreds of replications of the ultimatum game are however significantly different from the theoretical prediction. Responders reject with probability .4 to .6 proposals offering less than 20 percent of the available surplus, and the probability of rejection decreases as the size of the offer increases. About 70 percent of proposers offer between 40 and 50 percent of the available surplus.[6]

Early experimenters like the recent Nobel Prize winner Vernon Smith conducted market experiments and found that experimental markets converge quickly to the competitive equilibrium. These results have been interpreted as confirmation of the self-interest hypothesis. However, recently developed models of *inequity aversion* by Fehr

[6]See, for example, Fehr and Schmidt (2003) for an overview of the experimental evidence.

and Schmidt (1999) and Bolton and Ockenfels (2000) show that in market environments the standard competitive equilibrium may prevail even if agents are strongly concerned about fairness. The theory of inequity aversion assumes that some but not all agents suffer a utility loss if their own material payoffs differ from the payoffs of other agents in their reference groups. It can be shown that the interaction of the distribution of types (fair or selfish) with the strategic environment (market or non-market interaction) can explain why in some situations very unequal outcome may prevail (competitive equilibrium) while in others rather equitable outcomes are obtained (ultimatum game). The reason is that in environments with few players only, a fair agent may be able to enforce an equitable outcome while this is not possible in a market setting. The virtue of theories of inequity aversion is that they can – in contrast to other models of *social preferences* like *altruism* or *envy* – account for a large number of seemingly different results in wide array of experimental settings.

The model in Chapter 4 goes a step further and applies the theory of inequity aversion as formulated by Fehr and Schmidt (1999) to contract theory. We analyze how fairness concerns affect incentive provision in an otherwise standard moral hazard model with two risk-averse agents. In a classic contribution to the theory of incentives Holmström and Milgrom (1991, p. 24) state that "it remains a puzzle for this theory that employment contracts so often specify fixed wages and more generally that incentives within firms appear to be so muted, especially compared to those of the market." In Chapter 4 we show that inequity aversion can serve as an explanation for the scarcity of incentive contracts within firms.

In the standard principal-agent moral hazard model the optimal contract trades off incentive provision and agent's risk bearing. Effort choices cannot be contracted upon such that wages must condition on stochastic output realizations. In optimum the agent bears some risk for which he must be compensated. This constitutes the agency costs of providing incentives. We find that behindness aversion (an agent incurs a utility loss only when being worse off than the other agent) among agents unambiguously increases agency costs. This holds true if agents also suffer from being better off, unless they account for effort costs when comparing to the other agent. Fairness concerns increase agency costs because they impose an additional restriction on the design of optimal contracts. Inequity aversion effects an utility loss in states of the world with diverging output- and thus diverging wage realizations. The resulting, reduced utility levels could be generated without inequity aversion as well – simply by lowering the respective wage. Since these lower utility levels (at lower wage costs) were not optimal without inequity aversion, they cannot be optimal now.

We find that increased agency costs can undermine efficiency in two ways. First, inequity aversion may render equitable flat wage contracts optimal even though incentive contracts are optimal with selfish agents. Empirical studies (see, for example, Bewley (1999)) suggest that organizations likes firms are characterized by a dense network of social relations. Market interactions are, in contrast, rather anonymous. Taking this into account, our first main result offers an explanation of why incentives within firms are muted as compared to those in the marketplace. Second, to avoid social comparisons the principal may employ one agent only, thereby forgoing the efficient effort provision of the second agent. We call this the 'reference group effect'. This second result has implications for the internal organization and the boundary of the firm. Suppose the principal can set up different firms but doing so involves fixed costs. The principal now faces a trade-off. On the one hand, 'integrating' several agents within a single firm causes social comparisons and thus increased agency costs of providing incentives. On the other hand, 'separating' agents into different firms involves additional fixed costs. The solution to this trade-off thus defines an optimal size of the firm.

Chapter 1

IPO Pricing and Informational Efficiency: The Role of Aftermarket Short Covering*

1.1 Introduction

Since the Securities Act of 1934, it is legal practice in the U.S. that offering syndicates stabilize market prices of their recent public offerings. In their latest release the U.S. Securities and Exchange Commission (SEC) states: "Although stabilization is a price influencing activity intended to induce others to purchase the offered security, when appropriately regulated it is an effective mechanism for fostering the orderly distribution of securities and promotes the interests of shareholders, underwriters, and issuers."[8] With this paper we challenge the assertion that current regulation always serves the interests of all involved parties. We argue that issuing investment banks *can* combine two regulated stabilization tools to generate risk-free profits. Employing a model that captures the impact of this arbitrage opportunity on the offer price, we find that (a) either market transparency is lower or, on average, underpricing is exacerbated, and (b) the issuing investment bank's profits are boosted at the expense of issuer and investors.

*The chapter is based on joint work with Andreas Park from the University of Toronto.

[8] SEC (1997), Regulation M, Release No. 34-38067, p. 81. The Committee of European Securities Regulators (CESR (2002)) proposes rules which resemble the SEC regulations.

Current regulation allows investment banks to pursue the following three types of aftermarket activities. First, *stabilizing bids* can be posted at or below the offer price during the distribution period of the securities. Second, banks can establish a short position by selling securities in excess of the pre-announced amount. *Aftermarket short covering* refers to the practice of filling these positions in the aftermarket, which is done if the market price falls below the offer price. If the price instead rises, the bank is hedged by an *overallotment option* which grants the right to obtain typically up to 15% additional securities from the issuer at the offer price. Third, *penalty bids* are used to penalize customers who immediately resell their securities in the aftermarket.

Although on average IPOs have high first-day returns, there is a significant number of IPOs with negative returns. In these 'cold' IPOs, stabilizing bids and short covering should ensure liquidity for the security to offset potential selling pressure (in the first days after the float), and thus prevent sharp drops in prices. Penalty bids are meant to reduce selling pressure. In this paper we focus exclusively on the impact of short covering. An investment bank intending to support the security price adheres to the following procedure. It enters the aftermarket short. This position must be filled eventually. Suppose that the market price exceeds the offer price. Then there is supposedly no selling pressure and no need to provide extra liquidity. Covering the short position in the market, however, would be expensive. This is why almost all IPO contracts include a so-called 'Greenshoe' or overallotment option. It allows the bank to buy extra securities from the issuer at the offer price. In the bulk of offerings, the initial short position is perfectly hedged by this option. Increasing prices are therefore no risk for the bank. Suppose now that the price drops. The bank does provide liquidity, however, by doing so it also covers the short position in the market – at a price below the offer price. The difference between the market price and the offer price (minus the gross spread) is pure profit. In other words, the opportunity to enter the market short, paired with the overallotment option, provides investment banks with a second, risk-free potential source of income.

Only recently, new data became available that allowed to analyze investment banks' activities in the aftermarket directly. Aggarwal (2000) reports that underwriters utilize a combination of aftermarket short covering, penalty bids, and exercise of the overallotment option. Stabilizing bids are never observed. Ellis, Michaely, and O'Hara (2000) report that the lead underwriter always becomes the dominant market maker. They also find that market makers take large inventory positions, but reduce their risk by exercising the overallotment option.

There are two cases studies which support the casual observation that aftermarket trading can be very profitable. Jenkinson and Ljungqvist (2001) provide a study of the 1995 GenCo2 IPO (U.K.) during which the price fell in the aftermarket. The assigned investment banks Barclays de Zoete Wedd and Kleinwort Benson repurchased 45.7 million securities at the low market price to cover their short positions that were established at the offer price. Jenkinson and Ljungqvist (2001) conclude: "It demonstrates how valuable the over-allotment option potentially is to the syndicate of investment banks selling the issue. Since they will buy back the shares in the market only if the price is below the issue price, in closing (partially or in full) their short position they make profits. These profits accrue to the syndicate itself, as the holder of the option, rather than to the [...] vendors" (p. 180). Boehmer and Fishe (2001) analyze a case-study of an IPO in which the lead underwriter took a nearly perfectly hedged short position which was then covered in the aftermarket. The profits from trading amounted to 52% of the syndicate's overall profit from the offering. In their words: "[...][short covering activities] represent an economically significant profit opportunity for the Lead" (p. 4).[9]

The existing literature on the impact of price support on offer prices models stabi-

[9] Aggarwal (2000) finds that "short covering is not expensive for underwriters" (p. 1077). In more detail, she finds that for weak offerings investment banks make profits, for strong offerings however they may lose money. This stems from the fact that either the overallotment option is not fully exercised or investment banks had established a "naked short" prior to the offering such that short positions had to be covered at prices above the offer price. Our model cannot explain why investment banks sometimes establish "naked shorts" or do not fully exercise the overallotment option. We merely analyze the effects aftermarket short covering can have when investment banks utilize the possibility to make risk-free profits, i.e. when they do not establish "naked shorts" and fully exercise the overallotment option when prices rise. Ellis, Michaely, and O'Hara (2000) find that aftermarket activities of the lead underwriter are profitable and account for about 23% of the overall profit of underwriting. Reported profits stem from both market making and stabilizing activities (that is accumulating inventory positions). From the presentation in the paper it does not seem possible to disentangle whether stabilization contributed to or reduced trading profits. The claim that stabilization can be a profitable activity is thus not rejected by the data.

lization to be costly. The two seminal theoretical papers on stabilization, Benveniste, Busaba, and Wilhelm Jr. (1996) and Chowdhry and Nanda (1996), assume that banks post stabilizing bids to keep prices up. However, such stabilizing bids are never observed. Both models imply that stabilizing activities decrease underpricing – our model predicts the opposite. This paper thus contrasts the existing literature as we model explicitly that investment banks can earn money in the aftermarket, and to the best of our knowledge we are the first to do so in a theoretical framework.

We propose a stylized model of an offering procedure that is in accordance with empirical findings and perceived industry practice. We assume that both the investment bank and investors hold private information about the intrinsic value of the offered security. We assume this information asymmetry to arise at a point in time when all official, mandatory information has been released. Thus, any further public statement by bank or issuer will be perceived as cheap talk, and it is only actions, i.e. price-setting, that can convey additional information. We model the procedure as a signaling game in which the investment bank moves first and sets the offer price. It chooses the offer price strategically to maximize its profits form both the gross spread of the offer revenue and trading profits in the aftermarket. The bank anticipates investors' best replies to the offer price.

As a benchmark, we first analyze a setting without aftermarket activities and identify the conditions for the equilibrium to be both unique and separating (that is, a bank with different information sets different prices). We call a separating equilibrium *informationally efficient* since the bank's information is fully revealed by the offer price. After the offer is floated, prices adjust according to market demand. In equilibrium, the security can turn out to be either under- or overpriced, but investors account for this when ordering the security. We show that on average there is underpricing.

When introducing aftermarket short covering, relative to the benchmark one of two outcomes transpires: either the offer price falls on average, or separation breaks down and the offer-price equilibrium morphs into a pooling equilibrium. In the first case, an investment bank with favorable information distorts the price downwards and thereby, on average, *exacerbates underpricing*. In the second case investors are unable to infer the investment bank's signal from the offer price. This equilibrium is *informationally inefficient* since investors' decisions are based on private signals only and not also on the signal of the bank. A major objective of financial market regulation is market transparency. Without modelling an explicit payoff from higher transparency we simply assume that it is desirable if prices contain more rather than less information. Consequently, pooling equilibria are undesirable.

Furthermore, the price distortion leads to redistribution of wealth in favor of the investment bank. Looking at per-share profits, the issuer loses if separation prevails; in a pooling equilibrium he is better off. The issuer's losses are the investors' gains and vice versa. On the comparative statics side, an increase of the gross spread or the amount of overalloted securities reduces the parameter-set with informational efficiency.

The remainder of the paper is organized as follows. In Section 1.2 we introduce our model of the offering procedure without aftermarket short covering and identify necessary and sufficient conditions under which the investment bank reveals its private signals through separating offer prices. In Section 1.3 we introduce aftermarket short covering, identify the conditions under which the investment bank pools in the offer price and thus holds back its private information and show that, if separation is upheld, prices fall on average. We also provide results on comparative statics. In Section 1.4 we discuss the redistribution of profits. Section 1.5 concludes. Proofs and specifications of tools used in the equilibrium analysis are in the Appendix.

1.2 The Benchmark: Offer Prices in a Model without Aftermarket Short Covering

1.2.1 The Model Ingredients and Agents' Best Replies

Consider the following stylized model of the IPO process.

The Security. The security on offer can take values $V \in \mathbb{V} = \{0, 1\}$, both equally likely. The number of securities is denoted by S.

The Investors. There are N identical, risk neutral investors. N is assumed to be strictly larger than S. They can either order one unit of the security or none. Each investor receives a costless, private, conditionally i.i.d. signal $s_i \in \mathbb{V}$ about the value of the security. This information is noisy, i.e. $\mathsf{Pr}(s_i = v|V = v) = q_i$ with $q_i \in (\frac{1}{2}, 1)$. If an investor orders, he may or may not obtain the security during the offering procedure; if the issue is oversubscribed shares are distributed with uniform probability. If he does, his payoff is the market price minus the offer price. If the offer is not floated, his payoff is zero even if he ordered the security. An investor's type is his signal. We refer to the investor as a 'high-signal investor' if $s_i = 1$. For $s_i = 0$, it is a 'low-signal investor'.

The Issuer. We assume that the issuer has no strategic impact. He holds no private information about the value of the security. The issuer signs a contract with an investment bank that delegates the pricing decision and constitutes the amount of securities S to be sold.[10] It also specifies the gross spread β of the offer revenue that remains as remuneration at the bank. The issuer's payoff is thus fraction $(1 - \beta)$ of the offer revenue if the offer is floated, otherwise it is zero.

The Investment Bank. The risk neutral investment bank who signed the contract with the issuer receives a private signal $s_b \in \mathbb{V}$ about the value of the security. This signal is noisy and conditionally independent from investors' signals. Yet it is more informative, i.e. $q_b > q_i$, where $\mathsf{Pr}(s_b = v|V = v) = q_b$. Signals characterize a bank's type. If $s_b = 1$ we refer to the investment bank as a 'high-signal bank'. For $s_b = 0$, it is a 'low-signal bank'. The bank receives the signal after the contract has been signed and then announces the offer price p.[11] If demand is too weak to match supply, i.e. if the number of investors willing to buy is less than the number of securities to be sold, we assume that the offer is called off.[12] In case of excess demand securities are

[10] The two most widely used contracts between issuers and investment banks are firm commitment and best efforts contracts. These contracts differ with respect to risk allocation and incentive provision that may be necessary due to imperfectly observable distribution effort and asymmetric information about the value of the securities. However, in this stylized model we abstract form these complications.

[11] We discuss fixed-price offerings vs. bookbuilding at the end of this subsection.

[12] Busaba, Benveniste, and Guo (2001) report for a sample of 2,510 IPOs filed with the SEC from 1984 to 1994 that 14.3% of the offerings got called off. Issuers have the option to withdraw an offer

allocated at random. We assume that failure of the offering inflicts fixed costs C on the investment bank.[13] These costs are external to our formulation and can be thought of as deterioration of reputational capital. They may also capture the opportunity costs resulting from lost market share when being associated with an unsuccessful IPO.[14] Without loss of generality, we do not specify any costs the offering procedure itself may cause for the investment bank. Thus, if the offer is successful, the bank's payoff is $\beta \mathsf{pS}$; if it fails, its payoff is $-C$.

Signaling Value of the Offer Price. An investor bases his decision on his private information and on the information that the investment bank reveals about its own signal through the offer price. We denote this information by $\mu(\mathsf{p})$ and write $\mu(\mathsf{p}) = 1$ if the price reflects that the bank's signal is $s_b = 1$, $\mu(\mathsf{p}) = 0$ if the price reflects that $s_b = 0$, and $\mu(\mathsf{p}) = \frac{1}{2}$ to indicate that the price is uninformative. These three are the only relevant cases in equilibrium. We refer to μ as the *price-information* about the bank's signal.

The Aftermarket Price. The equilibrium market price is determined by the aggregate number of investors' favorable signals. In our model this number is always revealed, either directly through investor demand or immediately after the float through trading activities. Thus write $\mathsf{p}^m(d)$ for the market price as a function of $d \in \{0, \dots, N\}$, the number of high-signal investors. Appendix 1.6.1 fleshes out this argument and provides an extensive treatment of price formation.

Investors' Decisions and Expected Payoffs. We admit only symmetric, pure strategies; thus all investors with the same signal take identical decisions. These can

if the investment bank proposes a price that is perceived as too low. During the road show the bank learns about investors' valuations. In a firm commitment contract the bank uses this information to propose an offer price such that it can find enough investors to sell the entire offer; in a best efforts contract, such that selling all securities will not be too difficult. This model abstracts from the issuer's option to withdraw, and it leaves no room to the bank to adjust the offer price to investors' valuations.

[13] The model could be extended to allow the bank to buy up unsold securities. Costs then result from expensively bought inventory positions and not from failure. C would thus be 'smoothed'. This would, however, not alter our qualitative results but complicate the analysis considerably.

[14] Dunbar (2000), for instance, provides evidence that established investment banks lose market share when being associated with withdrawn offerings.

then be aggregated so that only three cases need to be considered. First, all investors buy, denoted $B_{0,1}$, second, only high-signal investors subscribe, denoted B_1, and third, no investor buys, denoted $B_\emptyset$. Thus, the set of potential collective best replies is $\mathbb{B} := \{B_{0,1}, B_1, B_\emptyset\}$.

To compute his expected payoff, an investors has to account for the probability of actually getting the security. There are two cases to consider. In the first, all investors buy. Thus, market demand is N and all investors receive the security with equal probability S/N. In the second case, only high-signal investors buy. If $d-1$ others buy, then an investor receives the security with probability $\mathsf{S}/(d)$. If overall demand d is smaller than the number of shares on offer, $d < \mathsf{S}$ the IPO fails and the investor who ordered gets it with probability 0.

Investors order the security whenever their expected payoff from doing so is non-negative. Suppose only high-signal investors buy, B_1. After observing the offer price, an investor's information set contains both his signal s_i and the information inferred from the offer price, $\mu(\mathsf{p})$. Since signals are conditionally i.i.d., for every $V \in \mathbb{V}$ there is a different distribution over the number of favorable signals ($s_i = 1$), which we denote $f(d|V)$. The investors' posterior distribution over demands is given by

$$\begin{aligned} g(d-1|s_i, \mu(\mathsf{p})) := \mathsf{Pr}(V = s_i|s_i, \mu(\mathsf{p})) \cdot f(d-1|V = s_i) \\ +\mathsf{Pr}(V \neq s_i|s_i, \mu(\mathsf{p})) \cdot f(d-1|V \neq s_i). \end{aligned} \tag{1.1}$$

Then for a high-signal investor, at price p his rational-expectation payoff from buying has to be non-negative,

$$\sum_{d=\mathsf{S}}^{N} \frac{\mathsf{S}}{d} \cdot (\mathsf{p}^m(d) - \mathsf{p}) \cdot g(d-1|s_i = 1, \mu(\mathsf{p})) \geq 0. \tag{1.2}$$

Likewise for $B_{0,1}$, in which case the summation runs from 1 to N, S/d is substituted with S/N, and $s_i = 1$ is replaced by $s_i = 0$.

Threshold Prices. Denote by $p_{s_i,\mu}$ the highest price that an investor is willing to pay in equilibrium if all investors with signal $\tilde{s}_i \geq s_i$ order, given signal s_i and price-information μ. Thus $p_{1,1}$ is the highest (separating) price with B_1, $p_{1,\frac{1}{2}}$ the highest (pooling) price with B_1, $p_{0,\frac{1}{2}}$ the highest (pooling) price with $B_{0,1}$, and $p_{0,0}$ the highest (separating) price with $B_{0,1}$. Note that at all these prices investors are aware that the security price may drop in the aftermarket and that they may not get the security. The threshold prices are formally derived in Appendix 1.6.2.

The Investment Bank's Expected Payoff. First consider case B_1. Variable d denotes the number of buys, i.e. the number of high-signal investors. If the true value is $V = 1$, we have

$$\mathsf{Pr}(d \geq \mathsf{S}|B_1) = \sum_{d=\mathsf{S}}^{N} \binom{N}{d} q_i^d (1-q_i)^{N-d}, \tag{1.3}$$

analogously for $V = 0$. A bank with signal s_b assigns probability $\alpha_{s_b}(\mathsf{S})$ to the event that at least S investors have the favorable signal. Since the investment bank receives its signal with quality q_b, for $s_b = 1$,

$$\alpha_1(\mathsf{S}) = q_b \cdot \sum_{d=\mathsf{S}}^{N} \binom{N}{d} q_i^d (1-q_i)^{N-d} + (1-q_b) \cdot \sum_{d=\mathsf{S}}^{N} \binom{N}{d} (1-q_i)^d q_i^{N-d}. \tag{1.4}$$

$\alpha_0(\mathsf{S})$ is defined analogously. If the bank charges a price at which only high-signal investors buy, its expected profit is

$$\Pi(\mathsf{p}|s_b, B_1) = \alpha_{s_b}(\mathsf{S}) \cdot \beta\mathsf{p}\mathsf{S} - (1-\alpha_{s_b}(\mathsf{S})) \cdot C. \tag{1.5}$$

Consider now $B_{0,1}$, the case where the offer price is low enough so that all investors are willing to buy, irrespective of their signals. The offer never fails, thus payoffs are given by $\Pi(\mathsf{p}|B_{0,1}) = \beta\mathsf{p}\mathsf{S}$. If the price is set so high that no investor buys, as in case $B_\emptyset$, a loss of C results with certainty.

Simplifying Assumptions. The unconditional distribution over favorable signals is a composite of the two conditional distribution and thus bimodal To obtain closed form solutions (or rather approximations) for success-probabilities and prices, we make two simplifying assumptions: the first simplifies computations, since the two modes of the distribution over favorable signals are centered around $N(1-q_i)$ and Nq_i. The results of the paper will also hold if it was not satisfied, as long as $\mathsf{S} < N/2$, but the assumption allows us to get closed form solutions for success-probabilities. The second assumption ensures that we can analyze the two underlying conditional distributions separately.

Assumption 1.1 $\mathsf{S} = (1-q_i)N$.

For every signal quality q_i, there exists an $\bar{N}(q_i)$ so that for all $N > \bar{N}(q_i)$ the two conditional distributions over favorable signals generated by $V = 0$ and $V = 1$ do not 'overlap.'[15] By standard results from statistics, sufficient for $\bar{N}(q_i)$ is $\bar{N}(q_i) > 64q_i(1-q_i)/(2q_i-1)^2$.

Assumption 1.2 *The number of investors N is larger than $\bar{N}(q_i)$.*

As a consequence of the second assumption we can apply the Law of Large Numbers and DeMoivre-Laplace's Theorem.[16] Since we assume that the IPO fails whenever $d < \mathsf{S}$, Assumption 1.1 implies $\alpha_0(\mathsf{S}) = (2-q_b)/2$ and $\alpha_1 = (1+q_b)/2$; in what follows we thus omit S. A consequence of the Law of Large Numbers is that $\mathsf{p}^m(d) \in \{0,1\}$ for almost all values of d.[17]

Fixed Price Offerings vs. Bookbuilding. On most stock exchanges in the world IPOs are sold through bookbuilding (for instance in the US, the UK, Germany, but not in France), whereas our model is a fixed-price offering. Current regulation allows risk-free aftermarket short covering profits and this paper tries to capture their strategic

[15]To be more precise: We need to ensure that if $V = 1$, the probability of demand $d < \mathsf{S}$ is zero.

[16]For instance, the mode of a binomial distribution is generally not exactly symmetric. However, if N is large enough, we can apply DeMoivre-LaPlace $(0 < q_i \pm 2\sqrt{q_i(1-q_i)/N} < 1)$ and employ the normal distribution instead. Thus we can treat each mode to be symmetric. The number traders has to large enough so that for $V = 0$, there are almost never more than $N/2$ traders with a favorable signal and vice versa for $V = 1$.

[17]To be more precise, for $d \gg N/2$, $\mathsf{p}^m(d) = 1$, and for $d \ll N/2$, $\mathsf{p}^m(d) = 0$. Thus to get interesting equilibria, it is necessary that S is strictly smaller than $N/2$. If it was not, an IPO where only $s_i = 1$ investors buy, would never be at risk of being overpriced as it fails in all overpriced cases.

impact. These potential profits depend primarily on price movements and thus one should study the the offer price as the strategic decision variable. In any imaginable framework the investment bank faces a trade-off between higher revenue and likelihood of failure. Thus it is reasonable to assert that the offer price or, depending on the formulation, the bookbuilding span has signaling value. A fixed-price mechanism is, arguably, the simplest possible way to capture the price's strategic dimension.

A hypothetical bookbuilding model will capture the strategic dimension in a similar fashion, yet the analysis would become less tractable without adding insight: In bookbuilding, the investment bank must set a bookbuilding span. This span can certainly have signaling value because it is, arguably, similar to setting a single price (a degenerate span). Suppose bookbuilding spans have to be sufficiently tight so that they are strictly in the $[0, 1]$-interval's interior. During the bookbuilding period, investors submit their orders which (potentially) reveal their private information – just as with our fixed price mechanism. At the end of the bookbuilding period the investment bank will set the final selling price somewhere in the span, distribute the shares, and reveal overall demand. As long as the span and thus the issue price in the span is strictly in the interior of the $[0, 1]$-interval, secondary market prices will adjust to a price outside the span. Our stylized, parsimonious model is rich enough to capture the same result that a more complicated bookbuilding model would yield.

1.2.2 Derivation of the Separating Equilibrium

The focus of this paper is the pricing decision of the investment bank given its signal. In the following we identify the conditions under which a profit maximizing investment bank will reveal its information through the offer price. A separating equilibrium is defined as *informationally efficient* since investors can derive the bank's signal from the offer price. In a pooling equilibrium information is shaded and thus it is *informationally inefficient*. In this case, investors decide only on the basis of their private signals.

The Equilibrium Concept and Selection Criteria. The equilibrium concept for this signaling game is, naturally, the Perfect Bayesian Equilibrium (PBE). A common problem with PBEs, however, is their multiplicity, stemming equilibria being supported by "unreasonable" out-of-equilibrium beliefs. The common way to overcome this problem is to apply an equilibrium selection rule such as the *Intuitive Criterion* (IC), introduced by Cho and Kreps (1987). We follow this line of research and consider only equilibria that do not fail the IC. All of these PBE selection devices favour separating over pooling equilibria. It will turn out, however, that in our framework under certain conditions the IC cannot rule out pooling price equilibria. Moreover, from the perspective of the investment bank the pooling equilibrium then Pareto dominates any separating equilibrium. It would thus be unreasonable not to assume that these equilibria will be picked. Thus in what follows, we will only consider equilibria that *satisfy the IC* and among these, we consider those that are *Pareto efficient* for the bank

A *pooling equilibrium* is specified through (i) an equilibrium offer price p^* from which investors infer (ii) price-information $\mu = \frac{1}{2}$, and (iii) investors' best replies given their private signals, μ, and p^*. A *separating equilibrium* is (i) a system of prices $\{\underline{\mathsf{p}}^*, \bar{\mathsf{p}}^*\}$ and price-information such that (ii) at $\mathsf{p}^* = \bar{\mathsf{p}}^*$, the *high separation price*, the price-information is that the bank has the favorable signal, $\mu = 1$, at $\mathsf{p}^* = \underline{\mathsf{p}}^*$, the *low separation price*, the price-information is that the bank has the low signal, $\mu = 0$, and (iii) investors' best replies given their private signals, μ, and p^*. In both separating and pooling equilibria, for $\mathsf{p} \notin \{\bar{\mathsf{p}}^*, \underline{\mathsf{p}}^*\}$ out-of-equilibrium public beliefs are chosen 'appropriately.' The following result is a straightforward consequence of signaling, the proof of which is in Appendix 1.6.5.

Lemma 1.1 [The Highest Possible Low Separating Price] *There exists no separating offer price* $\underline{\mathsf{p}}^* > p_{0,0}$.

In any separating equilibrium, therefore, the low price must be such that all investors buy, and the highest such separating price, given price-information $\mu = 0$, is $\underline{\mathsf{p}}^* = p_{0,0}$. In what follows we refer to $p_{0,0}$ as *the* low separation price.

Signaling equilibria in our setting come in one of three guises: The already mentioned separating equilibrium, a pooling equilibrium in which only high-signal investors buy, and a pooling equilibrium in which all investors buy. In the following, we characterize the conditions guaranteeing that only separating equilibria survive our selection criterion.

Fix a potential price $\mathsf{p} \in [p_{0,0}, p_{0,\frac{1}{2}}]$, the interval of potential pooling prices at which all investors would buy. Define $\phi_1(\mathsf{p})$ as the price at which the high-signal bank would be indifferent between charging a *risky* price $\phi_1(\mathsf{p})$ at which only high-signal investors buy, B_1, and a *safe* pooling price p with $B_{0,1}$ (all investors buy). Formally,

$$\alpha_1\beta\phi_1(\mathsf{p})\mathsf{S} - (1-\alpha_1)C = \beta\mathsf{p}\mathsf{S} \quad \Leftrightarrow \quad \phi_1(\mathsf{p}) = \frac{\mathsf{p}}{\alpha_1} + \frac{1-\alpha_1}{\alpha_1}\frac{C}{\beta\mathsf{S}}. \tag{1.6}$$

Price $\phi_0(\mathsf{p})$ is defined analogously for the low-signal bank. Thus price $\phi_{s_b}(\mathsf{p})$ is the lowest risky price that a bank with signal s_b is willing to deviate to from safe price p.[18] In what follows we refer to $\phi_1(\mathsf{p})$ as the *high-signal bank's deviation price*, and to $\phi_0(\mathsf{p})$ as the *low-signal bank's deviation price*. It is straightforward to see that $\phi_0(\mathsf{p}) > \phi_1(\mathsf{p})$ for all $\mathsf{p} \in [p_{0,0}, p_{0,\frac{1}{2}}]$, that is, the low-signal bank requires a higher price as compensation for risk taking. In addition, $\partial\phi_j(\mathsf{p})/\partial\mathsf{p} > 0$, $j \in \{0,1\}$, so the higher the pooling price, the higher the lowest profitable deviation price. We can now establish our first major result.

Proposition 1.1 (Conditions for Informationally Efficient Prices)
If (i) the high-signal bank's deviation price from the highest safe pooling price is not higher than the highest separating price, $\phi_1(p_{0,\frac{1}{2}}) \leq p_{1,1}$, and if (ii) the low-signal bank's deviation price from the low separating price is not smaller than the highest risky pooling price, $\phi_0(p_{0,0}) \geq p_{1,\frac{1}{2}}$ then there exists a unique PBE that satisfies the Intuitive Criterion and it is the separating equilibrium $\{(\underline{\mathsf{p}}^* = p_{0,0}, \mu = 0, B_{0,1}); (\bar{\mathsf{p}}^* = \min\{p_{1,1}, \phi_0(p_{0,0})\},\ \mu = 1, B_1); (\mathsf{p} \neq \{\underline{p}^*, \bar{p}^*\}, \mu = 0, B_{0,1}$ if $\mathsf{p} \leq p_{0,0}, B_1$ if $p_{0,0} < \mathsf{p} \leq p_{1,0}, B_\emptyset$ else$)\}$.

Interpretation of the Proposition. The first condition, $\phi_1(p_{0,\frac{1}{2}}) \leq p_{1,1}$, together with the IC is necessary and sufficient to rule out pooling equilibria in which all in-

[18]Deviation to a high, risky price can lead to increased overpricing, which is commonly perceived to be bad for a bank's reputation. Nanda and Yun (1997) analyze the impact of IPO mispricing on the market value of investment banks. They find that overpriced offerings result in decreased lead-underwriter market value. In our model, however, investors fully take into account that the offer price may drop in the aftermarket. Modelling such reputation effects would thus be contradictory in our setting.

vestors buy, irrespective of their signals. The second condition, $\phi_0(p_{0,0}) \geq p_{1,\frac{1}{2}}$, ensures that there is no pooling where only high-signal investors buy, B_1. The IC itself ensures that the bank with $s_b = 1$ always charges the highest sustainable separating price. The high separation price $\bar{\mathsf{p}}^*$ is the minimum of $p_{1,1}$ and $\phi_0(p_{0,0})$. The bank cannot charge more than $p_{1,1}$, and it cannot credibly charge more than $\phi_0(p_{0,0})$ as otherwise the bank with $s_b = 0$ would deviate. Finally, since $\phi_1(p_{0,0}) < \phi_1(p_{0,\frac{1}{2}}) \leq p_{1,1}$, the bank with $s_b = 1$ is willing to separate. The proof's details are in Appendix 1.6.5. A definition of the IC can be found, for instance, in Fudenberg and Tirole (1991)[p.448].

Underpricing. In the context of this model the first-day return is the difference between market price and offer price. We can establish the following proposition. The proof is in Appendix 1.6.5.

Proposition 1.2 (Underpricing)
In a separating equilibrium, on average, securities are underpriced.

Interpretation of the Result. The intuition behind the result is clear: Both types of investors only buy if their expected payoff is non-negative. At $p_{0,0}$ the low-signal investor just breaks even in expectation but the high-signal investor expects a strictly positive payoff. At $p_{1,1}$ the high-signal investor just breaks even and the low-signal investor abstains. Thus, ex-ante, expected payoff is positive, i.e. there is underpricing.

1.2.3 An Intuitive Characterization of the Equilibrium

The concept of deviation prices ϕ_{s_b} is a convenient tool to describe restrictions. We will now reformulate the conditions from Proposition 1.1 in terms of exogenous costs C. This allows us to derive a simple linear descriptive characterization of the equilibrium. Consider first condition (i), $\phi_1(p_{0,\frac{1}{2}}) \leq p_{1,1}$. If C is so high that

$$\phi_1(p_{0,\frac{1}{2}}) = \frac{p_{0,\frac{1}{2}}}{\alpha_1} + \frac{1-\alpha_1}{\alpha_1}\frac{C}{\beta \mathsf{S}} > p_{1,1} \tag{1.7}$$

then a separating equilibrium cannot be sustained. Even a high-signal bank then prefers to sell the security at a price where all investors buy.

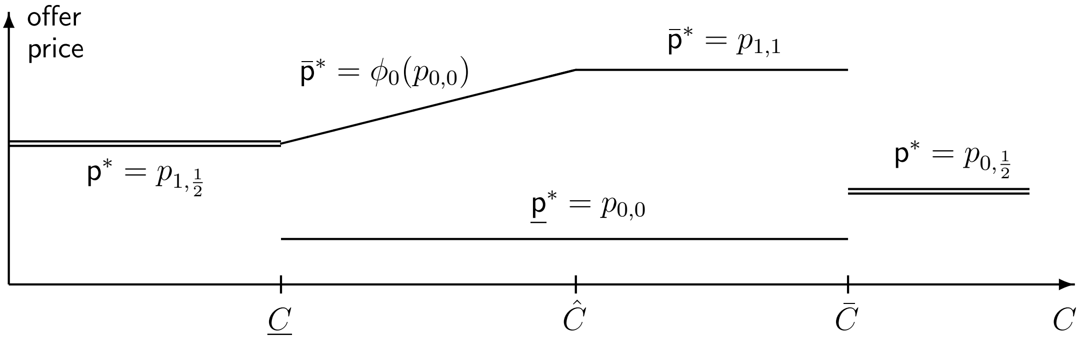

Figure 1.1: **Threshold Costs and Equilibrium Prices.** For costs smaller than $\underline{C}$, it holds that $\phi_0(p_{0,0}) < p_{1,1/2}$ so that the low-signal bank chooses a risky price. A pooling equilibrium in $p_{1,1/2}$ results. If $C \in (\underline{C}, \bar{C})$ a separating equilibrium results. For $C \in (\underline{C}, \hat{C})$ the high-signal bank cannot charge the highest separation price $p_{1,1}$ but must set a lower price $\phi_0(p_{0,0})$ to prevent the low-signal bank from mimicking. For $C \in [\hat{C}, \bar{C})$ the high signal bank can charge $\bar{\mathsf{p}}^* = p_{1,1}$. Finally, if $C > \bar{C}$ it holds that $\phi_1(p_{0,1/2}) > p_{1,1}$ so that even the high-type bank prefers a safe price and pooling in $p_{0,\frac{1}{2}}$ results.

Consider now condition (ii), $\phi_0(p_{0,0}) \geq p_{1,\frac{1}{2}}$. If C is so low that

$$\phi_0(p_{0,0}) = \frac{p_{0,0}}{\alpha_0} + \frac{1-\alpha_0}{\alpha_0}\frac{C}{\beta \mathsf{S}} < p_{1,\frac{1}{2}} \tag{1.8}$$

then a separating equilibrium, again, cannot be sustained (by the SIC). In this case, even a low-signal bank is willing to choose a high, risky pooling price and the high-signal bank can thus not credibly signal its information. If C is so high that $\phi_0(p_{0,0}) > p_{1,1}$ then for the low-signal bank it does not even pay to deviate to the highest separating price, $p_{1,1}$. This bound on C is given by

$$\hat{C} := \frac{\alpha_0 p_{1,1} - p_{0,0}}{1-\alpha_0}\,\beta \mathsf{S}. \tag{1.9}$$

Define, analogously, $\bar{C}$ and $\underline{C}$ such that (1.7) and (1.8) hold with equality. We get $\underline{C} < \hat{C} < \bar{C}$. The following Corollary to Proposition 1.1 summarizes the above characterization.

Corollary 1.1 (Proposition 1.1 in Terms of Costs)
If $C \in (\underline{C}, \bar{C})$ then the unique equilibrium is the separating equilibrium stated in Proposition 1.1. If $C \in (\underline{C}, \hat{C})$ then $\bar{\mathsf{p}}^* = \phi_0(p_{0,0})$, and if $C \in [\hat{C}, \bar{C})$ then $\bar{\mathsf{p}}^* = p_{1,1}$.

It has often been argued that certifying agents, here the investment bank, must have ‘enough’ reputational capital at stake to make certification credible. In this context, also ‘too much’ reputation can inhibit certification (separation from a low-signal bank) if it becomes to expensive to jeopardize one’s reputation at a high, risky offer price. Figure 1.1 plots threshold costs and corresponding equilibrium prices.

1.3 The Impact of Aftermarket Short Covering

In this section we extend the model and allow the investment bank to pursue aftermarket short covering. We analyze its effect on the investment bank’s pricing decision and investigate under which conditions informational efficiency will be undermined. We find that, in general, the conditions for a separating equilibrium become more restrictive. Upholding separation may come at a cost – thus on average the investment bank has to distort prices down, which causes more underpricing.

1.3.1 Overview of Short Covering and a Bank’s Strategy

With aftermarket short covering the investment bank has the opportunity to allot a predetermined amount of up to O securities on top of the principal volume of securities S. This amount O is referred to as the overallotment facility. It typically constitutes 15% of the number of initial securities S. The investment bank goes short in a position of this size. If the market price falls below the offer price, the bank fills its short positions in the aftermarket. This practise is referred to as aftermarket short covering. If the price is below the offer price, the bank makes a profit. If the market price rises above the offer price, the bank exercises a so-called overallotment option, the right to obtain up to O securities from the the issuer at the offer price. The option is only valid if the bank had indeed established a short position. Consequently, the bank is perfectly hedged against rising prices. We restrict attention to the case where either the entire amount of $\mathsf{S} + \mathsf{O}$ securities is sold or, if only fewer securities can be sold, the IPO fails; the restriction merely simplifies the analysis and does not affect the qualitative results. The bank receives the gross spread only on the securities that actually remain floated.

Intuitively, the size of a potential price drop and thus of profits from aftermarket short covering is larger the higher the offer price. In the benchmark case's separating equilibrium, a low-signal bank would not mimic a high-signal bank because it fears costs from a potential IPO failure. With short covering expected aftermarket profits are higher the larger the potential price drop. Moreover, a bank with a low signal considers such a drop more likely. It is then possible, that potential losses from a failed offering are offset by higher expected aftermarket gains. Two scenarios are possible: In equilibrium, the high-signal bank sets a *lower* price to separate from a low-signal bank. The high-signal bank, however, is only willing to do so as long as separation pays. Thus there is a point where defending separation becomes too costly so that the high-signal bank *pools* with the low-signal bank and an informationally inefficient outcome results.

1.3.2 Equilibrium Analysis

We write $\Pi^1(\mathsf{p}^*, B, s_b)$ for the investment bank's expected profits from their share of the offer revenue. Let $\Pi^2(\mathsf{p}^*, B, s_b)$ denote the expected second period profit from filling the short position at lower prices. In case of a separating equilibrium these are

$$\Pi^2(\bar{p}^*, B_1, s_b = 1) = \sum_{d=\mathsf{S}+\mathsf{O}}^{N} \mathsf{O} \cdot \max\{\bar{p}^*(1-\beta) - \mathsf{p}^m(d), 0\} \cdot \mathsf{Pr}(d|s_b = 1) \qquad (1.10)$$

if the price is risky and $s_b = 1$. For safe prices, the summation in $\Pi^2(\underline{p}^*, B_{0,1}, s_b = 0)$ is from 0 to N, as the IPO never fails. The conditional distribution $\mathsf{Pr}(d|s_b)$ of demand d is the distribution derived for $\alpha_{s_b}(\mathsf{S})$. Note that a high-signal bank sums from $\mathsf{S}+\mathsf{O}$, since lower demand leads to a failure of the IPO. An investment bank with $s_b = 0$, on the other hand, sums from 0 since the IPO is always successful. The bank also accounts for the foregone gross spread β when buying back in the market.

The market price after the offering $\mathsf{p}^m(d)$ adjusts according to investors' signals and with respect to these signals it is informationally efficient. The bank cannot stabilize 'against' this efficient price, but, of course, if the price is efficient, it need not and must not be 'stabilized'. In our model it is, therefore, not possible to study potentially beneficial effects of price stabilization. More generally, however, if one believes in efficient markets, stabilization is undesirable and, if at all, it can have no more than a short-term impact.

With short covering, a high separation price, $\bar{\mathsf{p}}^*$ has to be *small* enough so a low-signal bank cannot profitably deviate from the low, riskless price, $p_{0,0}$. Thus the investment bank with $s_b = 1$ has to determine $\phi_0'(p_{0,0})$ so that

$$\Pi^1(\phi_0'(p_{0,0})|s_b = 0, B_1) + \Pi^2(\phi_0'(p_{0,0})|s_b = 0) = \Pi^1(p_{0,0}|s_b = 0, B_{0,1}) + \Pi^2(p_{0,0}|s_b = 0). \quad (1.11)$$

In what follows, we make two further assumptions. The first states that the overall amount of shares that can be issued remains constant relative to the scenario without aftermarket short-covering. This simplifies computations and later allows us to compare the relative payoffs in both scenarios. The second requires that together signals of investors and bank are sufficiently informative. Figure 1.2 has an illustration of Assumption 1.4.

Assumption 1.3 $\mathsf{S} + \mathsf{O} = (1 - q_i)N$.

Assumption 1.4 *q_i and q_b are large enough so that $p_{1,\frac{1}{2}} > 2p_{0,0}$.*

Using Assumptions 1.3 and 1.4, we can prove the following lemma.

Lemma 1.2 (The Low-Signal Bank's Deviation Price Drops)
The low-signal bank's deviation price with short covering is smaller than without short covering, $\phi_0(\mathsf{p}) \geq \phi_0'(\mathsf{p}) \quad \forall\ \mathsf{p} \in [p_{0,0}, p_{0,\frac{1}{2}}]$.

In the proof we show that for any low-signal bank's deviation price $\phi_0(\mathsf{p})$, second period profits from aftermarket short covering for the low-signal bank are higher at the high, risky price $\bar{\mathsf{p}}^*$. Consequently, this bank has an additional incentive to deviate. The low-signal bank considers it more likely that the price drops, hence its potential gain from short covering is large, in particular, relative to what it can gain by setting the low separation price. To prevent a low-signal bank from mimicking, the high-signal bank has to reduce its offer price. The proof is in Appendix 1.6.5. In what follows, if there is a switch from separating to pooling, we restrict attention to those switches that are to the risky pooling price $p_{1,\frac{1}{2}}$.[19] We can now establish the main result. Analogously to Corollary 1 we spell it out in terms of separation costs as this allows for a more straightforward interpretation. The proof can be found in Appendix 1.6.5.

Proposition 1.3 (Equilibrium with Short Covering Relative to Benchmark)

1. *There exists a lower bound threshold cost* $\underline{C}' > \underline{C}$ *such that for all costs* $C \in [\underline{C}, \underline{C}')$*, the only equilibrium that satisfies the Intuitive Criterion and Pareto efficiency is a pooling equilibrium at the highest risky pooling price* $p_{1,\frac{1}{2}}$*.This price is informationally inefficient.*

2. *There exists an upper bound threshold cost* $\bar{C}'$ *such that for all costs* $C \in [\underline{C}', \bar{C}']$ *the unique equilibrium that satisfies the Intuitive Criterion and Pareto efficiency is a separating equilibrium. For the high separating price* $\bar{\mathsf{p}}^*$ *there exists a threshold cost level* $\hat{C}' \in [\hat{C}, \bar{C}')$ *so that*

   (a) for costs $C \in [\underline{C}', \hat{C}')$ *the high separation price is the low-signal bank's deviation price from the low separating price,* $\bar{\mathsf{p}}^* = \phi_0'(p_{0,0})$*,* $p_{1,\frac{1}{2}} < \phi_0'(p_{0,0}) < p_{1,1}$*, and*

   (b) for costs $C \in [\hat{C}', \bar{C}']$ *the high separation price is the highest possible risky price* $\bar{\mathsf{p}}^* = p_{1,1}$*.*

   On average, underpricing in the separating equilibrium is exacerbated.

Interpretation of the Result. The first part of the proposition states that for all costs smaller than $\underline{C}'$, both types of the bank prefer to pool and hence prices are informationally inefficient. Since $\underline{C}' > \underline{C}$ pooling occurs for a region of parameters where without aftermarket short covering there was separation. That is, the cost

[19] Our results on informational efficiency are not affected by this restriction. On the contrary, taking pooling in a risk-free price also into account would strengthen our findings. In addition, if there is a choice between the high, risky pooling price, $p_{1,\frac{1}{2}}$, and the low, safe pooling price, $p_{0,\frac{1}{2}}$, the former will always generate more ex-ante revenue. We thus focus on high pooling prices to keep the analysis simple.

region for which we get informational efficiency becomes more restrictive. The second part of the proposition outlines the region in which separation is sustained. For all costs smaller than threshold $\hat{C}'$, the investment bank with the good signal charges $\phi_0'(p_{0,0})$, which, by Lemma 1.2, is smaller than the price charged in that corresponding parameter region without short covering. In other words, for costs between $\underline{C}'$ and $\hat{C}'$ offer prices drop. By Proposition 1.2, there is underpricing in a separating equilibrium. Thus, on average, underpricing is exacerbated when separation is sustained. At first glance this result is surprising since second period expected gains are larger the higher the offer price. One might expect that agents are then more inclined to set higher prices. In our model, this casual intuition fails.

The Impact on the Upper Threshold Level for Costs. So far we have focussed on the relation of lower bound threshold costs $\underline{C}$ and $\underline{C}'$ and 'middle' bound threshold costs $\hat{C}$ and $\hat{C}'$. Surely, if $\hat{C}'$ increases relative to $\hat{C}$ (by Lemma 1.2) and $\underline{C}'$ increases relative to $\underline{C}$, then also $\bar{C}'$ should increase relative to $\bar{C}$. But this is not necessarily true – it may actually decrease. Furthermore, if it does increase, it is irrelevant. This is why: Keeping N, β, and O fixed, $\bar{C}$ and $\bar{C}'$ are functions of the signal qualities q_b and q_i. For low signal qualities, $\bar{C}'$ actually decreases. For such values the high separation price $p_{1,1}$ and the low, risk-free pooling price $p_{0,\frac{1}{2}}$ are close. Expected aftermarket profits are higher for the risk-free price and this outweighs the lower expected pooling revenue. For high values of q_b and q_i, both $\bar{C}$ and $\bar{C}'$ exceed the 'natural' upper bound for costs: The worst that can happen, is that a bank loses all (discounted) future business. This upper bound on C can be estimated. In Appendix 1.6.4 we go into the details of this argument, but in what follows we restrict attention to $\underline{C}$, $\underline{C}'$, $\hat{C}$, and $\hat{C}'$. To summarize: The first case of a decreasing upper bound strengthens our result, the second case does not weaken our argument.

Comparative Statics. We can express the overallotment option O as share r of S, that is $\mathsf{S} + \mathsf{O} = (1 + r)\mathsf{S}$. Thus, $r = 0$ is the benchmark case without short covering. Potential policy variables in this setup are the bank's share of the revenue, β, and the size of the overallotment option, r. The proof of the following Proposition is in Appendix 1.6.5.

Proposition 1.4 (Comparative Statics)
The conditions for informational efficiency become more restrictive for the gross spread, β, or the amount of the overallotment facility, r, increasing.

Interpretation of the Proposition. A higher level of β or an increased amount of r strengthen an investment bank's incentive to set higher prices. For a high-signal bank it is thus more difficult to defend a high separation price, consequently, more pooling results.

1.3.3 How would the result change without signaling?

In order to understand the impact of signaling, consider the case where the investment bank gets no signal at all. This is equivalent to the case of a neutral signal $q_b = 1/2$. The conditional probability of there being at least S high-signal investors is

$$\alpha(\mathsf{S}) = \sum_{d=\mathsf{S}}^{N} \binom{N}{d} \frac{1}{2} \left(q_i^d (1-q_i)^{N-d} + (1-q_i)^d q_i^{N-d} \right). \tag{1.12}$$

Here an offer price has no signaling value, investors learn nothing from it. If an investor has favorable signal $s_i = 1$, he buys the security if $\mathsf{p} \leq p_{1,\frac{1}{2}}$, if he has $s_i = 0$ he buys if $\mathsf{p} \leq p_{0,\frac{1}{2}}$. Thus price $p_{0,\frac{1}{2}}$ is risk-free. The investment bank then sets risky price $p_{1,\frac{1}{2}}$, if its expected payoffs are higher than those for the risk-free price,

$$\alpha(\mathsf{S})\beta p_{1,\frac{1}{2}}\mathsf{S} - (1-\alpha(\mathsf{S}))C \geq \beta p_{0,\frac{1}{2}}\mathsf{S}, \tag{1.13}$$

and it sets $p_{0,\frac{1}{2}}$ otherwise. Thus there exists a threshold $\tilde{C}$, such that for all costs $C \leq \tilde{C}$, the investment bank would charge the high price $p_{1,\frac{1}{2}}$, and for all $C > \tilde{C}$, it would play safe and charge $p_{0,\frac{1}{2}}$. However, once short covering is introduced, this second profit opportunity may enable the investment bank to charge a higher price. Simulation of prices show that, $\Pi^2(p_{1,\frac{1}{2}}) > \Pi^2(p_{0,\frac{1}{2}})$. Thus there exists a threshold cost $\tilde{C}'$ larger than $\tilde{C}$ such that the investment bank charges the higher, riskier price where it used to charge the low price. In this case, there would be more *overpricing*, for $C \in [\tilde{C}, \tilde{C}')$. This contrasts our signaling model, which produces the opposite effect: For a non trivial region of parameters we expect to observe, on average, more *underpricing*.

1.4 Payoff Analysis

Although the investment bank has a second source of profits, it is not immediately obvious that it will indeed be better off – if it has the high signal, it may have to distort prices downwards. The bank will thus receive lower expected revenues that may not be outweighed by short covering profits.

Measuring Payoffs. The investment bank's expected payoffs can be measured at two points in time: *Ex-ante*, that is before the bank receives its private information, and *interim*, that is after the signals are realized but before investors take decisions. Issuers have no private information, so their information is exclusively determined ex-ante. As a convention, we compare per-share profits and costs.

Table 1.1 summarizes a bank's conditional signal probabilities, the prices that are charged for each signal, the conditional probabilities of a successful IPO and, given it is indeed successful, the probability of short covering and its profitability. For instance, take $V = 0$ and $s_b = 1$, which occurs with probability $1 - q_b$. In a separating equilibrium the high-signal bank charges $\bar{\mathsf{p}}$ without and $\bar{\mathsf{p}}'$ with short covering. The IPO is successful with probability 1/2 and, given this, there is short covering with probability 1. With probability 1/2 the IPO fails and the bank incurs cost C. Note that if $V = 1$, by the Law of Large Numbers, the IPO almost never fails.

1.4.1 Payoff Comparison for the Investment Bank

We will trickle down from the the strongest to the weakest case: First we analyze the interim type-specific payoffs. This is the strongest case, because we determine when

Without Aftermarket Short Covering						
	$s_b = 1$			$s_b = 0$		
	$\mathsf{Pr}(s_b\|V)$	Price	Pr(success)	$\mathsf{Pr}(s_b\|V)$	Price	Pr(success)
$V = 1$	q_b	$\bar{\mathsf{p}} = \min\{p_{1,1}, \phi_0(p_{0,0})\}$	1	$1 - q_b$	$p_{0,0}$	1
$V = 0$	$1 - q_b$	$\bar{\mathsf{p}} = \min\{p_{1,1}, \phi_0(p_{0,0})\}$	$\frac{1}{2}$	q_b	$p_{0,0}$	1

With Aftermarket Short Covering					
	$s_b = 1$				
	$\mathsf{Pr}(s_b\|V)$	Price	Pr(success)	Pr(short cov.)	Profit $\mathsf{p} - p^m$
$V = 1$	q_b	$\bar{\mathsf{p}}' = \min\{p_{1,1}, \phi_0'(p_{0,0})\}$	1	0	0
$V = 0$	$1 - q_b$	$\bar{\mathsf{p}}' = \min\{p_{1,1}, \phi_0'(p_{0,0})\}$	$\frac{1}{2}$	1	$\kappa\bar{\mathsf{p}}'$
	$s_b = 0$				
	$\mathsf{Pr}(s_b\|V)$	Price	Pr(success)	Pr(short cov.)	Profit $\mathsf{p} - p^m$
$V = 1$	$1 - q_b$	$p_{0,0}$	1	0	0
$V = 0$	q_b	$p_{0,0}$	1	1	$\kappa p_{0,0}$

Table 1.1: **Summary of State-Profits.** The table summarizes the probabilities of signals given values, the separating prices that are charged in each case, the probabilities of a successful IPO and, given that, the probability of short covering, and its profitability. κ is defined as $(1-\beta)r/\beta(1+r)$.

types gain individually. We then proceed with the ex-ante payoff gains. We specify under which conditions the bank would prefer a setting with short covering. Payoffs are then averaged over signal-types because ex-ante, the signal is unknown. This is the weaker case. We focus on the extreme scenarios, that is (a) on the costs with the largest price drops after regime shifts and (b) on costs for which ex-ante payoff *with* short covering is lowest.

To derive the results, we construct the payoff differences from both settings at a given threshold cost and then substitute in closed form approximations of the threshold prices. Details of the formulae can be derived straightforwardly from Table 1.1. Appendix 1.6.3 outlines how the risky threshold prices can be approximated. The resulting risky threshold prices that we find can be interpreted as

$$p_{s,\mu} = \frac{\text{expected liquidation value given the price's information content}}{\text{fraction of cases where this information can be used}} .$$

So for instance, if $\mu(\mathsf{p}) = 1$, the unconditional value of this information piece is the q_b, the quality of the bank's signal. The fraction of cases where this information can be used is the probability of a successful IPO, given $\mu(\mathsf{p}) = 1$: Here it is α_1. Thus

$p_{1,1} = q_b/\alpha_1$. By the same token $p_{1,\frac{1}{2}} = .5/(3/4)$. For the statements below we computed payoff differences for $\beta = 7\%$ and $\mathsf{O} = .15\ \mathsf{S}$, which are the empirically most commonly observed parameters.[20] In summary:

Interim Payoffs for the Low-Signal Bank. Suppose that separation is maintained. Without short covering, per-share profits are $p_{0,0}$. With short covering there are additional expected aftermarket profits of $q_b\kappa p_{0,0}$, as can be seen from Table 1.1. Suppose now that a pooling equilibrium results. Then, by definition of the pooling equilibrium, the low-signal bank benefits. In both cases the low-signal bank is better off with short covering.

Interim Payoffs for the High-Signal Bank. If costs are lower than $\underline{C}$ or if costs are higher than $\hat{C}'$, the bank always wins: In both cases expected revenue remains constant and the bank also gets short covering profits. Suppose now that costs are in $(\underline{C}, \hat{C}')$ and that the price decrease is strongest, from from $p_{1,1}$ to $p_{1,\frac{1}{2}}$. Then for signal qualities in areas A and B in the Left Panel of Figure 1.2, the bank is always better off, despite the maximal price decrease; in areas C and D the bank loses. It may not be better off in all cases, but the smaller the price decrease, the smaller areas C and D become.

Ex-ante Payoffs There are two subcases to consider: (i) The threshold costs for which the highest price decrease occurs, which is $\hat{C}$. (ii) The threshold cost for which the ex-ante payoff *with* short covering is lowest, which is $\underline{C}'$.

(i) Suppose at $\hat{C}$, prices drop from separation in $p_{1,1}$ and $p_{0,0}$ to pooling in $p_{1,\frac{1}{2}}$. At $\hat{C}$, without short covering the low type is indifferent between riskless $p_{0,0}$ and risky $p_{1,1}$. Using the risky payoffs, without short covering payoffs are $(\alpha_1 - \alpha_0)p_{1,1} - costs$.

[20]Chen and Ritter (2000) report that β is almost always 7 percent. Naturally, when both O is small and β is large, some of the statements below may change. Clearly, if these contract variables are such that there is very little to be won in the aftermarket (low O) but a lot to be lost in revenue (high β), then matters may change. However, the essence of the arguments below is that even at the most extreme price drops there is a non-trivial parameter space where the bank is always better off. Taking also parameters sets for β and O into the description of the analysis would merely complicate the exposition.

With short covering payoffs are $(\alpha_1 - \alpha_0)p_{1,\frac{1}{2}} - cost + short\ covering\ profits$. So costs cancel, revenues are lower, but, as it turns out, the short-covering profits *always* overcompensate for the loss in revenue. Thus despite the maximum price decrease at $\hat{C}$, ex-ante the investment bank is always better off with short-covering.

(ii) The bank has lowest ex-ante payoffs *with* short covering at $\underline{C}'$, the costs where the low bank is indifferent between choosing risky separation price $p_{1,\frac{1}{2}} = \phi_0'(p_{0,0})$ and riskless price $p_{0,0}$. The most extreme drop in revenue happens when $\hat{C} < \underline{C}'$, so that without short covering, the bank plays a separation equilibrium. Note that at this cost $\underline{C}'$, without short covering the low type is *not* indifferent between a risky and a riskless price as $\phi_0(p_{0,0}) > p_{1,1}$. Payoffs without short covering are of the order $\alpha_1 p_{1,1} + p_{0,0} - costs$, with short covering they are $\alpha_1 p_{1,\frac{1}{2}} + p_{0,0} - costs + short\ covering\ profits$. The investment bank is better off for parameters q_b, q_i in areas A, B, and C, but not D (Left Panel, Figure 1.2). With respect to signal qualities, this area appears to be large. However, taking (i) into account, this is only relevant for a strict subinterval of $[\hat{C}, \hat{C}']$ and if also $\underline{C}' > \hat{C}$. For all other costs, the bank is ex-ante always better off. The Right Panel of Figure 1.2 illustrates this point.

To summarize, in most cases the investment bank is ex-ante and interim better off.

1.4.2 Payoff Comparison for Issuer and Investors

Given our model specification we can only compare the *revenue* that the issuer receives in settings with and without short covering.[21] Suppose with short covering, separation is maintained. If the separation price decreases, $\bar{\mathsf{p}}' < \bar{\mathsf{p}}$, the issuer loses. Suppose now, there is a switch from separation to pooling. The high separation price decreases from $\bar{\mathsf{p}}$ to $p_{1,\frac{1}{2}}$, but at the same time the low separation price rises from $p_{0,0}$ to $p_{1,\frac{1}{2}}$. Comparison of expected payoffs shows that in this case the issuer is better off for *all* parameter values.[22] Investors' profits are directly opposed to the issuer's profit. Whenever the issuer gains (in expectation) investors lose and vice versa.

[21] This is equivalent to expected profits: Profit here would be defined as the difference between revenue per share and the true value, which, by the LLN, is identical to the aftermarket price. We do not take other factors such as, for e.g., costs for alternative financing (if the IPO fails) into account.

[22] Recall that we restrict the analysis to per-share profits. Taking into account that the number of securities eventually sold to the market will be lower with short covering it can be the case that whenever, simultaneously, q_i is very small and q_b is very large the issuer is worse off even with pooling.

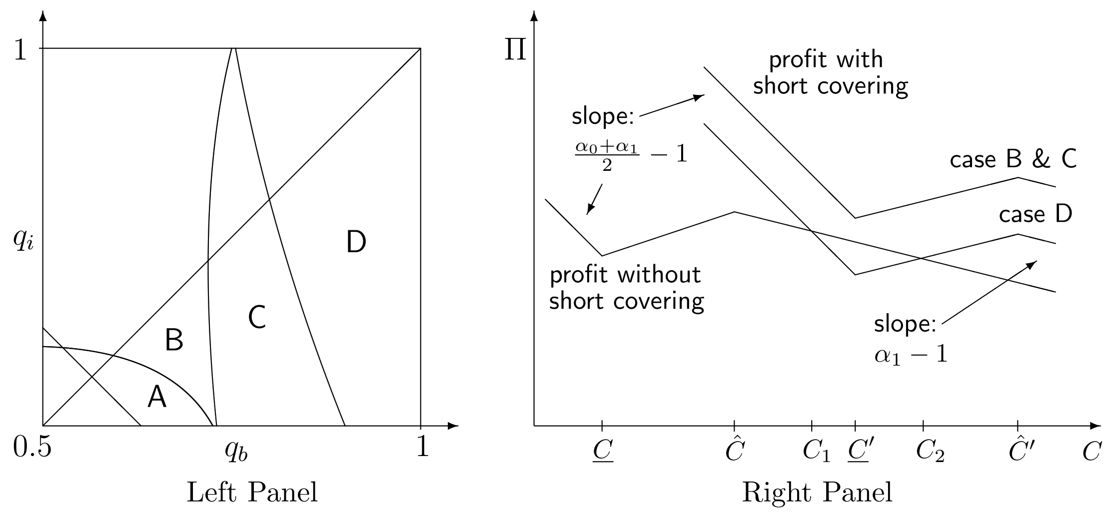

Figure 1.2: **Informational Efficiency and Sign of Bank's Profit Change.**
Left Panel: Areas A, B, C, and D indicate permitted values of q_i and q_b, i.e. $q_b > q_i > .5$ and $p_{1,\frac{1}{2}} > 2p_{0,0}$. For $C = \hat{C}$, A indicates where an informational efficient separating equilibrium is uphold with short covering; in B, C, and D a pooling equilibrium results. The high-signal bank is better off in A and B, and worse off in C and D at $C = \hat{C}$. Ex-ante, the bank is better off in A, B, and C, for all C; in D there exist $C_1 \in [\hat{C}, \underline{C}']$ and $C_2 \in [\underline{C}', \hat{C}']$ such that for $C \in [C_1, C_2]$ it may lose. The figure is based on simulated values for $\beta = .07$, $r = .15$, and $N = 1000$.
Right Panel: The lower line indicates *ex-ante* profits of the bank as a function of C without short covering. The higher lines indicate profits with aftermarket short covering. For the values of q_i and q_b in areas B and C these profits are always higher; in area D it may be the case that for $C \in [C_1, C_2]$ these profits are lower.

Even though this section is merely concerned with redistribution, it yields an interesting insight. The investment bank is nearly always better off with aftermarket short covering, in many cases irrespective of its signal. The issuer never gains but often loses if separation is upheld, but always wins if separation morphs into pooling; the effect on investors' payoffs is the opposite.

1.5 Conclusion

Investment banks legally pursue supposedly price stabilizing activities in the post-offer market. In this paper we analyze how these aftermarket activities influence the setting of the offer price in the first place. We take a different perspective from existing theoretical work as we build the model around the stylized fact that investment banks *can* realize risk-free profits through aftermarket short covering. The current model cannot assess why some investment banks expose themselves to risk and establish 'naked shorts', or why they do not exercise the overallotment option in full even when prices rise above the offer price. This paper only explains the strategic impact of the *possibility* of risk-free profits. The investment bank's behavior must not be perceived as rogue or fraud, but as a rational response to a change in the environment. Investors anticipate the bank's behavior and react rationally to it.

We propose a stylized model of an offering procedure that is in accordance with empirical findings and perceived industry practice. We assume that both the investment bank and investors hold private information about the intrinsic value of the offered security. Prices are set so that rational-expectation investors only order the security if they expect to make a profit, taking into account the behavior of the investment bank. The market price after the offering will adjust according to investors' signals. As these are conditionally i.i.d., the price almost surely reflects the fundamental value of the security. The bank cannot stabilize 'against' this fully efficient price, but, of course, if the price is efficient, it need not and must not be 'stabilized'. So in the best of worlds, one with full transparency, the bank can make an extra profit through short covering. In the real world the IPO process is opaque; neither investors nor regulators nor researchers know precisely the banks' strategies. It is certainly reasonable to assume that in such an 'imperfect world' the strategic impact of the second source of profits is rather more than less important.

There is little empirical support that stabilization is possible and has desirable, positive effects. Indeed it is somewhat surprising that regulators allow price manipulations. It is sometimes argued that investment banks will not always stabilize to avoid a moral hazard problem with investors who believe being fully protected against over-pricing. It is likely that this reduces the effect of potential aftermarket profits as described in this paper. The result itself, however, obviously still holds — there are still hardly any costs involved. In fact, from the regulators perspective price distortions can easily be ruled out if the bank is prohibited from filling the short position at prices below $1 - \beta$ times the offer price. As long as the banks can keep the existence of a short position secret from investors, a moral hazard problem would not occur.

In our setting, the security may turn out to be overpriced. Investors, however, have already taken this into account. Investment banks always set the highest feasible price and thus acts in the issuer's interest. It is important to notice that in our setting the investment bank does not temper prices to rob issuers. The informational asymmetry in the paper arises at a point in time when all official, mandatory information has been released and any other public statement by investment bank or issuer will be perceived as cheap talk. Only actions, that is prices, can carry a meaningful message.

The offering procedure was modelled as a signaling game. The investment bank moves first and strategically chooses the offer price to maximize its profits from both the gross spread of the offer revenue and profits from short covering in the aftermarket. We establish a benchmark by analyzing the situation without aftermarket activities, and identify the conditions under which the equilibrium is both unique and separating. A separating equilibrium is referred to as informationally efficient since the investment bank's information is fully revealed by the offer price. We further show that, on average, securities are underpriced in the separation equilibrium. With the introduction of aftermarket short covering payoff functions and, consequently, the strategic environment change. As a result, either the offer price falls on average, or a pooling equilibrium results. In the first case, an investment bank with favorable information distorts the price downwards and thereby, on average, exacerbates underpricing. In the second case investors are unable to infer the investment bank's signal from the offer price. This equilibrium is informationally inefficient since investors' decisions are based on private signals only and not also on the signal of the investment bank.

The intuition behind the results can be best explained by relating this paper to job-market signaling with two types of workers. In the so-called Riley-outcome, the low type chooses education level zero, and the high type chooses his education just high enough so that it does not pay for the low type to deviate to his level of education. In our paper this corresponds to a low-signal bank choosing a low, risk-free price. At this price all investors want to buy the security and consequently the offering will never fail. Nevertheless, in the aftermarket any offering can turn out to be overpriced. The high-signal bank chooses a high, risky price just low enough so that the risky price does not pay for the low-signal bank. A price is risky when it is so high that only high-signal investors buy; in this case the offering will fail if there are not enough investors with the favorable signal. When introducing profits from short covering, the effect is that of a personal extra benefit from education. Suppose this perk is higher for the low type of worker than for the high type worker. As a result, the high type has to choose a higher level of education to maintain separation. In our model, the low-signal bank considers a price drop in the aftermarket more likely, thus the potential profits from short covering are higher than for the high signal bank. And so the high-signal bank has to distort prices downwards in order to maintain separation. At first sight this is a surprising result, as casual intuition suggests that potential aftermarket profits should result in more *over*-pricing. There may also come a point where it does not pay for the high signal bank to maintain separation, and so it settles for pooling. The result is informational inefficiency.

The investment bank enjoys higher payoffs with short covering for the vast majority of parameter constellations. Looking at per-share profits, the issuer never gains but often loses if separation prevails; but if there is a switch to a pooling equilibrium he is always better off. Investors' payoffs are directly opposed to the issuer's gains or losses. An increase in the investment bank's share of the revenue or an increase in the amount of overalloted securities reduces the parameter-set with informational efficiency.

Our analysis is in accordance with recent empirical analyzes but contrasts the existing theoretical literature which argues that stabilizing activities in the aftermarket serve efficiency. We therefore challenge financial market authorities' view that current regulations simultaneously serve the interests of issuers, investors, and investment banks.

1.6 Appendix

1.6.1 Aftermarket Price Formation

The finally prevailing market price depends on the number of positive signals about the value of the security. In determining the price we have to distinguish between cases B_1 and $B_{0,1}$.

Consider first case B_1. Since only high-signal investors buy, aggregated demand d indicates the number of high-signal investors. Suppose $d \geq \mathsf{S}$, i.e. the IPO is successful. Investors are assumed to take the aggregated information about signals into account and update their expectations accordingly. At this updated expectation all investors irrespective of their private signals are indifferent between selling and holding or buying and abstaining, depending on whether they own a security or not, respectively. The updated expectation thus becomes the aftermarket price, denoted by $\mathsf{p}^m(d)$. We will later show that case B_1 will occur at the high price of a separating equilibrium only, i.e. investors know that the bank's signal is $s_b = 1$. Taking further into account that the true value of the security is either 0 or 1, we can write $\mathsf{p}^m(d|\mu = 1) = \mathsf{Pr}(V = 1|d, \mu = 1)$. Using Bayes' rule, we can express the aftermarket price as

$$\mathsf{p}^m(d|\mu = 1) = \frac{\mathsf{Pr}(d|V = 1)\mathsf{Pr}(s_b = 1|V = 1)}{\mathsf{Pr}(d|V = 1)\mathsf{Pr}(s_b = 1|V = 1) + \mathsf{Pr}(d|V = 0)\mathsf{Pr}(s_b = 1|V = 0)}. \quad (1.14)$$

Due to the binomial structure of the prior distributions over signals, the conditional distribution for demand realization d is, for $V = 1$,

$$f(d|V = 1) := \mathsf{Pr}(d|V = 1) = \binom{N}{d} q_i^d (1 - q_i)^{N-d}, \quad (1.15)$$

and for $V = 0$ analogously. The price-information about s_b is unambiguous in a separating equilibrium. We can therefore replace it with the conditional probability of the bank's signal being correct, which is q_b or $1 - q_b$. Bayes' rule yields

$$\mathsf{p}^m(d|\mu = 1) = \frac{q_b q_i^{2d-N}}{q_b q_i^{2d-N} + (1 - q_b)(1 - q_i)^{2d-N}}. \quad (1.16)$$

Consider now case $B_{0,1}$ in which all investors order the security, i.e. stated demand is N and securities are allocated at random. The demand is uninformative since it does not reveal the number of high-signal investors. Suppose that we are at the low price of a separating equilibrium. Note that high-signal investors expect the security to be of higher value than low-signal investors. Hence, there exists a price larger than the offer price, $\tilde{\mathsf{p}} > \mathsf{p}^*$ at which high-signal investors who were not allocated a security would be willing to buy the security, and low-signal investors would be willing to sell, in case they were allocated a security. Without modelling the price-finding procedure explicitly we assume that the following intermediate process takes place. Those high-signal investors who did not receive the security in the offering submit a unit market-buy-order. Those low-signal investors who obtained the security in the offering submit a unit market-sell-order. All other investors abstain. The number of investors who want to buy or to sell is denoted by $\tilde{d}$ and $\tilde{S}$, respectively. Aggregate demand of high-signal investors is then $d = \tilde{d} + \mathsf{S} - \tilde{S}$ and the market price p^m can be determined as before. The same procedure can be applied to determine the first period market clearing price in the case of a pooling equilibrium. The conditional expectation which determines the price, however, will then not contain the component about the signal of the investment bank.

1.6.2 Threshold Prices

Denote by $p_{s_i,\mu}$ the maximum price at which an investor with signal s_i and price information μ buys, given all investors with $\tilde{s}_i \geq s_i$ buy. At this price the investor's expected return from buying the security is zero, normalizing outside investment opportunities accordingly.

Define $\psi(1|1,1) := \mathsf{Pr}(V = 1|s_i = 1, \mu = 1)$ and $\psi(0|1,1) := \mathsf{Pr}(V = 0|s_i = 1, \mu = 1)$. Consider now the structure of the conditional distribution $f(d-1|V)$. For $V = 1$, this is a binomial distribution over $\{0, \ldots, N-1\}$ with center $(N-1)q_i$, and likewise for $V = 0$ with center $(N-1)(1-q_i)$. Since by Assumption 1.2, N is 'large enough' for every q_i, $f(d-1|1) = 0$ for $d < N/2$ and $f(d|0) = 0$ for $d > N/2$. When combining both $f(d-1|1)$ and $f(d-1|0)$, we obtain a bi-modal function. In $g(\cdot|s_i, \mu)$, investors' posterior distribution over demands, these are weighted with $\psi(1|s_i,\mu)$ and $\psi(0|s_i,\mu)$. Assumption 1.2 now satisfies two purposes. The first is to ensure that we pick N large enough, so that the two modes do not overlap. The second can be seen from the following lemma.

Lemma 1.3 *For any $q_i > \frac{1}{2}$, there exists a number of investors $N(q_i)$, such that $\mathsf{p}^m(d) \cdot g(d-1|s_i,\mu) \in \{0, g(d-1|s_i,\mu)\}$ almost everywhere.*

The lemma states that market prices are mostly 0 or 1, if they are not, then the weight of this demand is negligible. To see this consider the following heuristic argument.

Proof: $\mathsf{p}^m(d)$ is a s-shaped function in d, given by equation (1.16). For large N, $\mathsf{p}^m(d) \in \{0,1\}$ almost everywhere. Define $\mathbb{I}^*$ as the interval of d around $N/2$ s.t. for $d \in \mathbb{I}^*$ we have $\mathsf{p}^m(d) \notin \{0,1\}$. $\mathsf{p}^m(d)$ is multiplied with density $g(d-1|s_i,\mu)$, which peaks at $(N-1)(1-q_i)$ and $(N-1)q_i$. For N increasing $\mathbb{I}^*/N \to 0$ and the bi-modal distribution becomes more centered around $(N-1)(1-q_i)$ and $(N-1)q_i$. Hence, for every q_i there is an $(N-1)(q_i)$ such that for $d \in \mathbb{I}^*$, $g(d|s_i,\mu) \cdot \mathsf{p}^m(d) = 0$, i.e. the weight on $\mathsf{p}^m(d) \notin \{0,1\}$ can be made arbitrarily small. $\square$

Using Lemma 1.3 we can determine the threshold prices as follows. Consider first $p_{1,1}$.

$$\begin{aligned} 0 &= (1-p_{1,1}) \sum_{d=N/2}^{N-1} \frac{\mathsf{S}}{d+1}\, g(d-1|1,1) - p_{1,1} \sum_{d=\mathsf{S}-1}^{N/2} \frac{\mathsf{S}}{d+1}\, g(d-1|1,1) \\ \Leftrightarrow p_{1,1} &= \frac{\sum_{d=N/2}^{N-1} \frac{\mathsf{S}}{d+1}\, g(d-1|1,1)}{\sum_{d=\mathsf{S}-1}^{N-1} \frac{\mathsf{S}}{d+1}\, g(d-1|1,1)}. \end{aligned} \tag{1.17}$$

For $d > N/2$, $g(d-1|s_i,\mu) = \psi(1|s_i,\mu) f(d-1|1)$ and for $d < N/2$, $g(d-1|s_i,\mu) = \psi(0|s_i,\mu) f(d-1|0)$. Also define

$$\Sigma_0 := \sum_{d=\mathsf{S}-1}^{N/2} \frac{f(d-1|0)}{d+1} \quad \text{and likewise} \quad \Sigma_1 := \sum_{d=N/2}^{N-1} \frac{f(d-1|1)}{d+1}, \quad \text{and} \quad \sigma := \Sigma_0/\Sigma_1.$$

Also write $\ell(\mu) := \psi(0|1,\mu)/\psi(1|1,\mu)$. Thus for the combination of signal s_i and price-information μ with B_1 we can write

$$p_{1,1} = (1+\sigma\ell(1))^{-1} \quad \text{and likewise} \quad p_{1,\frac{1}{2}} = (1+\sigma\ell(\tfrac{1}{2}))^{-1}. \tag{1.18}$$

Consider now the case for $p_{0,0}$. At this price all agents receive the security with equal probability and we sum from 0 to $N-1$. Thus

$$0 = (1-p_{0,0}) \sum_{d=N/2}^{N-1} \frac{\mathsf{S}}{N}\, g(d-1|0,0) - \\ p_{0,0} \sum_{d=0}^{N/2} \frac{\mathsf{S}}{N}\, g(d-1|0,0) \Leftrightarrow p_{0,0} = \psi(1|0,0). \tag{1.19}$$

Likewise we have

$$p_{0,\frac{1}{2}} = \psi(1|0,\tfrac{1}{2}). \tag{1.20}$$

1.6.3 Approximate Closed Form Solutions

We will now derive *approximate* closed form solutions so that we can solve our model analytically. In this appendix we let d denotes the number of *other* investors with favourable information — this contrasts the exposition of the main text, but it simplifies the notation here. First consider the strategy of agent number N. There are $N-1$ other investors. Given that he invests and the true value is, say, $V=1$, then by the law of large numbers, demand/the number of favorable signals will always be larger than $N/2$. Furthermore, the market price is almost surely $p^m(d)=1$. If d others order, then when buying he gets the asset with probability $1/(d+1)$. Thus his payoff for price p

$$(1-\mathsf{p}) \sum_{d=(1-q_i)N-1}^{N-1} \frac{1}{d+1}\binom{N-1}{d} {q_i}^d (1-q_i)^{N-1-d} = \\ (1-\mathsf{p}) \sum_{d=N/2}^{N-1} \frac{1}{d+1}\binom{N-1}{d} {q_i}^d (1-q_i)^{N-1-d}. \tag{1.21}$$

To compute the sum we proceed in a similar manner as one would to compute the expected value of a binomial distribution: First observe that because N is large,

$$\sum_{d=N/2}^{N-1} \frac{1}{d+1}\binom{N-1}{d}{q_i}^d(1-q_i)^{N-1-d} = \sum_{d=0}^{N-1} \frac{1}{d+1}\binom{N-1}{d}{q_i}^d(1-q_i)^{N-1-d} \quad (1.22)$$

Then we can compute

$$\begin{aligned}
&\sum_{d=0}^{N-1} \frac{1}{d+1}\binom{N-1}{d}{q_i}^d(1-q_i)^{N-1-d} = \frac{1}{q_iN}\sum_{d=0}^{N-1} \frac{N!}{(N-d)!(d+1)!}\ {q_i}^{d+1}(1-q_i)^{N-1-d} \\
=\ & \frac{1}{q_iN}\left(\sum_{l=0}^{N}\binom{N}{l}{q_i}^l(1-q_i)^{N-l} - \binom{N}{0}{q_i}^0(1-q_i)^{N-0}\right) \\
=\ & \frac{1}{q_iN}(1-(1-q_i)^N). \qquad (1.23)
\end{aligned}$$

In the second step we made a change of variable, $l = d+1$, but through this change, we had to subtract the element of the sum for $l = 0$. Consequently, for large N, we can say that

$$\sum_{d=N/2}^{N-1} \frac{1}{d+1}\binom{N-1}{d}{q_i}^d(1-q_i)^{N-1-d} \approx \frac{1}{q_iN}. \quad (1.24)$$

Using the same arguments, we could also show that

$$\sum_{d=0}^{N-1} \frac{1}{d+1}\binom{N-1}{d}{q_i}^{N-1-d}(1-q_i)^d \approx \frac{1}{(1-q_i)N}. \quad (1.25)$$

Use now familiar notation to denote the combination of private and public beliefs $\phi_{s,\mu}$. Recall that we can write $p_{1,1}$ as

$$p_{1,1} = \frac{1}{1+\ell(1)\ \frac{\Sigma_0}{\Sigma_1}}. \quad (1.26)$$

What we now need to find is a closed form for

$$\Sigma_0 = \sum_{d=N(1-q_i)-1}^{N/2} \frac{1}{d+1}\binom{N-1}{d}{q_i}^{N-1-d}(1-q_i)^d. \quad (1.27)$$

For increasing N one can see that $\frac{1}{d+1}\binom{N-1}{d}{q_i}^{N-1-d}(1-q_i)^d$ gets *numerically* symmetric around $(1-q_i)N-1$. Thus we can express

$$\begin{aligned}\Sigma_0 &= \frac{1}{2}\sum_{d=0}^{N/2}\frac{1}{d+1}\binom{N-1}{d}{q_i}^{N-1-d}(1-q_i)^d = \frac{1}{2}\sum_{d=0}^{N}\frac{1}{d+1}\binom{N-1}{d}{q_i}^{N-1-d}(1-q_i)^d \\ &\approx \frac{1}{2}\frac{1}{(1-q_i)N}.\end{aligned} \tag{1.28}$$

Assembling, we obtain

$$p_{1,1} = \frac{1}{1+\ell(1)\,\frac{\Sigma_0}{\Sigma_1}} \approx \frac{1}{1+\frac{(1-q_i)(1-q_b)}{q_i q_b}\frac{q_i N)}{2(1-q_)N)}} = \frac{2q_b}{1+q_b} \equiv \frac{q_b}{\alpha_1}. \tag{1.29}$$

Equivalently, we get

$$p_{1,\frac{1}{2}} \approx \frac{1}{1+\frac{1-q_i}{q_i}\frac{q_i N}{2(1-q_i)N}} = \frac{2}{3}, \text{ and } p_{0,1} \approx \frac{1-q_b}{\alpha_0}. \tag{1.30}$$

The information content of a high pooling price is 1/2, and knowing this information, the probability of the offering being successful is 3/4. Thus the interpretation of risky prices is thus the ratio of the expected liquidation value given price-information to the share of successful offerings given this information

$$p_{1,\mu} = \frac{\mathsf{E}[\,V \mid \mu\,]}{\mathsf{Pr}(\text{IPO successful} \mid \mu)}. \tag{1.31}$$

1.6.4 Maximal Reputation Costs

If an IPO fails, the worst that can happen is that the investment bank loses all future IPO business, i.e. it is out of the market. Assuming that future business takes place in the same environment (e.g. the quality of signals remains constant), the bank can maximally lose all discounted future profits. Assume that the bank discounts future profits at rate δ. Consider the case of highest potential costs $\bar{C}$ that can occur from a failing IPO in a separating equilibrium. An upper bound for costs is given by the discounted lost future profits if $\bar{\mathsf{p}} = p_{1,1}$. Then ex-ante profits of a single IPO are

$$\Pi(p_{0,0}, p_{1,1}, C) = \frac{1}{2}\ (\mathsf{S}+\mathsf{O})\beta\ (p_{0,0} + \frac{1+q_b}{2} p_{1,1}) - \frac{1-q_b}{4}\ C. \tag{1.32}$$

Assuming that an investment bank would conduct one IPO each period and accounting for the fact that in a separating equilibrium the ex-ante probability of the IPO to be successful is $(3+q_b)/4$ we get

$$C_{\max} = \sum_{t=0}^{\infty} (1-\delta)^t \cdot ((3+q_b)/4)^t \cdot \Pi(p_{0,0}, p_{1,1}, C_{\max}). \tag{1.33}$$

Thus maximal possible costs can be solved to be

$$C_{\max} = 2(\mathsf{S}+\mathsf{O})\beta \frac{p_{0,0} + \frac{1+q_b}{2} p_{1,1}}{\delta(3+q_b) + 2(1-q_b)}. \tag{1.34}$$

Comparing values of $C_{\max}$ to those of $\bar{C}$ shows that for q_i and q_b sufficiently large $\bar{C} \gg C_{\max}$. Furthermore, for reasonable values of the discount rate, the reverse relation holds true only for values of q_i and q_b where we get $\bar{C}' < \bar{C}$. That is, either $\bar{C}' < \bar{C}$ and informational inefficiencies result, or $\bar{C}$ is so large that it lies outside the relevant parameter region in the context of this model.

1.6.5 Omitted Proofs

Proof of Lemma 1.1

Suppose $\underline{\mathsf{p}}^* > p_{0,0}$. At this price only high-signal investors buy. A high-signal bank will always set a price where at least high-signal investors buy. Hence, high-signal investors buy at both prices $\underline{\mathsf{p}}^*$ and $\bar{\mathsf{p}}^*$. A low-signal bank can now increase its payoff by setting a higher price as α_0 is not affected by this, a contradiction. □

Proof of Proposition 1.1

First we will argue that the only separating equilibrium surviving the IC is the one outlined in the proposition. Then we will argue that pooling cannot occur.

Step 1 (*Separating*) First observe that there cannot be a separating price $\bar{\mathsf{p}}^*$ where investors choose $B_{0,1}$ because otherwise the low-signal bank would deviate to this price. Note that no separating price with $\bar{\mathsf{p}}^* > \phi_0(p_{0,0})$ can exist because at this price, the low-signal bank would prefer to deviate. No price $\bar{\mathsf{p}}^* > p_{1,1}$ can exist since not even high-signal investors would buy. Furthermore, $\bar{\mathsf{p}}^* \geq \phi_1(p_{0,0})$ must be satisfied since otherwise the high-signal bank would prefer to deviate to $p_{0,0}$. Finally no price $\bar{\mathsf{p}}^*$ below $p_{1,0}$ is reasonable because the high-signal bank would then deviate to this price. Take $\tilde{p}$, with $\max\{\phi_1(p_{0,0}), p_{1,0}\} \leq \tilde{p} \leq \min\{p_{1,1}, \phi_0(p_{0,0})\}$. Note that such a $\tilde{p}$ always exists as long as $\phi_1(p_{0,0}) \leq p_{1,1}$ and $p_{1,0} \leq \phi_0(p_{0,0})$. The conditions stated in Proposition 1.1 ensure this is the case because $\phi_1(p_{0,\frac{1}{2}}) > \phi_1(p_{0,0})$ and $p_{1,\frac{1}{2}} > p_{1,0}$.

We analyze the candidate separating equilibrium

$$\{(\underline{\mathsf{p}}^* = p_{0,0}, \mu = 0, B_{0,1}); (\bar{\mathsf{p}}^* = \tilde{p}, \mu = 1, B_1);$$
$$(\mathsf{p}^* \notin \{\underline{\mathsf{p}}^*, \bar{\mathsf{p}}^*\}, \mu = 0, B_{0,1} \text{ if } \mathsf{p} \leq p_{0,0}, B_1 \text{ if } p_{0,0} < \mathsf{p} \leq p_{1,0}, B_\emptyset \text{ else})\}.$$

By definition of $\phi_0(p_{0,0})$ it holds that

$$\beta p_{0,0}\mathsf{S} = \alpha_0\beta\phi_0(p_{0,0})\mathsf{S} - (1-\alpha_0)C > \alpha_0\beta\tilde{p}\mathsf{S} - (1-\alpha_0)C$$

so that the low-signal bank would not deviate to $\tilde{p}$. Since $\max\{\phi_1(p_{0,0}), p_{1,0}\} \leq \tilde{p}$, the high-signal bank would also not deviate. Hence this is a PBE.

Now consider the application of the IC. Suppose a high separation price $\bar{\mathsf{p}} = \tilde{\tilde{p}}$ with $\tilde{p} < \tilde{\tilde{p}} \leq \min\{p_{1,1}, \phi_0(p_{0,0})\}$ is observed. This price is equilibrium dominated for a bank with $s_b = 0$ by definition of $\phi_0(p_{0,0})$. The low-signal bank can therefore be excluded the set of potential deviators. The only remaining agent is the high-signal bank. The best response of high-signal investors then is to buy at $\bar{\mathsf{p}} = \tilde{\tilde{p}}$, i.e. B_1. Hence the PBE with $\bar{\mathsf{p}}^* = \tilde{p}$ does not survive the IC. Applying this reasoning repeatedly, all separating prices with $\bar{\mathsf{p}} < \min\{p_{1,1}, \phi_0(p_{0,0})\}$ can be eliminated.

Step 2a (*Pooling with* $B_{0,1}$) For all investors to buy we must have $\mathsf{p} \leq p_{0,\frac{1}{2}}$. Suppose there was deviation to $\mathsf{p} = \phi_1(p_{0,\frac{1}{2}}) < \phi_0(p_{0,\frac{1}{2}})$. For the low-signal bank this would not be profitable by definition of $\phi_0(p_{0,\frac{1}{2}})$. But for some beliefs about the signal of the bank and corresponding best responses, high-signal investors could be better off. The best response for investors with beliefs on the remaining set of types, i.e. $\mu = 1$, however, is B_1 as we have $\phi_1(p_{0,\frac{1}{2}}) < p_{1,1}$. Hence, applying IC there cannot be a pooling equilibrium with $B_{0,1}$.

Step 2b (*Pooling with* B_1) We must have $\mathsf{p} \leq p_{1,\frac{1}{2}}$. Since $\phi_0(p_{0,0}) > p_{1,\frac{1}{2}}$, the low-signal bank would prefer to deviate to $p_{0,0}$, hence this cannot be an equilibrium.

To summarize, restrictions $\phi_1(p_{0,\frac{1}{2}}) < p_{1,1}$ and $\phi_0(p_{0,0}) > p_{1,\frac{1}{2}}$ ensure that the only equilibrium surviving the IC is the one depicted in Proposition 1.1. □

Proof of Proposition 1.2

Consider the highest possible separating offer prices. The market price will by the Law of Large Numbers resemble the true value of the security. Assumptions 1.1 and 1.2 imply that the IPO fails with probability 0.5 if the true value is $V = 0$ and the high separation price is set. If the true value is $V = 1$ the IPO never fails. Thus, ex-ante there is underpricing if $\frac{1}{2}(1 - p_{0,0} - \alpha_1 p_{1,1}) > 0$. Substituting in closed form solutions for threshold prices $p_{1,1}$ and $p_{1,\frac{1}{2}}$ from Appendix 1.6.3 this can be written as

$$\frac{(1-q_b)(1-q_i)}{q_b q_i + (1-q_b)(1-q_i)} + q_b \leq 1 \tag{1.35}$$

Recall that $\alpha_1 = \frac{1+q_b}{2}$. Numerically, it is straightforward to check that the inequality holds for all $q_b, q_i \in (.5, 1)$. □

Proof of Lemma 1.2

We will analyze two cases. Firstly we will show that at $C = \hat{C}$, $\bar{\mathsf{p}}^* = p_{1,1} = \phi_0(p_{0,0})$ can no longer be a sustained as a separating equilibrium if short covering is possible.

Secondly we will show that at $C = \underline{C}$, $\bar{\mathsf{p}}^* = p_{1,\frac{1}{2}} = \phi_0(p_{0,0})$ cannot be sustained as the separating equilibrium.

We will regard situations in which with respect to the offering price the low-signal bank is indifferent between charging $p_{0,0}$ with all investors buying, $B_{0,1}$, and $\bar{\mathsf{p}}^*$ where only high-signal investors buy, B_1. If the payoffs from short covering are higher in the case of deviating to price $\bar{\mathsf{p}}^*$, then this price can no longer be sustained as a separating price and then, naturally, $\phi_0'(p_{0,0}) < \phi_0(p_{0,0})$. To get this we need to show $\Pi^2(\bar{\mathsf{p}}^*|B_1, s_b = 0) > \Pi^2(\underline{\mathsf{p}}^*|B_{0,1}, s_b = 0)$. Defining $\Delta(\mathsf{p})$ s.t. $\forall \quad d \leq \Delta(\mathsf{p})$ the aftermarket price is not above $(1-\beta)\mathsf{p}$ this is equivalent to

$$
\begin{aligned}
\sum_{d=\mathsf{s}+\mathsf{o}}^{\Delta(\bar{p}^*)} \mathsf{O} \cdot \{(1-\beta)\bar{\mathsf{p}}^* - \mathsf{p}^m(d)\} \cdot \mathsf{Pr}(d|s_b = 0) \quad &> \quad \sum_{d=0}^{\Delta(\underline{p}^*)} \mathsf{O} \cdot \{(1-\beta)\underline{\mathsf{p}}^* - \mathsf{p}^m(d)\} \cdot \mathsf{Pr}(d|s_b = 0) \\
\Leftrightarrow \quad \left.\begin{array}{c} (1-\beta)\bar{\mathsf{p}}^* \sum\limits_{d=\mathsf{s}+\mathsf{o}}^{\Delta(\bar{p}^*)} \mathsf{Pr}(d|s_b = 0) \\ - \sum\limits_{d=\mathsf{s}+\mathsf{o}}^{\Delta(\bar{p}^*)} \mathsf{p}^m(d) \cdot \mathsf{Pr}(d|s_b = 0) \end{array}\right\} \quad &> \quad \left\{\begin{array}{c} (1-\beta)\underline{\mathsf{p}}^* \sum\limits_{d=0}^{\Delta(\underline{p}^*)} \mathsf{Pr}(d|s_b = 0) \\ - \sum\limits_{d=0}^{\Delta(\underline{p}^*)} \mathsf{p}^m(d) \cdot \mathsf{Pr}(d|s_b = 0) \end{array}\right. \\
\tilde{\Leftrightarrow} \; (1-\beta)\bar{\mathsf{p}}^* \frac{q_b}{2} \quad &> \quad (1-\beta)p_{0,0}q_b. \qquad (1.36)
\end{aligned}
$$

The last step follows from Lemma 1.3 in Appendix 1.6.2. We can now check what happens at the threshold points. Suppose that $C = \underline{C}$ so that $\bar{\mathsf{p}}^* = p_{1,\frac{1}{2}}$. Then (1.36) translates to $p_{1,\frac{1}{2}}/2 > p_{0,0}$ which is ensured by Assumption 1.4. Recall that numerically this assumption requires that not both q_i and q_b are small. Suppose that $C = \hat{C}$ so that $\bar{\mathsf{p}}^* = p_{1,1}$. Then we need that $p_{1,1}/2 > p_{0,0}$. Informativeness of s_b implies $p_{1,1} > p_{1,\frac{1}{2}}$. □

Proof of Proposition 1.3

The second step of the proof of Lemma 1.2 ensures that $\underline{C}' \geq \underline{C}$. The model is setup so that all payoffs $\Pi^1 + \Pi^2$ can be dealt with as one. Hence the aforementioned procedure can be applied here as well. The proof of the pooling outcome goes exactly along the lines of the proof of Proposition 1.1. Take a separating equilibrium in which both agents make less profit than in the pooling equilibrium. Pareto Efficieny rules this equilibrium out. The existence of $\hat{C}' > \hat{C}$ is again ensured by Lemma 1.2. By definition, for $C > \hat{C}'$, the highest attainable price is $p_{1,1}$, and it is the only one selected by the IC. □

Proof of Proposition 1.4

From Proposition 1.3 we know that a pooling equilibrium results for all $C < \underline{C}'$. $\underline{C}'$ is defined as the value of C for which equation (1.11) is fulfilled with $\phi_0'(p_{0,0}) = p_{1,\frac{1}{2}}$. Solving for $\underline{C}'$ one obtains

$$\underline{C}' \propto \beta(\mathsf{S}+\mathsf{O})\left(\frac{2-q_b}{2}p_{1,\frac{1}{2}} - p_{0,0}\right) + (1-\beta)\mathsf{O}q_b\left(\frac{p_{1,\frac{1}{2}}}{2} - p_{0,0}\right). \tag{1.37}$$

Partially differentiating w.r.t. O we obtain

$$\frac{\partial \underline{C}'}{\partial \mathsf{O}} = \beta\left(\frac{2-q_b}{2}p_{1,\frac{1}{2}} - p_{0,0}\right) + (1-\beta)q_b\left(\frac{p_{1,\frac{1}{2}}}{2} - p_{0,0}\right). \tag{1.38}$$

Both terms in brackets are positive by Assumption 1.4 as long as $q_b < 1$. Partial differentiation w.r.t. β yields

$$\begin{aligned}
\frac{\partial \underline{C}'}{\partial \mathsf{O}} &= (\mathsf{S}+\mathsf{O})\left(\frac{2-q_b}{2}p_{1,\frac{1}{2}} - p_{0,0}\right) - q_b\left(\frac{p_{1,\frac{1}{2}}}{2} - p_{0,0}\right) \\
&\propto \left[\frac{p_{1,\frac{1}{2}}}{2}\left(2 - q_b - \frac{r}{1+r}q_b\right) - p_{0,0}\left(1 - \frac{r}{1+r}q_b\right)\right].
\end{aligned} \tag{1.39}$$

Since $2 - q_b - \frac{r}{1+r}q_b > 1 - \frac{r}{1+r}q_b$ whenever $q_b < 1$, Assumption 1.4 ensures that the term is positive. $\square$

Chapter 2

Investment Bank Compensation in Venture and Non-Venture Capital Backed IPOs*

2.1 Introduction

This paper proposes a signaling model of the initial public offering (IPO) process that can explain two related puzzles. First, why investment banks' share of IPO revenue, the gross spread, is so large that they are left with profits despite market competition and, second, why these spreads are significantly lower in venture capital (VC) backed IPOs than in non-VC backed IPOs. Megginson and Weiss (1991) report for a U.S. sample of 640 IPOs between 1983 and 1987 gross spreads of 7.4% for VC backed IPOs and 8.2% for non-VC backed IPOs. Data on the profitability of IPOs is hard to obtain. However, Chen and Ritter (2000) argue that there are economies of scale in underwriting IPOs. They show that spreads do not differ in offerings that raise between $20 million and $80 million. Since investment banks at least break even in small offerings large offerings must be profitable. They report that "investment bankers readily admit that the IPO business is very profitable" (p. 1105).

*The chapter is based on joint work with Andreas Park from the University of Toronto.

The cost of going public is subject of a long standing and lasting debate among financial economists. The literature focuses mainly on the underpricing of IPOs, an empirically well-documented phenomenon. Ritter and Welch (2002), as a recent example, report an average first-day return of 18.8% for 6,249 IPOs in the U.S. between 1980 and 2001. However, underpricing is not the only cost of going public. Issuers leave a fraction of the offer revenue, the gross spread, to investment banks (underwriters) as compensation for their services. Chen and Ritter (2000) find that the gross spread amounts to 7% on average for a U.S. sample of 3,203 IPOs between 1985 and 1998. Although investment bank compensation thus accounts for a substantial proportion of the cost of going public, it has hardly attracted attention by theorists so far. In light of the impact that the Chen and Ritter (2000) paper had, this is even more surprising. Not only do they find that gross spreads average 7% but that they are exactly 7% in most of the offerings. Hansen (2001) reports that these findings triggered 27 lawsuits against investment banks for not competing in price and a U.S. Department of Justice investigation of "alleged conspiracy among securities underwriters to fix underwriting fees." Thus, in practice, the size of the gross spread attracts considerable attention. Even though this paper does not explain the *clustering* of spreads that called attention in the first place, we explain the *level* of spreads – the underlying bone of contention.

Notwithstanding the legal debate on investment bank collusion and the seemingly obvious empirical evidence, our theoretical formulation allows a very different, subtle explanation for high spreads. We find that it can be in the best interest of the issuer to pay seemingly inflated spreads. That is, issuers would strategically pay high spreads even if a competing bank offered its service at a lower spread. Issuers hence do not bargain for lower spreads and, consequently, investment banks do not compete in them. In our model, there are two reasons for this. First, it is at the bank's discretion to set the offer price. At high offer prices the IPO can fail because there may not be enough investors willing to subscribe. Banks then have to bear reputation costs. To induce them to take the risk and set high offer prices issuers must set sufficiently high spreads. Banks earn a rent because given the spread they could set a low, risk-free price and receive their share in revenue with certainty. Secondly, we assume that investment banks hold private information about market's valuation of the firm on offer. In our model the size of the spread critically affects the bank's decision whether to reveal or to hide this information which in turn affects the offer price. Thus a lower spread may result in lower expected revenue for the issuer. However, competition could instead take place in features of the IPO contract that are left outside the model, as for example additional co-managers or analyst coverage. This fits the findings of Chen and Ritter (2000) as they report that not only the clustering of spreads at 7% has increased over time – indicating lack of competition in spreads – but that the number of co-managers in IPOs and analyst coverage has increased over time as well – indicating some competition in these features of the IPO contract.

Megginson and Weiss (1991) were the first to show that VC backed issuers pay lower gross spreads than non-VC backed issuers.[23] If the IPO is non-VC backed the founder of the firm takes all relevant decisions. In VC backed IPOs the venture capitalist usually holds all control rights, so we assume that in a VC backed IPO the venture capitalist decides on the level of the spread. Our second main result addresses differences between IPOs with uninformed and privately informed issuers. Assuming that, in contrast to the founder, the venture capitalist has private information about market sentiment, our model predicts that in equilibrium spreads are lower in VC than in non-VC backed IPOs.

We propose a simple stylized model of the offering procedure, cast into a signaling game. We assume that investment banks and investors have private but noisy information about the intrinsic value of the offered security which is either "good news" or "bad news". In a wider sense, this signal can also be understood as information about market sentiment. Investment banks strategically choose the offer price to maximize their expected profits from the offer revenue gross spread. A higher price does not necessarily increase revenue: at high prices the IPO may fail as there may not be enough investors to buy up the entire offering. Prior to the signaling game the issuer offers the investment bank a contract that specifies the gross spread level. This level critically affects banks' pricing decisions. Given the contract variables, the investment bank sets

[23]For more recent data see Francis and Hasan (2001)

an optimizing price that, first, either reveals (separation) or camouflages (pooling) its private information and, second, is either low so that all investors subscribe (risk free) or high so that only investors with "good news" buy (risky). If the issuer is also privately informed, the gross spread level can be either separating or pooling, too. Banks, in turn, account for the spread's information content when deciding on the offer price. Anticipating the bank's pricing decision, the issuer sets the level of the gross spread strategically so that the bank sets the offer price that gives the issuer highest expected profit. Investors are aware of this process and subscribe only if their expected profits are non-negative. At the equilibrium spread the investment bank makes positive profits.

For informed issuers we consider two cases: In the first, the issuer receives a private signal that is conditionally independent from the bank's signal (later interpreted as VC backed IPOs). In the second case, the issuer's signal is perfectly correlated with the investment bank's signal (strong banking ties). In the first case, the issuer will not reveal his private information and set a spread that hides his signal. Nevertheless, the spreads is set so that the investment bank will separate in prices. In the second case, an issuer with favorable information sets a spread that prevents its low-signal counterpart from mimicking. Spreads are thus separating and also indicate the bank's signal. Prices can hence carry no additional information. With uninformed issuers the spread cannot convey information. We show that spreads are then set so that both types of investment bank pool in the offer price, causing an informationally inefficient equilibrium. The pooling spreads with independently informed issuers (VC backed) are lower than the spread set by an uninformed issuer (non-VC backed). Furthermore, if bank's and issuer's signals are perfectly correlated, spreads are, on average, the highest.

The remainder of the paper is organized as follows. Section 2.2 presents our model of the IPO procedure. Section 2.3 derives the equilibrium prices set by the investment bank. Section 2.4 analyzes the strategic choice of the gross spread by uninformed and superiorly informed issuers. Section 2.5 presents the main results on levels and difference of gross spreads. Section 2.6 concludes.

2.2 A Stylized Model of the IPO Procedure†

Consider the following stylized model of the IPO process.

The Security. The security on offer can take values $V \in \mathbb{V} = \{0, 1\}$, both equally likely. The realization is not known to any player in the game.

The Investors. There are N identical, risk neutral investors. Each investor receives a costless, private, conditionally i.i.d. signal $s_i \in \mathbb{V}$ about the value of the security. This information is noisy but (for technical reasons) sufficiently informative, i.e. $\mathsf{Pr}(s_i = v|V = v) = q$, with $q \in (0.6, 1)$, where $v \in \mathbb{V}$.[24] If an investor orders, he may or may not obtain the security during the offering procedure. Shares are distributed with uniform probability in case the issue is oversubscribed. If an investor obtains the security his payoff is the market price minus the offer price. If he does not obtain the security or if the offer is not floated his payoff is zero. An investor's type is his signal. We refer to the investor as a 'high-signal investor' if $s_i = 1$. For $s_i = 0$, it is a 'low-signal investor'.

The Issuer. In general the issuer can be either informed or uninformed. For the latter we consider two subcases: in the first, the issuer (firm) receives a private signal $s_f \in \{0, 1\}$, in the second, the issuer and the investment bank (see below) receive the identical signal. Any signal is costless and conditionally independent from the investors' signals but, for simplicity, of the same quality, i.e. $\mathsf{Pr}(s_i = v|V = v) = q$. The uninformed issuer receives no signal. We will refer to these types of issuers as 'privately informed', 'identically informed', and 'uninformed'. In Section 2.5 we interpret the meaning of informative signals and relate informed and uninformed issuers to real-world types such as VC backed and non-VC backed issuers. The issuer is risk neutral and signs a contract with an investment bank that delegates the pricing decision and constitutes

†The model presented here is in part similar to the model presented in Chapter 1. However, to keep this chapter self-contained we accept some redundancies.

[24] As turns out in the analysis, for $q < .6$ we have to distinguish a large number of subcases. Avoiding these complications, we thus require sufficiently informative signals.

the amount of securities, S, to be sold. It also specifies the publicly announced gross spread $\beta \in (0,1)$, the share of the offer revenue that remains as remuneration at the bank.[25] The issuer chooses this spread. If the offer is floated, his profit is fraction $(1-\beta)$ of the offer revenue, otherwise it is zero.

Investment Banks. Investment banks are risk neutral. A bank that gets offered a contract (we will show that it always accepts) receives a costless, private signal $s_b \in \mathbb{V}$ about the value of the security. The signal is conditionally independent from the investors' and privately informed issuer's signals but, for simplicity, of the same quality, i.e. $\mathsf{Pr}(s_b = v|V = v) = q$. Signals characterize a bank's type: If $s_b = 1$ we refer to the investment bank as a 'high-signal bank', for $s_b = 0$, it is a 'low-signal bank'. After receiving the signal the bank chooses the offer price p. If there is excess demand, securities are allocated at random; if the number of investors willing to buy is less than the number of shares to be sold, the offer is called off. We assume that failure of the offering inflicts fixed costs C on the investment bank. These costs are external to our formulation and to be thought of as deterioration of reputational capital. They may also capture the opportunity costs resulting from lost market share when being associated with an unsuccessful IPO.[26] Without loss of generality, the offering procedure itself causes no costs for the investment bank. Thus, if the offer is successful, the bank's payoff is fraction β of the offer revenue; if it fails, a loss of C results.

Signaling Value of the Gross Spread and the Offer Price. The level of the gross spread and the offer price are announced first. Then investors decide whether or not to order, basing their decisions on their private information and on the information

[25] We want each agent to have only one choice variable: issuers choose the spread, banks the price and investors may or may not invest. Another candidate choice variable is the number of shares S, or even the number of potential investors N that are addressed, e.g. during the road-show. However, including these as choice variables would require a different, more elaborate modelling approach.

[26] There is an extensive large literature on investment bank reputation. Dunbar (2000), for e.g., provides evidence that established investment banks lose market share when being associated with withdrawn offerings. Booth and Smith (1986) argue that in the context of asymmetric information between insiders and outsiders the investment bank as a repeated player in the IPO market certifies that the issue is not overpriced. Following this argument, C can be interpreted as measuring the deterioration of the certification value of the investment bank's brand name.

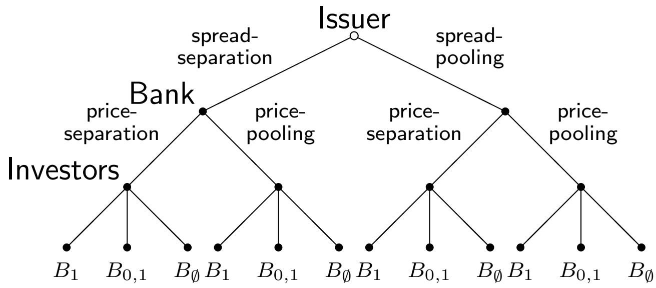

Figure 2.1: **Extensive Form of the Signaling Game.**

that issuer and bank reveal about their signals through the level of the gross spread and the offer price. We denote information contained in prices by $\mu(\mathsf{p})$, information in spreads $\nu(\beta)$. In case of the uninformed issuer, the spread is uninformative and only prices can carry information. In case of an identically informed issuer, the information contained in β is hierarchical to the information in p: Issuers with different signals may set different levels of the gross spread which then reveals the signal of the issuer and the bank; in this case, prices cannot carry further information. We write $\mu(\mathsf{p}) = 1$ if the price reflects that the bank's signal is $s_b = 1$, $\mu(\mathsf{p}) = 0$ if reveals that $s_b = 0$, and $\mu(\mathsf{p}) = \frac{1}{2}$ to indicate that the price is uninformative; likewise for $\nu(\beta)$. In equilibrium these will turn out to be the only relevant cases. Thus $\mu, \nu : [0, 1] \to \{0, 1/2, 1\}$. We refer to $\mu(\mathsf{p})$ as the *price-information* about the bank's signal and to $\nu(\beta)$ as the *spread information.*

The Aftermarket Price. The equilibrium market price is determined by the aggregate number of investors' favorable signals. In our model this number is always revealed, either directly through investor demand or immediately after the float through trading activities. Thus write $\mathsf{p}^m(d)$ for the market price as a function of $d \in \{0, \ldots, N\}$, the number of high-signal investors. Appendix 2.7.1 fleshes out this argument and provides an extensive treatment of price formation.

Investors' Decisions and Expected Payoffs. We admit only symmetric pure strategies; thus all investors with the same signal take identical decisions. These can then be aggregated so that only three cases need to be considered: First, all investors subscribe, $B_{0,1}$, second, only high-signal investors subscribe, B_1, and third, no investor subscribes, $B_\emptyset$. Thus, the set of potential collective best replies is $\mathbb{B} := \{B_{0,1}, B_1, B_\emptyset\}$.

To compute his expected payoff, an investor has to account for the probability of receiving the security. There are two cases to consider. In the first, all investors buy, i.e. $B_{0,1}$. Thus, market demand is N and all investors receive the security with equal probability S/N. In the second case, only high-signal investors buy. If d investors (including oneself) buy, then each one receives the security with probability S/d. If overall demand d is smaller than the number of shares on offer, $d < \mathsf{S}$, the IPO fails and investors who ordered the security get it with probability 0.

Investors order the security whenever their expected payoff from doing so is non-negative. Suppose only high-signal investors buy, i.e. B_1. After observing the offer price, an investor's information set contains both his signal s_i and the information inferred from the offer price and spread, $\mu(\mathsf{p})$ and $\nu(\beta)$. Since signals are conditionally i.i.d., for every $V \in \mathbb{V}$ there is a different distribution over the number $d-1$ of others' favorable signals ($s_i = 1$), which we denote as $f(d-1|V)$. Thus investors' posterior distribution that the number of others' favorable signals is $d-1$ is given by

$$g(d-1|s_i, \mu(\mathsf{p}), \nu(\beta)) := \sum_{V \in \mathbb{V}} \mathsf{Pr}(V|s_i, \mu(\mathsf{p}), \nu(\beta)) \cdot f(d-1|V). \tag{2.1}$$

If only investors with favorable signals order, then for a high-signal investor, at price p his rational-expectation payoff from buying has to be non-negative,

$$\sum_{d=\mathsf{S}-1}^{N-1} \frac{\mathsf{S}}{d} \cdot (\mathsf{p}^m(d) - \mathsf{p}) \cdot g(d-1|s_i = 1, \mu(\mathsf{p}), \nu(\beta)) \geq 0. \tag{2.2}$$

Likewise for the respective low-signal investors when all investors order, $B_{0,1}$, in which case the summation runs from 1 to N, and $s_i = 1$ is replaced by $s_i = 0$.

Threshold Prices. Denote by $p_{s_i,\mu,\nu}$ the highest price that an investor with signal s_i, price information $\mu(\mathsf{p})$ and spread information $\nu(\beta)$ is willing to pay in equilibrium if all investors with signal $\tilde{s}_i \geq s_i$ order. If the issuer is uninformed, ν is replaced with a diamond, $\diamond$; if issuer and bank get the same signal *and* if the issuer signals his private information, $\mu(\mathsf{p})$ is replaced with a diamond to indicate that the price cannot reveal further information. Suppose, the issuer reveals information ν. Then $p_{1,1,\nu}$ is the highest (price-separating) price with B_1, $p_{1,\frac{1}{2},\nu}$ the highest (price-pooling) price with B_1, $p_{0,\frac{1}{2},\nu}$ the highest (price-pooling) price with $B_{0,1}$, and $p_{0,0,\nu}$ the highest (price-separating) price with $B_{0,1}$. Note that at all these prices investors are aware that the security price may drop (or rise) in the aftermarket and that they may not get the security. The threshold prices are formally derived in Appendix 2.7.2.

The Investment Bank's Expected Payoff. First consider case B_1. Variable d denotes the number of orders, i.e. the number of high-signal investors. If the true value is $V = 1$, we have

$$\mathsf{Pr}(d \geq \mathsf{S}|B_1) = \sum_{d=\mathsf{S}}^{N} \binom{N}{d} q^d (1-q)^{N-d}, \tag{2.3}$$

analogously for $V = 0$. Suppose the issuer gets no signal or its private signal s_f. A bank with signal s_b assigns probability $\alpha_{s_b,\nu}(\mathsf{S})$ to the event that at least S investors have the favorable signal. If the investment bank has signal s_b and spread information ν, then

$$\alpha_{1,\nu}(\mathsf{S}) = \sum_{d=\mathsf{S}}^{N} \binom{N}{d} \left(\mathsf{Pr}(V=1|s_b,\nu) \cdot q^d (1-q)^{N-d} + \mathsf{Pr}(V=0|s_b,\nu) \cdot (1-q)^d q^{N-d} \right) \tag{2.4}$$

If the bank charges a price at which only high-signal investors buy, its expected profit is

$$\Pi(\mathsf{p}|s_b,\nu,B_1) = \alpha_{s_b,\nu}(\mathsf{S}) \cdot \beta \mathsf{pS} - (1 - \alpha_{s_b\nu}(\mathsf{S})) \cdot C. \tag{2.5}$$

Consider now $B_{0,1}$, the case where the offer price is low enough so that all investors are willing to buy, irrespective of their signals. The offer never fails, thus payoffs are given by $\Pi(\mathsf{p}|B_{0,1}) = \beta\mathsf{pS}$. If the price is set so high that no investor buys, as in case $B_\emptyset$, a loss of C results with certainty.

The unconditional distribution over favorable signals is a composite of the two conditional distribution and thus bimodal. To obtain (approximate) closed form solutions

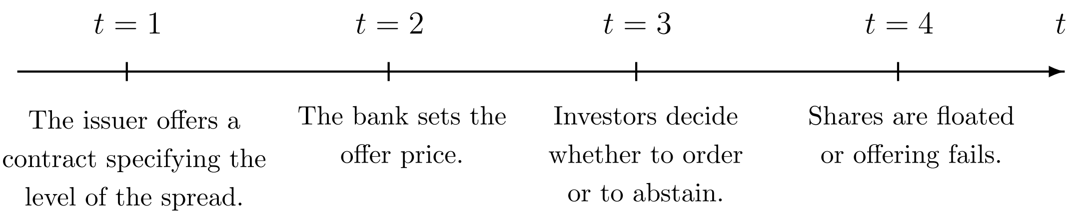

Figure 2.2: **The Timing of the Game.**

for success-probabilities and prices, we make two simplifying assumptions: the first simplifies computations, since the two modes of the distribution over favorable signals are centered around $N(1-q)$ and Nq. It allows us to get closed form solutions for success-probabilities. The second assumption ensures that we can analyze the two underlying conditional distributions separately.

Assumption 2.1 $\mathsf{S} = (1-q)N$.

For every signal quality q, there exists an $\tilde{N}(q)$ so that for all $N > \tilde{N}(q)$ the two conditional distributions over favorable signals generated by $V = 0$ and $V = 1$ do not 'overlap'. By standard results from statistics, a sufficient condition for $\tilde{N}(q)$ is given by $\tilde{N}(q) > 64q(1-q)/(2q-1)^2$.

Assumption 2.2 *The number of investors N is larger than $\tilde{N}(q)$.*

As a consequence of the second assumption we can apply the Law of Large Numbers and DeMoivre-Laplace's Theorem.[27] Since we assume that the IPO fails whenever $d < \mathsf{S}$, Assumption 2.1 implies, for instance, if the spread is uninformative, i.e. $\nu = 1/2$, then $\alpha_{0,\frac{1}{2}}(\mathsf{S}) = (2-q)/2$ and $\alpha_{1,\frac{1}{2}} = (1+q)/2$. In what follows we thus omit S. A consequence of the Law of Large Numbers is that $\mathsf{p}^m(d) \in \{0, 1\}$ for almost all values of d.[28]

Before proceeding we summarize the timing of the model in Figure 2.2. First, an (identically) informed issuer receives its signal and then all types of issuer offer a contract to a bank specifying the spread level. Second, the bank sets the offer

[27] For instance, the mode of a binomial distribution is generally not exactly symmetric. However, if N is large enough, we can apply DeMoivre-LaPlace ($0 < q \pm 2\sqrt{q(1-q)/N} < 1$) and employ the normal distribution instead. Thus we can treat each mode to be symmetric.

[28] To be more precise, for $d \gg N/2$, $\mathsf{p}^m(d) = 1$, and for $d \ll N/2$, $\mathsf{p}^m(d) = 0$. Thus to get interesting equilibria, it is necessary that S is strictly smaller than $N/2$. If it was not, an IPO where only $s_i = 1$ investors buy, would never be at risk of being overpriced as it fails in all overpriced cases.

price given the spread, the information contained therein, and its own signal. Third, investors decide whether to order given all available information. Finally, in case the IPO takes place, the number of favorable signals is revealed in the aftermarket and the price adjusts according to it.

In what follows proceed backwardly. In the next section we analyze the price setting of the investment bank given the level of the gross spread. In Section 2.4 we derive each type of issuer's choice of the spread in anticipation of the bank's price setting. Section 2.5 interprets the findings of Sections 2.3 and 2.4 and presents our main results on investment banking profits and differences in spreads between different classes of issuers.

2.3 Investment Banks' Equilibrium Price Choice

There are two cases to consider: First, spreads are uninformative or reflect the issuer's independent information. In that case, the investment bank plays a signaling game and needs to decide whether or not to reveal its private information. Second, in case of the identically informed issuer, spreads can reveal the bank's signal. Then the bank has no strategic decision problem but merely chooses the price that is optimal given all public information.

2.3.1 Uninformative Spreads or Spreads Reflecting the Issuer's Independent Signal

In the following we identify the conditions under which a profit maximizing investment bank will reveal its information through the offer price. A separating equilibrium is defined as *informationally efficient* since investors can derive the bank's signal from the offer price. In a pooling equilibrium information is shaded and thus it is *informationally inefficient*. In this case, investors decide only on the basis of their private signals. In what follows we take the information that may be contained in spreads, $\nu(\beta)$, as given. Separation and pooling thus always refers to prices.

The Equilibrium Concept and Selection Criteria. The equilibrium concept for this signaling game is, naturally, the Perfect Bayesian Equilibrium (PBE). A common problem with PBEs, however, is their multiplicity, stemming equilibria being supported by "unreasonable" out-of-equilibrium beliefs. The common way to overcome this problem is to apply an equilibrium selection rule such as the *Intuitive Criterion* (IC), introduced by Cho and Kreps (1987). We follow this line of research and consider only equilibria that do not fail the IC. All of these PBE selection devices favor separating over pooling equilibria. It will turn out, however, that in our framework under certain conditions pooling equilibria cannot be ruled out by the IC. Moreover, these pooling equilibria then Pareto dominate any separating equilibrium.[29] It would thus be unreasonable not to assume that these equilibria will be picked. Thus in what follows, we will only consider equilibria that *satisfy the IC* and among these, we consider those that are *Pareto efficient* for the agent who takes the signaling action. In this section this agent would be the bank.

A *pooling equilibrium* in prices is specified through (i) an equilibrium offer price p^* from which investors infer (ii) price-information $\mu = \frac{1}{2}$, and (iii) investors' best replies given their private signals, μ, and p^*. A *separating equilibrium* in prices is (i) a system of prices $\{\underline{\mathsf{p}}^*, \bar{\mathsf{p}}^*\}$ and price-information such that (ii) at $\mathsf{p}^* = \bar{\mathsf{p}}^*$, the *high separation price*, the price-information is that the bank has the favorable signal, $\mu = 1$, at $\mathsf{p}^* = \underline{\mathsf{p}}^*$, the *low separation price*, the price-information is that the bank has the low signal, $\mu = 0$, and (iii) investors' best replies given their private signals, μ, and $\bar{\mathsf{p}}^*$ or $\underline{\mathsf{p}}^*$. In both separating and pooling equilibria, for $\mathsf{p} \notin \{\bar{\mathsf{p}}^*, \underline{\mathsf{p}}^*\}$ or $\mathsf{p} \neq \mathsf{p}^*$, respectively, out-of-equilibrium public beliefs are chosen 'appropriately'. The following result is a straightforward consequence of signaling, the proof of which is in Appendix 2.7.4.

Lemma 2.1 (The Highest Possible Low Separating Price)
There exists no PBE (price-)separating offer price $\underline{\mathsf{p}}^* > p_{0,0,\nu}$.

In any separating equilibrium, therefore, the low price must be such that all investors buy, and the highest such separating price, given price-information $\mu = 0$, is $\underline{\mathsf{p}}^* = p_{0,0,\nu}$. In what follows we refer to $p_{0,0,\nu}$ as *the* low separation price.

[29]Here Pareto domination refers to the payoffs of the respective banks or issuers who take the decision first, not to investors who react.

Price-signaling equilibria in our setting come in one of three guises: The already mentioned separating equilibrium, a pooling equilibrium in which only high-signal investors buy, and a pooling equilibrium in which all investors buy. In the following, we characterize the conditions guaranteeing that only separating equilibria survive our selection criterion.

Fix a potential price $\mathsf{p} \in [p_{0,0,\nu}, p_{0,\frac{1}{2},\nu}]$, the interval of potential pooling prices at which all investors would buy.[30] Define $\phi_{1,\nu}(\mathsf{p})$ as the price at which the high-signal bank would be indifferent between charging a *risky* price $\phi_{1,\nu}(\mathsf{p})$ at which only high-signal investors buy, B_1, and a *safe* pooling price p with $B_{0,1}$ (all investors buy). Formally,

$$\alpha_{1,\nu}\beta\phi_{1,\nu}(\mathsf{p})\mathsf{S} - (1-\alpha_{1,\nu})\, C = \beta\mathsf{p}\mathsf{S} \quad \Leftrightarrow \quad \phi_{1,\nu}(\mathsf{p}) = \frac{\mathsf{p}}{\alpha_{1,\nu}} + \frac{1-\alpha_{1,\nu}}{\alpha_{1,\nu}}\frac{C}{\beta\mathsf{S}}. \tag{2.6}$$

Price $\phi_{0,\nu}(\mathsf{p})$ is defined analogously for the low-signal bank. Thus price $\phi_{s_b,\nu}(\mathsf{p})$ is the lowest risky price that a bank with signal s_b is willing to deviate to from safe price p.[31] In what follows we refer to $\phi_{1,\nu}(\mathsf{p})$ as the *high-signal bank's deviation price*, and to $\phi_{0,\nu}(\mathsf{p})$ as the *low-signal bank's deviation price.* It is straightforward to see that $\phi_{0,\nu}(\mathsf{p}) > \phi_{1,\nu}(\mathsf{p})$ for all $\mathsf{p} \in [p_{0,0,\nu}, p_{0,\frac{1}{2},\nu}]$, that is, the low-signal bank requires a higher price as compensation for risk taking. In addition, $\partial\phi_{s_b,\nu}(\mathsf{p})/\partial\mathsf{p} > 0$, $s_b \in \{0,1\}$, so the higher the pooling price, the higher the lowest profitable deviation price. Taking spread-information as given, we omit ν from the equilibrium specification. In what follows we analyze equilibria depending on two conditions on primitives.

Condition 1 The high-signal bank's deviation price from the highest safe pooling price is not higher than the highest separating price, $\phi_{1,\nu}(p_{0,\frac{1}{2},\nu}) \leq p_{1,1,\nu}$.

Condition 2 The low-signal bank's deviation price from the low separating price is not smaller than the highest risky pooling price, $\phi_{0,\nu}(p_{0,0,\nu}) \geq p_{1,\frac{1}{2},\nu}$.

[30]The order of prices is immediately obvious from Appendix 2.7.3.

[31]Deviation to a high, risky price can lead to increased overpricing, which is commonly perceived to be bad for a bank's reputation. Nanda and Yun (1997) analyze the impact of IPO mispricing on the market value of investment banks. They find that overpriced offerings result in decreased lead-underwriter market value. In our model, however, investors fully take into account that the offer price may drop in the aftermarket. Modelling such reputation effects would thus be contradictory in our setting.

Proposition 2.1 (Equilibrium Price Setting)

(a) If Condition 1 *and* Condition 2 *are both fulfilled then the unique PBE that satisfies the Intuitive Criterion is the separating equilibrium* $\{(\underline{\mathsf{p}}^* = p_{0,0,\nu}, \mu = 0, B_{0,1}); (\bar{\mathsf{p}}^* = \min\{p_{1,1,\nu}, \phi_{0,\nu}(p_{0,0,\nu})\},\ \mu = 1, B_1); (\mathsf{p} \neq \{\underline{p}^*, \bar{p}^*\}, \mu = 0, B_{0,1}$ *if* $\mathsf{p} \leq p_{0,0,\nu}, B_1$ *if* $p_{0,0,\nu} < \mathsf{p} \leq p_{1,0,\nu}, B_\emptyset$ *else*$)\}$.

(b) If Condition 1 *is not fulfilled then the only PBE that satisfies the Intuitive Criterium and Pareto efficiency is the pooling equilibrium* $\{(\mathsf{p}^* = p_{0,\frac{1}{2},\nu}, \mu = \frac{1}{2}, B_{0,1}); (\mathsf{p} \neq p_{0,\frac{1}{2},\nu}, \mu = 0, B_1$ *if* $\mathsf{p} \leq p_{1,0,\nu}, B_\emptyset$ *else*$)\}$ *in which all investors buy.*

(c) If Condition 2 *is not fulfilled then the only PBE that satisfies the Intuitive Criterium and Pareto Efficiency is the pooling equilibrium* $\{(\mathsf{p}^* = p_{1,\frac{1}{2},\nu}, \mu = \frac{1}{2}, B_1); (\mathsf{p} \neq p_{1,\frac{1}{2},\nu}, \mu = 0, B_1$ *if* $\mathsf{p} \leq p_{1,0,\nu}, B_\emptyset$ *else*$)\}$ *in which only high-signal investors buy.*

Interpretation of the Proposition. Condition 1, $\phi_{1,\nu}(p_{0,\frac{1}{2},\nu}) < p_{1,1,\nu}$, together with the IC is necessary and sufficient to rule out pooling equilibria in which all investors buy, irrespective of their signals. Condition 2, $\phi_{0,\nu}(p_{0,0,\nu}) > p_{1,\frac{1}{2},\nu}$, ensures that there is no pooling where only investors with "good news" buy, B_1. The IC itself ensures that the high-signal bank always charges the highest sustainable separating price. The high separation price $\bar{\mathsf{p}}^*$ is the minimum of $p_{1,1,\nu}$ and $\phi_{0,\nu}(p_{0,0,\nu})$. The bank cannot charge more than $p_{1,1,\nu}$ and it cannot credibly charge more than $\phi_{0,\nu}(p_{0,0,\nu})$ as otherwise the low-signal bank would deviate. Finally, since $\phi_{1,\nu}(p_{0,0,\nu}) < \phi_{1,\nu}(p_{0,\frac{1}{2},\nu}) < p_{1,1,\nu}$, the high-signal bank is willing to separate. If Condition 1 is violated not even the high-signal bank wants to take the risk of setting a price where only high-signal investors buy. A separating price pair with all investors buying at both prices cannot be an equilibrium. The bank charging the lower price always had an incentive to deviate to the higher price since the success probability remains unchanged. Pareto efficiency for banks together with the IC then ensures that the highest pooling price at which all investors buy results as unique equilibrium outcome. If Condition 2 is violated also the low-signal bank wants to set a high price at which only high-signal investors buy. A separating price pair with only high-signal investors buying at both prices cannot be an equilibrium. Again, the bank charging the lower price always had an incentive to deviate. Under efficiency only the highest such pooling price survives as the unique equilibrium. Finally, notice that it cannot be the case that both Condition 1 and Condition 2 are violated simultaneously.

2.3.2 Spreads Reflecting the Issuer's Identical Signal

The moment the identically informed issuer separates, the bank has no more control over the the price's signaling value. We signify this by including a diamond, $\diamond$, instead of $\mu(\mathsf{p})$ into prices, $p_{s_i,\diamond,\nu}$. The bank continues to choose the price that, given its private information, maximizes expected profit. However, a low-signal bank can no longer mimic a high-signal bank because investors have inferred the bank's signal form the spread.

Suppose the bank has signal $s_b = 0$ and spread-information is $\nu(\beta) = 0$. Then the highest price high-signal investors are willing to pay is $\mathsf{p}_{1,\diamond,0}$. This price is risky as only investors with signal $s_i = 1$ are willing to buy. Price $p_{0,\diamond,0}$ is the highest safe price at which all investors buy. However, if the spread is high enough, the risk of a failing IPO may still be outweighed by potential gains. If the spread β_0^s is large enough so that risky profits at $p_{1,\diamond,0}$ strictly exceed riskless profits at $p_{0,\diamond,0}$, i.e.

$$\alpha_{\diamond,0}\beta\mathsf{S}p_{1,\diamond,0} - (1 - \alpha_{\diamond,0})C > \beta\mathsf{S}p_{0,\diamond,0} \tag{2.7}$$

then the low-signal bank will choose risky price $p_{1,\diamond,0}$. The high-signal bank faces a similar choice: If the spread is too low, it would rather choose a safe price. Here, however, the highest riskless price is $p_{0,\diamond,1}$, as at this price investors with the low signal are willing to buy, given they believe that the bank's/issuer's signal is $s_b = 1$. So the high-signal bank only choose risky price $p_{1,\diamond,1}$ if spread β_1^s is high enough so that risky profits at $p_{1,\diamond,1}$ strictly exceed riskless profits at $p_{0,1}$, which is

$$\alpha_{\diamond,1}\beta\mathsf{S}p_{1,\diamond,1} - (1 - \alpha_{\diamond,1})C > \beta\mathsf{S}p_{0,\diamond,1}. \tag{2.8}$$

From (2.7) and (2.8) we can derive the respective threshold spreads

$$\beta_0^s = \frac{1-\alpha_{\diamond,0}}{\alpha_{\diamond,0}p_{1,\diamond,0} - p_{0,\diamond,0}}\,\frac{C}{S} \qquad \text{and} \qquad \beta_1^s = \frac{1-\alpha_{\diamond,1}}{\alpha_{\diamond,1}p_{1,\diamond,1} - p_{0,\diamond,1}}\,\frac{C}{S}. \tag{2.9}$$

It is straightforward to compute that $\beta_0^s - \beta_1^s = (2(1-q))^{-1}C/\mathsf{S} > 0$. Consequently, if spreads are separating and sufficiently large, then banks will set risky prices.

2.4 The Issuer's Strategic Choice of the Spread

As with the investment bank, the analysis is split into two parts. In the first, the issuer is uninformed and thus not involved in a strategic situation. He will set prices so as to make the bank set equilibrium prices that are revenue-maximizing. In the second part, the issuer does have private information. The issuer then has the power to play a signaling game. He anticipates the behavior of the investment bank, and thus he will set spreads strategically to maximize expected revenue.

2.4.1 Equilibrium Spreads if the Issuer is Uninformed

For the investment bank, the choice of equilibrium prices critically depends on Conditions 1 and 2 from Proposition 2.1. In the following we give an intuitive interpretation of the equilibrium outcome in terms of the gross spread, demonstrating how the spread affects these conditions. We then derive the uninformed issuer's decision about the spread level. As before we indicate that the issuer is uninformed by replacing ν with a diamond.

An Intuitive Characterization of the Equilibrium. The concept of deviation prices $\phi_{s_b,\diamond}$ is a convenient tool to describe restrictions. We will now reformulate Conditions 1 and 2 from Proposition 2.1 in terms of the gross spread β. This allows us to derive a simple linear descriptive characterization of the equilibrium. Consider first

Condition 1, $\phi_{1,\diamond}(p_{0,\frac{1}{2},\diamond}) \leq p_{1,1,\diamond}$. If β is so low that

$$\phi_{1,\diamond}(p_{0,\frac{1}{2},\diamond}) = \frac{p_{0,\frac{1}{2},\diamond}}{\alpha_{1,\diamond}} + \frac{1-\alpha_{1,\diamond}}{\alpha_{1,\diamond}} \frac{C}{\beta \mathsf{S}} > p_{1,1,\diamond} \tag{2.10}$$

then the separating equilibrium cannot be sustained and the pooling equilibrium in $p_{0,\frac{1}{2},\diamond}$ prevails. In other words, if the gross spread is rather low then the incentive to set a high price and take the risk of failure is reduced whereas the cost of failure remains unchanged. The threshold value for β s.t. not even the high-signal bank sets a risky price is given by

$$\beta^s_\diamond = \frac{1-\alpha_{1,\diamond}}{\alpha_{1,\diamond} p_{1,1,\diamond} - p_{0,\frac{1}{2},\diamond}} \frac{C}{\mathsf{S}} = \frac{1}{2} \frac{1}{2q-1} \frac{C}{N}. \tag{2.11}$$

Moreover, if β is so high that

$$\phi_{0,\diamond}(p_{0,0,\diamond}) = \frac{p_{0,0,\diamond}}{\alpha_{0,\diamond}} + \frac{1-\alpha_{0,\diamond}}{\alpha_{0,\diamond}} \frac{C}{\beta \mathsf{S}} < p_{1,\frac{1}{2},\diamond} \tag{2.12}$$

then a separating equilibrium, again, cannot be sustained and the pooling equilibrium in $p_{1,\frac{1}{2},\diamond}$ prevails. In this case the gross spread is so high that even the low-signal bank is willing to take the risk of failure and set a high price at which only high-signal investors buy. For the high-signal bank it becomes too costly to uphold separation, i.e. it would have to lower the high separation price so much that it prefers pooling. This threshold value for a pooling β is given by

$$\beta^p_\diamond = \frac{1-\alpha_{0,\diamond}}{\alpha_{0,\diamond} p_{1,\frac{1}{2},\diamond} - p_{0,0,\diamond}} \frac{C}{\mathsf{S}} = \frac{1}{2} \frac{q/(1-q)}{\alpha_0 p_{1,\frac{1}{2},\diamond} - p_{0,0,\diamond}} \frac{C}{N}. \tag{2.13}$$

Finally, there exists a $\hat{\beta}^s_\diamond \in [\beta^s_\diamond, \beta^p_\diamond]$ such that the deviation price of the low-signal bank is just $p_{1,1,\diamond}$, i.e. for values of β above $\hat{\beta}^s_\diamond$ the high-signal bank has to lower the high separation price in order to uphold separation. The following Corollary to Proposition 2.1 summarizes the above characterization; Figure 2.3 offers an illustration of the corollary.

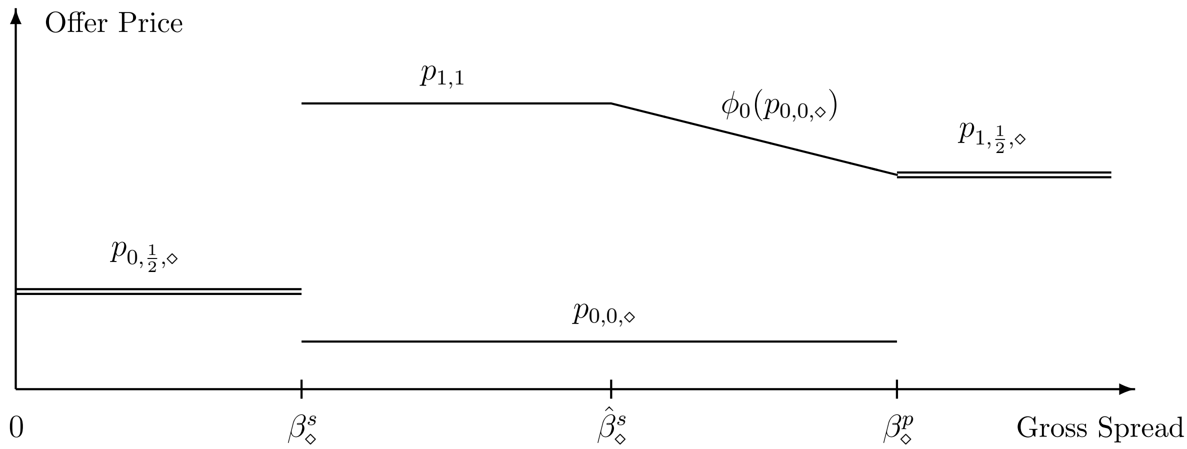

Figure 2.3: **Equilibrium Offer Prices at Different Levels of Uninformative Spreads.** For levels of the spread below $\beta^p_\diamond$, both types of banks pool in $\mathsf{p}^* = p_{0,\frac{1}{2},\diamond}$. For $\beta \in [\beta^s_\diamond, \hat{\beta}^s_\diamond]$ there is separation: The low-signal bank always sets $\underline{\mathsf{p}}^* = p_{0,0,\diamond}$, and the high-signal bank sets $\bar{\mathsf{p}}^* = p_{1,1,\diamond}$ for $\beta \in [\beta^s_\diamond, \hat{\beta}^s_\diamond]$ and $\bar{\mathsf{p}}^* = \phi_{0,\diamond}(p_{0,0,\diamond})$ for $\beta \in (\hat{\beta}^s_\diamond, \beta^p_\diamond)$. If $\beta \geq \beta^p_\diamond$ there is pooling in $p_{1,\frac{1}{2},\diamond}$.

Corollary 2.1 (Proposition 2.1 in Terms of the Gross Spread)
If spreads are uninformative and $\beta \in [\beta^s_\diamond, \beta^p_\diamond)$ then the unique equilibrium is the separating equilibrium stated in Proposition 2.1. If $\beta \in [\beta^s_\diamond, \hat{\beta}^s_\diamond]$ then $\bar{\mathsf{p}}^* = p_{1,1,\diamond}$, and if $\beta \in (\hat{\beta}^s_\diamond, \beta^p_\diamond)$ then $\bar{\mathsf{p}}^* = \phi_{0,\diamond}(p_{0,0,\diamond})$. If $\beta < \beta^s_\diamond$ then pooling in $p_{0,\frac{1}{2},\diamond}$ prevails. If $\beta \geq \beta^p_\diamond$ there is pooling in $p_{1,\frac{1}{2},\diamond}$.

Implicitly, we assumed a tie-breaking rule specifying that at spread thresholds $\beta^p_\diamond$ banks set a risky pooling price and at $\beta^s_\diamond$ they set separating prices. At $\beta^p_\diamond$ both types of banks set $p_{1,\frac{1}{2},\diamond}$ even though the low-signal bank is indifferent between $p_{1,\frac{1}{2},\diamond}$ and $p_{0,0,\diamond}$. The latter, however, cannot be an equilibrium: Suppose the low-signal bank sets $p_{0,0,\diamond}$. Then the issuer could raise the spread by an arbitrarily small $\epsilon > 0$ making the low-signal bank strictly prefer $p_{1,\frac{1}{2},\diamond}$. However, by lowering the spread to $\beta^p_\diamond + \epsilon/2$, the issuer could raise his profits and yet the low-signal bank would still set $p_{1,\frac{1}{2},\diamond}$, and so on. The similar reasoning applies to price setting at $\beta^s_\diamond$.

Strategic Choice of the Gross Spread. If the underlying issuer is uninformed, his strategic choice of spreads conveys no information. For every spread, however, the issuer knows the best response of both types of banks. Consequently, the issuer has to choose the level of the gross spread that maximizes his overall expected payoff. If he sets the spread too low, even a bank with favorable information chooses a low, riskless price. If spreads are high, the issuers get a smaller share of the revenue. Furthermore, for large spreads the high-signal bank may be unable to set a separating price. Pareto efficiency for the first mover (the issuer) ensures that out of all βs triggering separation or pooling, the issuer will always choose the *smallest* one. In particular, to get pooling in the riskless price $p_{0,\frac{1}{2},\diamond}$, the issuer can set spread 0. The issuer then has the choice between the following expected profits

$$(1-\beta^s_\diamond)\,\frac{\alpha_{1,\diamond}p_{1,1,\diamond}+p_{0,0,\diamond}}{2}\,\mathsf{S},\quad p_{0,\frac{1}{2},\diamond}\,\mathsf{S},\quad \text{and}\quad (1-\beta^p_\diamond)\,\frac{\alpha_{0,\diamond}+\alpha_{1,\diamond}}{2}\,p_{1,\frac{1}{2},\diamond}\,\mathsf{S} \tag{2.14}$$

in separation, low riskless pooling, and high risky pooling, respectively. To find the equilibrium spreads, one has to compare the issuers' payoffs for given equilibrium spreads. For given parameters q, C, N, the issuer will always choose the spread with maximal expected payoffs.

1. Pooling in $p_{1,\frac{1}{2},\diamond}$ is better than separation if

$$(1-\beta^p_\diamond)\,\frac{\alpha_{0,\diamond}+\alpha_{1,\diamond}}{2}\,p_{1,\frac{1}{2},\diamond}\,\mathsf{S} > (1-\hat{\beta})\,\frac{\alpha_{1,\diamond}p_{1,1,\diamond}+p_{0,0,\diamond}}{2}\,\mathsf{S} \;\Leftrightarrow\; \frac{C}{N} < R_1(q). \tag{2.15}$$

2. Pooling in $p_{1,\frac{1}{2},\diamond}$ is better than pooling in $p_{0,\frac{1}{2},\diamond}$ if

$$(1-\beta^p_\diamond)\,\frac{\alpha_{0,\diamond}+\alpha_{1,\diamond}}{2}\,p_{1,\frac{1}{2},\diamond} > p_{0,\frac{1}{2},\diamond} \;\Leftrightarrow\; \frac{C}{N} < R_2(q), \tag{2.16}$$

where $R_1(q)$ and $R_2(q)$ are derived by reformulating the inequalities; they depend on agents' signal quality. We state their precise form at the end of Appendix 2.7.4.

The above transformations make use of the closed form expressions for prices and success probabilities. Numerically it can easily be checked that $R_1(q) < R_2(q)$ for all $q \in (.6, 1)$, that is if high risky pooling is better than separation, it is also better than low, riskless pooling. We can now state the following proposition.

Proposition 2.2 (Gross Spreads with Uninformed Issuers)
Assume $C/N < R_1(q)$. There is a unique equilibrium that satisfies the IC and efficiency: The uninformed issuer offers a contract with $\beta = \beta^p_\diamond$ and both types of investment banks set pooling offer price $p_{1,\frac{1}{2},\diamond}$.

Proof: The choice of β follows directly from the comparison of the respective expected profits. The resulting price-setting by banks follows from Proposition 2.1. □

Interpretation of the Proposition. The ratio C/N measures the failure costs that the investment bank incurs per potential investor. If failure costs per investor are small, the level of the spread that triggers the high pooling price is most profitable. Proposition 2.2 is not a complete equilibrium analysis. However, if we impose the restriction that maximal spreads cannot exceed 10%, the corresponding ratio C/N will never exceed $R_1(q)$. Spreads above 10% are hardly observed,[32] thus for the empirically relevant parameter it is reasonable to restrict attention to the equilibrium characterized in the proposition.

2.4.2 Equilibrium Spreads if the Issuer is Independently Informed

Suppose now that the issuer gets his own, private signal, s_f, conditionally independent from all signals s_i and s_b. Then the signaling game has two stages. In the first, the issuer may or may not signal his information. Bank and investors incorporate this information. In the second stage, the bank chooses its equilibrium price, which may or may not reveal the bank's private signal. There are multiple different constellations imaginable:

1. The issuer pools in spreads and the bank separates in prices, pools in a risk-less price $p_{0,\frac{1}{2},\frac{1}{2}}$ or pools in a risky price $p_{1,\frac{1}{2},\frac{1}{2}}$.

[32] See for e.g. Chen and Ritter (2000).

2. The issuer separates in spreads, and

   (a) given a low-signal issuer $\nu = 0$, the bank separates in $p_{1,1,0}$ or $p_{0,0,0}$, pools in $p_{1,\frac{1}{2},0}$, or pools in $p_{0,\frac{1}{2},0}$, and

   (b) given a high-signal issuer signal, $\nu = 1$, the bank separates in $p_{1,1,1}$ or $p_{0,0,1}$, pools in $p_{1,\frac{1}{2},1}$, or pools in $p_{0,\frac{1}{2},1}$.

Equilibrium prices for given spread-information are covered by Proposition 2.1, it remains to analyze the issuer's optimal spread-choice.

Analogously to Corollary 2.1 we can determine threshold levels for the gross spread such that investment banks just set the low pooling price, separating prices, or the high pooling price. The lowest spread that induces banks to set the low pooling price is still $\beta = 0$. The two other threshold levels are denoted by β^s_ν for separation and β^p_ν for risky pooling.

The issuer's strategic choice of the spread follows from the comparison of the respective profits. As it turns out, there are no spread-separating equilibria — issuers always pool in the spread. Furthermore, the equilibrium pooling spread induces the bank to play a separating equilibrium in prices.

Proposition 2.3 (Gross Spreads with Independently Informed Issuers) *Assume $C/N < R_1(q)$. Then there exists a unique, spread-pooling equilibrium that satisfies the IC: Both types of issuers offer a contract with $\beta = \beta^s_{\frac{1}{2}}$ and investment banks separate by setting prices $p_{1,1,\frac{1}{2}}$ and $p_{0,0,\frac{1}{2}}$. At $p_{1,1,\frac{1}{2}}$, investors hold price-spread information $\mu = 1$ and $\nu = \frac{1}{2}$ and only investors with $s_i = 1$ buy. At $p_{0,0,\frac{1}{2}}$, investors hold price-spread information $\mu = 0$ and $\nu = \frac{1}{2}$ and all investors buy.*

Interpretation of the Proposition. To prove the claim we proceed counterfactual: We describe spreads and price-choices in a spread-separating equilibrium and show that a spread-separating situation is not incentive compatible for the low-signal issuer. He would always deviate and mimic the high-signal issuer. The intuition is straightforward: With separating spreads both low- and high-signal issuer prefer to play a spread that induces risky pooling prices; clearly a low-signal issuer would prefer the higher price though. Furthermore, the high-signal issuer cannot defend his position by setting different spreads, even when trying to play a spread that induces separation-pricing. We then show that only spread-pooling can result. There will be two candidates for spread-pooling: The first spread induces price-pooling, the second price-separation. However, only price-separation is IC-proof. Details of the proof are in Appendix 2.7.4.

2.4.3 Equilibrium Spreads if the Issuer is Identically Informed

If the issuer pools in spreads price setting by banks is as in Subsection 2.3.1. If the spread, however, is informative the bank has no strategic considerations to take care of in its optimal price choice: Its signal is the same as the issuer's who has just revealed his information. The high-signal bank does not have defend itself against deviation of the low-signal bank. In Subsection 2.3.2 we have already described banks' price setting.

We derived threshold spreads which induce the banks to choose risky prices, β^s_ν, $\nu \in \{0, 1\}$. The following order holds for spreads: $\beta^s_0 > \beta^p_\diamond > \beta^s_1 > \beta^s_{\frac{1}{2}} = \beta^s_\diamond$. If the issuer signals, the bank's pricing decision carries no informational value; we indicate this by substituting μ with a diamond, $\diamond$. Empirically, the gross spread almost never exceeds 10%, we thus restrict attention to parameter constellations so that $\beta^s_0 \leq 10\%$ (which implies $C/N < R_1(q)$). We can now show the following proposition.

Proposition 2.4 (Gross Spreads with Identically Informed Issuers)
Assume spreads do not exceed 10%. *Then there exists a unique, separating equilibrium that satisfies the IC and is first-mover efficient:*

(a) The identically informed low-signal issuer offers a contract with spread β^s_0 *and the investment bank sets price* $p_{1,\diamond,0}$. *Investors derive information* $\nu(\beta) = 0$, *and only those with signal* $s_i = 1$ *buy.*

(b) The identically informed high-signal issuer sets spread β^s_1 *and the bank sets price* $p_{1,\diamond,1}$. *Investors derive information* $\nu(\beta) = 1$, *and only those with signal* $s_i = 1$ *buy.*

Interpretation of the Proposition. We have derived the threshold spread in Subsection 2.3.2. The proof follows in three steps. First, we derive conditions under which each issuer is satisfied with the bank choosing a risky price at the proposed spreads β_0^s, β_1^s. The conditions will ensure that expected payoffs are higher than profits from setting zero spreads. In this step we will use that spreads must not exceed 10%. Second, we show that these spreads are proof to derivations, so that no type of issuer wants to mimic the other, and no type favors playing out of equilibrium spreads. Third we argue that with identically informed issuers there can be no pooling equilibrium (under the given restriction on β). Details are in Appendix 2.7.4.

To summarize, if the issuers have the same signal as the bank, they play a separation equilibrium in which both low- and high-signal issuer set spreads at which the bank sets a risky price. Notice that this is the only informationally efficient case where prices contain all existing information. In the case with uninformed issuers, banks pool in prices; in the case with independently informed issuers, spreads are pooling.

2.5 Results and Interpretation

We claimed to address two issues: First, why do investment banks make positive profits in a competitive market and second, why do VC backed IPOs have lower spreads. We deal with these issues in this section.

Even though it is hard to obtain data on banking profits, Chen and Ritter (2000) report that "investment bankers readily admit that the IPO business is very profitable" (p. 1105). Chen and Ritter argue that there are economies of scale in underwriting IPOs. They show that spreads do not differ in offerings that raise between \$20 million and \$80 million. Since banks at least break even in small offerings large offerings must be profitable. Megginson and Weiss (1991) were the first to report that spreads are significantly lower in VC backed IPOs than in non-VC backed IPOs. They show for a U.S. sample of 640 IPOs between 1983 and 1987 that gross spreads for VC backed issuers amount to 7.4% whereas they are 8.2% for non-VC backed issuers. Francis and Hasan (2001) find smaller but significant differences between spreads of VC and non-VC backed IPOs for their U.S. sample of 843 IPOs between 1990 and 1993 as well. In the following we show that our model can help explain both these phenomena.

In addition, we address implications of the model on the level of spreads when a commercial bank conducts the IPO of a former client. We finally show that our model is consistent with underpricing.

2.5.1 Positive Profits for Investment Banks

Equilibrium spreads allow investment banks positive profits. Issuers first announce the level of the spread and then banks choose prices at their discretion. They can always set a low, riskless price at which all investors buy, so that they receive their revenue share with certainty. Issuers, on the other hand, have a keen interest that banks set high prices, receiving the bulk of the revenue (almost always more than 90%). At high prices, however, only high-signal investors buy, making such prices risky. Spreads, therefore, have to be sufficiently high so that, banks are compensated for the risk of failure. This effect alone should leave them with zero expected profits. Moreover, spreads must be incentive compatible so that banks set high prices and do not deviate to a risk-free low price. An investment bank's expected profit, therefore, is always at least what they would gain by deviating to a low risk free price. Since we assume that the offering procedure itself causes no costs for the investment bank, it follows that investment banks earn positive profits.

Proposition 2.5 (Positive Profits for Investment Banks)
Investment banks enjoy positive profits that will not disappear in the face of competition.

Suppose a competing investment bank offered to conduct an IPO at a lower spread than specified in the contract the issuer offered initially. The issuer would not accept: even though he would get a higher fraction of the revenue, lower spreads trigger different equilibrium prices, leading to lower payoffs. In our model, banks have full discretion over the offer price. Issuers must, therefore, set incentive compatible spreads. In reality banks do not have full discretion over prices: many offerings fail because issuer and bank cannot agree on the offer price.[33] However, the qualitative result does not hinge

[33] See Busaba, Benveniste, and Guo (2001).

upon the assumption that banks have *full* discretion. Banks have a good deal of power when it comes to price setting, and this is all we need for the qualitative result to hold.

Issuers thus have no incentive to bargain for lower spreads. Competition, however, may take place in features of the IPO contract that we do not model. Chen and Ritter (2000), for example, report that over time the number of co-managers in IPOs and thus analyst coverage has increased over time as well. These findings complement our results nicely: Chen and Ritter state that, apparently, issuers cannot negotiate the spread; we find assert that they do not want to.

2.5.2 VC Issuers set Lower Spreads than Non-VC Issuers

In a strict sense, signals provide information about the asset's true liquidation value. In a wider sense, signals can be seen as information about market sentiment – market prices determine an investor's payoff, the true liquidation value only affects market prices through the distribution of signals. In this way it is not unreasonable to assert that an issuer is uninformed whereas banks and investors are informed. Certainly, some entrepreneurs have little experience with financial markets. Venture capitalists, on the other hand, are financial institutions and so they should be able to assess market sentiment. As the venture capitalist usually holds all relevant control rights, we interpret the independently informed issuer to be a VC-backed issuer. The uninformed issuer we interpret to be the single non-VC backed entrepreneur. In this model investment banks also hold private information. Before setting the offer price they closely interact with investors, for example during the road show, and thus are informed about the market's valuation of the firm on offer.

Proposition 2.6 (VC Backed Issuers set Lower Gross Spreads)
Assume spreads do not exceed 10%. Then VC backed issuers set lower levels of the gross spread than non-VC backed issuers.

Proof: If spreads do not exceed 10%, $C/N < R_1(q)$, then Proposition 2.2 states that uninformed issuers set spread level $\beta^p_\diamond$. From Proposition 2.3, the only IC-proof and efficient equilibrium spread with independently informed issuers is pooling spread $\beta^s_{\frac{1}{2}}$. Numerically it is straightforward to show that $\beta^s_{\frac{1}{2}} < \beta^p_{\frac{1}{2}} = \beta^p_\diamond$. □

A VC backed issuer holds private information before setting the spread level. An issuer with “good news” regards it as likely that the bank will also receive “good news”, and he wants the bank to transform this information to investors via separating prices. The high-signal issuer also considers it likely that there are enough high-signal investors such that the IPO will not fail at the risky separation price. An issuer with “bad news” will always mimic the high-signal issuer. Issuers receive almost always more than 90% of the offer revenue and thus have a strong interest in high prices. The reduction in offer price from signaling “bad news” is thus too costly for the low-signal issuer – even if he is forced to set the price-separation inducing spread as well.

2.5.3 Strong Commercial Banking Ties

Before going public many companies have strong, long-lasting ties with commercial banks, for instance through credit-financing. Thus if a commercial bank organizes a long-term client’s IPO, it is reasonable to believe that they truly have identical information. Only recently US regulators allowed commercial banks to offer investment banking services, including IPO underwriting. Our model predicts that bank spreads in such IPOs will be, on average, higher than in uninformed (non-VC backed) issuer’s or VC-backed IPOs. In particular, if bank’s and issuer’s signal are unfavorable, the issuer is willing to set a high spread so that the bank still chooses the risky price.

Proposition 2.7 (Identically Informed Issuers set Higher Average Spreads)

Assume spreads do not exceed 10%. Then on average, identically informed issuers set higher levels of the spread than uninformed (non-VC backed) or VC backed issuers.

When restricting the analysis to the ‘empirically relevant’ parameter space where spreads do not exceed 10% the uninformed issuer sets the spread such that investment banks with different signals pool in a high, risky offer price. If issuers receive the same signal as the bank, they separate in spreads. In both cases, however, spreads are so high that banks set the high risky price irrespective of their signals. If investors observe the low separation spread they infer that the issuer’s inside information is bad.

But then even the high risky price at which only high-signal investors buy is relatively low. The issuer thus has to set a relatively high spread to make the bank set the risky price. The opposite effect occurs when the high separation spread is set. However, the first effect dominates so that, on average, spreads are higher with identically informed issuers than in non-VC or VC backed offerings.

One may then conjecture that the low-signal issuer contemplates abandoning its commercial bank to look for an independent third-party bank. In equilibrium, it turns out, however, that this deviation is not profitable. Let a deviation be common knowledge. The resulting beliefs will render this deviation unprofitable. It is numerically straightforward to show that the high signal issuer would not be interested in this move: The best that can happen to him is that he is perceived as a high type issuer. But then even the highest expected payoff he'll get from working with an independent bank is, in expectation, lower than what he gets from his commercial bank. The reason is that with a third party, there is a risk that the bank gets an unfavorable signal and charges the low price.

Thus if the high type would not change, any change of banking-partner would be perceived as coming from a low type bank. It is straightforward to check numerically that the low-type bank then would not want to deviate either.

2.5.4 Underpricing

Even though this paper is not mainly concerned with explaining underpricing, in equilibrium the model is consistent with the empirical findings on first-day returns. In the context of this model underpricing is the difference between offer price and market price. We can establish the following proposition.

Proposition 2.8 (Underpricing)
(a) If spreads are uninformative but prices are separating then, on average, securities are underpriced. (b) If spreads are informative and all equilibrium prices risky then, on average, ordering the security yields zero profits.

The intuition behind the result is simple. Both types of investors only buy if their expected payoff is non-negative. At $p_{0,0}$ the low-signal investor just breaks even in expectation but the high-signal investor expects a strictly positive payoff. At $p_{1,1}$ the high-signal investor just breaks even and the low-signal investor abstains. Ex-ante expected payoffs are positive, hence underpricing. If, however, spreads are separating and prices risky, then only investors with the favorable signal buy. With them buying, prices are defined so that they yield zero profit.

2.6 Conclusion

We addressed and answered two puzzles of the IPO literature. First, why do investment banks earn positive profits in a competitive market as argued, for example, by Chen and Ritter (2000)? Second, why do banks receive lower gross spreads in VC backed than in non-VC backed IPOs as argued, for example, by Megginson and Weiss (1991)? Although investment bank compensation accounts for a substantial proportion of the cost of going public, it has hardly attracted theorists' attention. Our model is, to the best of our knowledge, the first to explain the level of the gross spread. We model the IPO procedure as a two-stage signaling game and find, first, that investment banks are left with profits and, second, that signaling considerations under different information constellations cause spreads to differ between classes of issuers.

In the first signaling stage, the issuer decides on the spread. Assuming that venture capitalists decide on spreads in VC backed offerings and that they have private information about market sentiment, they set different spreads than a non-VC backed issuer without much experience with financial markets. In the second stage, investment banks set offer prices given their private information, the spread, and the information contained in the spread. Issuers anticipate the bank's pricing decision and set the gross spread to maximize their expected revenue. Finally, investors decide whether to subscribe or refrain. They are aware of the IPO process details and order only if their expected profit is positive. In equilibrium, VC backed issuers offer lower spreads than non-VC backed issuers. Furthermore, issuers must offer high enough spreads to ensure that banks set high prices – allowing them substantial profits. Issuers act rationally, they do not want to cut spreads since lower spreads induce different, less favorable equilibrium offering prices.

2.7 Appendix

2.7.1 Aftermarket Price Formation

The finally prevailing market price depends on the number of positive signals about the value of the security. In determining the price we have to distinguish between cases B_1 and $B_{0,1}$.

Consider first case B_1. Since only high-signal investors buy, aggregated demand d indicates the number of high-signal investors. Suppose $d \geq \mathsf{S}$, i.e. the IPO is successful. Investors are assumed to take the aggregated information about signals into account and update their expectations accordingly. At this updated expectation all investors irrespective of their private signals are indifferent between selling and holding or buying and abstaining, depending on whether they own a security or not, respectively. The updated expectation thus becomes the aftermarket price, denoted by $\mathsf{p}^m(d)$. We will later show that case B_1 will occur at the high price of a separating equilibrium only, i.e. investors know that the bank's signal is $s_b = 1$. Taking further into account that the true value of the security is either 0 or 1, we can write $\mathsf{p}^m(d|\mu = 1) = \mathsf{Pr}(V = 1|d, \mu = 1)$. Using Bayes' rule, we can express the aftermarket price as

$$\mathsf{p}^m(d|\mu = 1) = \frac{\mathsf{Pr}(d|V = 1)\mathsf{Pr}(s_b = 1|V = 1)}{\mathsf{Pr}(d|V = 1)\mathsf{Pr}(s_b = 1|V = 1) + \mathsf{Pr}(d|V = 0)\mathsf{Pr}(s_b = 1|V = 0)}. \tag{2.17}$$

Due to the binomial structure of the prior distributions over signals, the conditional distribution for demand realization d is, for $V = 1$,

$$f(d|V = 1) := \mathsf{Pr}(d|V = 1) = \binom{N}{d} q^d (1-q)^{N-d}, \tag{2.18}$$

and for $V = 0$ analogously. The price-information about s_b is unambiguous in a separating equilibrium. We can therefore replace it with the conditional probability of the bank's signal being correct, which is q or $1 - q$. Bayes' rule yields

$$\mathsf{p}^m(d|\mu = 1) = \frac{qq^{2d-N}}{qq^{2d-N} + (1-q)(1-q)^{2d-N}}. \tag{2.19}$$

Consider now case $B_{0,1}$ in which all investors order the security, i.e. stated demand is N and securities are allocated at random. The demand is uninformative since it does not reveal the number of high-signal investors. Suppose that we are at the low price of a separating equilibrium. Note that high-signal investors expect the security to be of higher value than low-signal investors. Hence, there exists a price larger than the offer price, $\tilde{\mathsf{p}} > \mathsf{p}^*$ at which high-signal investors who were not allocated a security would be willing to buy the security, and low-signal investors would be willing to sell, in case they were allocated a security. Without modelling the price-finding procedure explicitly we assume that the following intermediate process takes place. Those high-signal investors who did not receive the security in the offering submit a unit market-buy-order. Those low-signal investors who obtained the security in the offering submit a unit market-sell-order. All other investors abstain. The number of investors who want to buy or to sell is denoted by $\tilde{d}$ and $\tilde{S}$, respectively. Aggregate demand of high-signal investors is then $d = \tilde{d} + \mathsf{S} - \tilde{S}$ and the market price p^m can be determined as before. The same procedure can be applied to determine the first period market clearing price in the case of a pooling equilibrium. The conditional expectation which determines the price, however, will then not contain the component about the signal of the investment bank.

2.7.2 Threshold Prices

Denote by $p_{s_i,\mu,\nu}$ the maximum price at which an investor with signal s_i and price-information μ and spread information ν buys, given all investors with $\tilde{s}_i \geq s_i$ buy. At this price the investor's expected return from buying the security is zero, normalizing outside investment opportunities accordingly.

Define $\psi(1|1,1,\nu) := \mathsf{Pr}(V = 1|s_i = 1, \mu = 1, \nu)$ and $\psi(0|1,1,\nu) := \mathsf{Pr}(V = 0|s_i = 1, \mu = 1, \nu)$. Consider now the structure of the conditional distribution $f(d-1|V)$. For $V = 1$, this is a binomial distribution over $\{0, \ldots, N-1\}$ with center $(N-1)q$, and likewise for $V = 0$ with center $(N-1)(1-q)$. Since by Assumption 2.2, N is 'large enough' for every q, $f(d-1|1) = 0$ for $d < N/2$ and $f(d|0) = 0$ for $d > N/2$.

When combining both $f(d-1|1)$ and $f(d-1|0)$, we obtain a bi-modal function. In $g(\cdot|s_i,\mu,\nu)$, investors' posterior distribution over demands, these are weighted with $\psi(1|s_i,\mu,\nu)$ and $\psi(0|s_i,\mu,\nu)$. Assumption 2.2 now satisfies two purposes. The first is to ensure that we pick N large enough, so that the two modes do not overlap. The second can be seen from the following lemma.

Lemma 2.2 *For any $q > \frac{1}{2}$, there exists a number of investors $N(q)$, such that $\mathsf{p}^m(d) \cdot g(d|s_i,\mu,\nu) \in \{0, g(d|s_i,\mu,\nu)\}$ almost everywhere.*

The lemma states that market prices are mostly 0 or 1, if they are not, then the weight of this demand is negligible. To see this consider the following heuristic argument.

Proof: $\mathsf{p}^m(d)$ is a s-shaped function in d, given by equation (2.19). For large N, $\mathsf{p}^m(d) \in \{0,1\}$ almost everywhere. Define $\mathbb{I}^*$ as the interval of d around $N/2$ s.t. for $d \in \mathbb{I}^*$ we have $\mathsf{p}^m(d) \notin \{0,1\}$. $\mathsf{p}^m(d)$ is multiplied with density $g(d|s_i,\mu,\nu)$, which peaks at $(N-1)(1-q)$ and $(N-1)q$. For N increasing $\mathbb{I}^*/N \to 0$ and the bi-modal distribution becomes more centered around $(N-1)(1-q)$ and $(N-1)q$. Hence, for every q there is an $(N-1)(q)$ such that for $d \in \mathbb{I}^*$, $g(d|s_i,\mu,\nu) \cdot \mathsf{p}^m(d) = 0$, i.e. the weight on $\mathsf{p}^m(d) \notin \{0,1\}$ can be made arbitrarily small. $\square$

Using Lemma 2.2 we can determine the threshold prices as follows. Consider first $p_{1,1,\nu}$.

$$\begin{aligned} 0 &= (1-p_{1,1,\nu}) \sum_{d=N/2}^{N-1} \frac{\mathsf{S}}{d+1}\, g(d-1|1,1,\nu) - p_{1,1,\nu} \sum_{d=\mathsf{S}-1}^{N/2} \frac{\mathsf{S}}{d+1}\, g(d-1|1,1,\nu) \\ \Leftrightarrow p_{1,1,\nu} &= \frac{\sum_{d=N/2}^{N-1} \frac{\mathsf{S}}{d+1}\, g(d-1|1,1,\nu)}{\sum_{d=\mathsf{S}-1}^{N-1} \frac{\mathsf{S}}{d+1}\, g(d-1|1,1,\nu)}. \end{aligned} \tag{2.20}$$

For $d > N/2$, $g(d-1|s_i,\mu,\nu) = \psi(1|s_i,\mu,\nu) f(d-1|1)$ and for $d < N/2$, $g(d-1|s_i,\mu,\nu) = \psi(0|s_i,\mu,\nu) f(d-1|0)$. Define

$$\Sigma_0 := \sum_{d=\mathsf{S}-1}^{N/2} \frac{f(d-1|0)}{d+1} \quad \text{and likewise} \quad \Sigma_1 := \sum_{d=N/2}^{N-1} \frac{f(d-1|1)}{d+1}, \quad \text{and} \quad \sigma := \Sigma_0/\Sigma_1.$$

Also write $\ell(\mu,\nu) := \psi(0|1,\mu,\nu)/\psi(1|1,\mu,\nu)$. Thus for the combination of signal s_i, price-information μ and spread information ν with B_1 we can write

$$p_{1,1,\nu} = (1+\sigma\ell(1,\nu))^{-1} \quad \text{and likewise} \quad p_{1,\frac{1}{2},\nu} = (1+\sigma\ell(\tfrac{1}{2},\nu))^{-1}. \tag{2.21}$$

Consider now the case for $p_{0,0,\nu}$. At this price all agents receive the security with equal probability and we sum from 0 to $N-1$. Thus

$$0 = (1-p_{0,0,\nu}) \sum_{d=N/2}^{N-1} \frac{\mathsf{S}}{N}\, g(d-1|0,0,\nu) - p_{0,0,\nu} \sum_{d=0}^{N/2} \frac{\mathsf{S}}{N}\, g(d-1|0,0,\nu) \Leftrightarrow p_{0,0,\nu} = \psi(1|0,0,\nu). \tag{2.22}$$

Likewise we have

$$p_{0,\frac{1}{2},\nu} = \psi(1|0,\tfrac{1}{2},\nu). \tag{2.23}$$

2.7.3 Approximate Closed Form Solutions

We will now derive *approximate* closed form solutions so that we can solve our model analytically. In this appendix we let d denotes the number of *other* investors with favorable information — this contrasts the exposition of the main text, but it simplifies the notation here. First consider the strategy of agent number N. There are $N-1$ other investors. Given that he invests and the true value is, say, $V=1$, then by the law of large numbers, demand/the number of favorable signals will always be larger than $N/2$. Furthermore, the market price is almost surely $p^m(d)=1$. If d others order, then when buying he gets the asset with probability $1/(d+1)$. Thus his payoff for price p

$$(1-\mathsf{p}) \sum_{d=(1-q)N-1}^{N-1} \frac{1}{d+1}\binom{N-1}{d} q^d (1-q)^{N-1-d} =$$
$$(1-\mathsf{p}) \sum_{d=N/2}^{N-1} \frac{1}{d+1}\binom{N-1}{d} q^d (1-q)^{N-1-d}. \tag{2.24}$$

To compute the sum we proceed in a similar manner as one would to compute the expected value of a binomial distribution: First observe that because N is large,

$$\sum_{d=N/2}^{N-1} \frac{1}{d+1}\binom{N-1}{d} q^d (1-q)^{N-1-d} = \sum_{d=0}^{N-1} \frac{1}{d+1}\binom{N-1}{d} q^d (1-q)^{N-1-d} \tag{2.25}$$

Then we can compute

$$\begin{aligned}
&\sum_{d=0}^{N-1}\frac{1}{d+1}\binom{N-1}{d}q^d(1-q)^{N-1-d} = \frac{1}{qN}\sum_{d=0}^{N-1}\frac{N!}{(N-d)!(d+1)!}\,q^{d+1}(1-q)^{N-1-d}\\
&= \frac{1}{qN}\left(\sum_{l=0}^{N}\binom{N}{l}q^l(1-q)^{N-l}-\binom{N}{0}q^0(1-q)^{N-0}\right)\\
&= \frac{1}{qN}(1-(1-q)^N). \qquad (2.26)
\end{aligned}$$

In the second step we made a change of variable, $l = d+1$, but through this change, we had to subtract the element of the sum for $l = 0$. Consequently, for large N, we can say that

$$\sum_{d=N/2}^{N-1}\frac{1}{d+1}\binom{N-1}{d}q^d(1-q)^{N-1-d} \approx \frac{1}{q_iN}. \qquad (2.27)$$

Using the same arguments, we could also show that

$$\sum_{d=0}^{N-1}\frac{1}{d+1}\binom{N-1}{d}q^{N-1-d}(1-q)^d \approx \frac{1}{(1-q)N}. \qquad (2.28)$$

Use now familiar notation to denote the combination of private and public beliefs $\phi_{s,\mu}$. For the time being, assume the issuer is uninformed so that ν is replaced with a diamond. Recall that we can write $p_{1,1,\diamond}$ as

$$p_{1,1,\diamond} = \left(1+\ell(1,\diamond)\,\frac{\Sigma_0}{\Sigma_1}\right)^{-1}. \qquad (2.29)$$

What we now need to find is a closed form for

$$\Sigma_0 = \sum_{d=N(1-q)-1}^{N/2}\frac{1}{d+1}\binom{N-1}{d}q^{N-1-d}(1-q)^d. \qquad (2.30)$$

For increasing N one can see that $\frac{1}{d+1}\binom{N-1}{d}q^{N-1-d}(1-q)^d$ gets *numerically* symmetric around $(1-q)N-1$. Thus we can express

$$\begin{aligned}\Sigma_0 &= \frac{1}{2}\sum_{d=0}^{N/2}\frac{1}{d+1}\binom{N-1}{d}q^{N-1-d}(1-q)^d = \frac{1}{2}\sum_{d=0}^{N}\frac{1}{d+1}\binom{N-1}{d}q^{N-1-d}(1-q)^d \\ &\approx \frac{1}{2}\frac{1}{(1-q)N}. \end{aligned} \tag{2.31}$$

Assembling, we obtain

$$p_{1,1} = \left(1+\ell(1,\diamond)\,\frac{\Sigma_0}{\Sigma_1}\right)^{-1} \approx \left(1+\frac{(1-q)^2}{q^2}\frac{qN}{2(1-q)N}\right)^{-1} = \frac{2q}{1+q} \equiv \frac{q}{\alpha_{1,\diamond}}. \tag{2.32}$$

Equivalently, we get

$$p_{1,\frac{1}{2},\diamond} \approx \left(1+\frac{1-q}{q}\frac{qN}{2(1-q)N}\right)^{-1} = \frac{2}{3}, \text{ and } p_{0,1,\diamond} \approx \frac{1-q}{\alpha_{0,\diamond}}. \tag{2.33}$$

The information content of a high pooling price is 1/2, and knowing this information, the probability of the offering being successful is 3/4. Thus the interpretation of risky prices is thus the ratio of the expected liquidation value given price- and spread-information to the share of successful offerings given this information

$$p_{1,\mu,\nu} = \frac{\mathsf{Pr}(V=1 \mid \mu,\nu)}{\mathsf{Pr}(\text{IPO successful} \mid \mu,\nu)}. \tag{2.34}$$

2.7.4 Omitted Proofs

Proof of Lemma 2.1

Suppose $\underline{\mathsf{p}}^* > p_{0,0,\nu}$. At this price only high-signal investors buy. A high-signal bank will always set a price where at least investors with signal $s_i = 1$ buy. Hence, investors with signal $s_i = 1$ buy at both prices $\underline{\mathsf{p}}^*$ and $\bar{\mathsf{p}}^*$. A low-signal bank can now increase its payoff by setting a higher price as $\alpha_{0,\nu}$ is not affected by this, a contradiction. □

Proof of Proposition 2.1

(a) First we will argue that given Conditions 1 and 2 the only separating equilibrium surviving the Intuitive Criterion (IC) is the one outlined in Proposition 2.1(a). Then we will argue that pooling cannot occur.

Step 1 (*Separating*) First observe that there cannot be a separating price $\bar{\mathsf{p}}^*$ where investors choose $B_{0,1}$ because otherwise the low-signal bank would deviate to this price. Note that no separating price with $\bar{\mathsf{p}}^* > \phi_{0,\nu}(p_{0,0,\nu})$ can exist because at this price, the low-signal bank would prefer to deviate. No price $\bar{\mathsf{p}}^* > p_{1,1,\nu}$ can exist since not even investors with $s_i = 1$ would buy. Furthermore, $\bar{\mathsf{p}}^* \geq \phi_{1,\nu}(p_{0,0,\nu})$ must be satisfied since otherwise the high-signal bank would prefer to deviate to $p_{0,0,\nu}$. Finally no price $\bar{\mathsf{p}}^*$ below $p_{1,0,\nu}$ is reasonable because the high-signal bank would then deviate to this price. Take $\tilde{p}$, with $\max\{\phi_{1,\nu}(p_{0,0,\nu}), p_{1,0,\nu}\} \leq \tilde{p} \leq \min\{p_{1,1,\nu}, \phi_{0,\nu}(p_{0,0,\nu})\}$. Note that such a $\tilde{p}$ always exists as long as $\phi_{1,\nu}(p_{0,0,\nu}) \leq p_{1,1,\nu}$ and $p_{1,0,\nu} \leq \phi_{0,\nu}(p_{0,0,\nu})$. The conditions stated in Proposition 2.1 ensure this is the case because $\phi_{1,\nu}(p_{0,\frac{1}{2},\nu}) > \phi_{1,\nu}(p_{0,0,\nu})$ and $p_{1,\frac{1}{2},\nu} > p_{1,0,\nu}$.

We analyze the candidate separating equilibrium

$$\begin{aligned}&\{(\underline{\mathsf{p}}^* = p_{0,0,\nu}, \mu = 0, B_{0,1}); (\bar{\mathsf{p}}^* = \tilde{p}, \mu = 1, B_1);\\&(\mathsf{p}^* \notin \{\underline{\mathsf{p}}^*, \bar{\mathsf{p}}^*\}, \mu = 0, B_{0,1} \text{ if } \mathsf{p} \leq p_{0,0,\nu}, B_1 \text{ if } p_{0,0,\nu} < \mathsf{p} \leq p_{1,0,\nu}, B_\emptyset \text{ else})\}.\end{aligned}$$

By definition of $\phi_{0,\nu}(p_{0,0,\nu})$ it holds that

$$\beta p_{0,0,\nu}\mathsf{S} = \alpha_{0,\nu}\beta\phi_{0,\nu}(p_{0,0,\nu})\mathsf{S} - (1-\alpha_{0,\nu})C > \alpha_{0,\nu}\beta\tilde{p}\mathsf{S} - (1-\alpha_{0,\nu})C \quad (2.35)$$

so that the low-signal bank would not deviate to $\tilde{p}$. Since $\max\{\phi_{1,\nu}(p_{0,0,\nu}), p_{1,0,\nu}\} \leq \tilde{p}$, the high-signal bank would also not deviate. Hence this is a PBE.

Now consider the application of the IC. Suppose a high separation price $\bar{\mathsf{p}} = \tilde{\tilde{p}}$ with $\tilde{p} < \tilde{\tilde{p}} \leq \min\{p_{1,1,\nu}, \phi_{0,\nu}(p_{0,0,\nu})\}$ is observed. This price is equilibrium dominated for a bank with $s_b = 0$ by definition of $\phi_{0,\nu}(p_{0,0,\nu})$. The low-signal bank can therefore be excluded the set of potential deviators. The only remaining agent is the high-signal bank. The best response of investors with signal $s_i = 1$ then is to buy at $\bar{\mathsf{p}} = \tilde{\tilde{p}}$, i.e. B_1. Hence the PBE with $\bar{\mathsf{p}}^* = \tilde{p}$ does not survive the IC. Applying this reasoning repeatedly, all separating prices with $\bar{\mathsf{p}} < \min\{p_{1,1,\nu}, \phi_{0,\nu}(p_{0,0,\nu})\}$ can be eliminated.

Step 2 (*Pooling with* $B_{0,1}$) For all investors to buy we must have $\mathsf{p} \leq p_{0,\frac{1}{2},\nu}$. Suppose there was deviation to $\mathsf{p} = \phi_{1,\nu}(p_{0,\frac{1}{2},\nu}) < \phi_{0,\nu}(p_{0,\frac{1}{2},\nu})$. For the low-signal bank this would not be profitable by definition of $\phi_{0,\nu}(p_{0,\frac{1}{2},\nu})$. But for some beliefs about the signal of the bank and corresponding best responses, investors with $s_b = 1$ could be better off. The best response for investors with beliefs on the remaining set of types, i.e. $\mu = 1$, however, is B_1 as we have $\phi_{1,\nu}(p_{0,\frac{1}{2},\nu}) < p_{1,1,\nu}$. Hence, applying the IC, there cannot be a pooling equilibrium with $B_{0,1}$.

Step 3 (*Pooling with* B_1) We must have $\mathsf{p} \leq p_{1,\frac{1}{2},\nu}$. Since $\phi_{0,\nu}(p_{0,0,\nu}) > p_{1,\frac{1}{2},\nu}$, the low-signal bank would prefer to deviate to $p_{0,0,\nu}$, hence this cannot be an equilibrium.

(b) We will first argue that if Condition 1 is not fulfilled each separating equilibrium is Pareto dominated by pooling in the risk-less price. Then we will show that also a pooling price at which only high-signal investors buy is Pareto dominated. We will finally argue that among all PBE pooling equilibrium prices at which all investors buy only the one outlined in Proposition 2.1 is Pareto efficient.

Step 1 (*Separating*) If Condition 1 is not fulfilled we have

$$\beta p_{0,\frac{1}{2},\nu}\mathsf{S} = \alpha_{1,\nu}\beta\phi_{1,\nu}(p_{0,\frac{1}{2},\nu})\mathsf{S} - (1-\alpha_{1,\nu})C > \alpha_{1,\nu}\beta p_{1,1,\nu}\mathsf{S} - (1-\alpha_{1,\nu})C \quad (2.36)$$

so the high-signal bank prefers pooling in $p_{0,\frac{1}{2},\nu}$ to the highest possible separation price $p_{1,1,\nu}$. Likewise, since $p_{0,\frac{1}{2},\nu} > p_{0,0,\nu}$ the risk-free pooling price is Pareto dominating for the $s_b = 0$ bank. Thus separation is always Pareto dominated and deselected.

Step 2 (*Pooling with* B_1) Since the high-signal bank can profitably deviate from $p_{1,1,\nu}$ it will and can do so from $p_{1,\frac{1}{2},\nu} < p_{1,1,\nu}$. Pooling with B_1 can thus be no equilibrium.

Step 3 (*Pooling with* $B_{0,1}$) Not even the high-signal bank wants to set a price where only high-signal investors buy. Candidate prices for an equilibrium are thus only prices with $B_{0,1}$. Consider $\mathsf{p} = \tilde{p} < p_{0,\frac{1}{2},\nu}$. Since both types of banks would prefer $\mathsf{p} = \tilde{\tilde{p}}$ with $\tilde{p} < \tilde{\tilde{p}} < p_{0,\frac{1}{2},\nu}$ Pareto efficiency prescribes that investors must hold $\mu = \frac{1}{2}$ and thus all investors will buy at $\mathsf{p} = \tilde{\tilde{p}}$. Applying this reasoning repeatedly, all prices with $\mathsf{p} < p_{0,\frac{1}{2},\nu}$ can be eliminated.

(c) We will first argue that if Condition 2 is not fulfilled every separating equilibrium is Pareto dominated. We will then argue that the only pooling equilibrium in which only high-signal investors buy is the one outlined in Proposition 2.1. We finally show that pooling in a price where all investors buy cannot be an equilibrium.

Step 1 (*Separating*) Since Condition 2 does not hold we have

$$\beta p_{0,0,\nu}\mathsf{S} = \alpha_{0,\nu}\beta\phi_{0,\nu}(p_{0,0,\nu})\mathsf{S} - (1-\alpha_{0,\nu})C < \alpha_{0,\nu}\beta p_{1,1,\nu}\mathsf{S} - (1-\alpha_{0,\nu})C \quad (2.37)$$

so the low-signal bank will mimic the high-signal bank at any price $\tilde{p} \geq \phi_{0,\nu}(p_{0,0,\nu})$. To uphold separation the high-signal bank must lower its price below $\phi_{0,\nu}(p_{0,0,\nu}) < p_{1,\frac{1}{2},\nu}$. However, a high separation price below $p_{1,\frac{1}{2},\nu}$ cannot be an efficient equilibrium since both types of banks would prefer pooling price $p_{1,\frac{1}{2},\nu}$. There can thus be no separating equilibrium.

Step 2 (*Pooling with B_1*) From Step 1 we know that both types of banks prefer pooling in $\tilde{p} \in [\phi_{0,\nu}(p_{0,0,\nu}), p_{1,\frac{1}{2},\nu}]$ even to the separating equilibrium with the highest possible $\bar{\mathsf{p}}$. Consider the candidate pooling price $\tilde{\tilde{p}}$ with $\tilde{p} < \tilde{\tilde{p}} < p_{1,\frac{1}{2},\nu}$. Since both types prefer $\tilde{\tilde{p}}$ to $\tilde{p}$ efficiency prescribes $\mu = 0.5$ and thus $\tilde{p}$ cannot be an equilibrium. Applying this reasoning repeatedly, all prices with $\mathsf{p} < p_{1,\frac{1}{2},\nu}$ can be eliminated. The only pooling equilibrium surviving is thus the one depicted in Proposition 2.1.

Step 3 (*Pooling with $B_{0,1}$*) Suppose that $p_{0,\frac{1}{2},\nu}$ was an equilibrium, supported by out-of-equilibriums belief that any deviation is by a low-signal bank. Then consider a deviation to $\phi_{1,\nu}(p_{0,\frac{1}{2},\nu})$. Naturally, $\phi_{1,\nu}(p_{0,\frac{1}{2},\nu}) < \phi_{0,\nu}(p_{0,\frac{1}{2},\nu})$, and thus, applying the IC, this deviation can only be triggered by a high-signal bank. It is straightforward to check that, numerically, a violation of Condition 2 implies that

Condition 1 holds, i.e. $\phi_{1,\nu}(p_{0,\frac{1}{2},\nu}) < p_{1,1,\nu}$. Furthermore,

$$\phi_{1,\nu}(p_{0,\frac{1}{2},\nu}) = \frac{p_{0,\frac{1}{2},\nu}}{\alpha_{1,\nu}} + \frac{1-\alpha_{1,\nu}}{\alpha_{1,\nu}} \frac{C}{\beta \mathsf{S}}, \tag{2.38}$$

which is increasing in costs C. The largest C so that Condition 2 just holds $\underline{C} = \beta\mathsf{S}(\alpha_{0,\nu} p_{0,\frac{1}{2},\nu} - p_{0,0,\nu}/(1-\alpha_{0,\nu}))$. Any C violating Condition is smaller than $\underline{C}$. Numerically, then $\phi_{1,\nu}(p_{0,\frac{1}{2},\nu}) < p_{1,\frac{1}{2},\nu}$, thus efficiency holds. □

Proof of Proposition 2.3

To prove this result, we proceed in five steps: In the first we derive the issuer's optimal spread choice under the assumption that spreads are separating. The issuer then chooses the spread that maximizes his payoff; the spread will induce the bank to set either a separating or a pooling price. This step serves as benchmark for comparing deviation payoffs. The first-mover Pareto efficiency requirement ensures, that in any spread-separating equilibrium, the low-signal issuer will always set his preferred spread, irrespective of the high-signal issuer's choice. In the second step, we argue that the low-signal issuer will always mimic the high-signal issuer's optimal choice. In the third step we show that the high-signal issuer cannot defend separation in spreads by choosing a different level of the spread. This step consists of three sub-steps in which we show that neither constellation (price-separation inducing or price-pooling inducing spreads) can be upheld. In the fourth step we show that pooling in spreads is indeed an equilibrium, but we also show that there can be two equilibria. In the fifth step we argue that only the price-separation inducing spread satisfies the Intuitive Criterion (IC).

The results can only be obtained numerically: When comparing different payoffs, the decisive equations are complicated polynomials, that cannot be expressed in an appealing simple form. Explicit solutions, however, can be obtained from the authors upon request. Furthermore, throughout the proof we use the restriction that $\beta < 10\%$.

Table 2.1 describes how an issuer computes his expected payoffs. In this proof we let β^s_ν denote the spread that yields separation given spread information ν.[34]

[34]We emphasize that this is *not* the same as the spreads defined in Subsection 2.3.2. Nevertheless,

$\mathsf{Pr}(V\|s_f=1)$	q $V=1$			$1-q$ $V=0$		
	$\mathsf{Pr}(s_b\|V)$	Price	Pr(IPO successful$\|V$)	$\mathsf{Pr}(s_b\|V)$	Price	Pr(IPO successful$\|V$)
$s_b=1$	q	$p_{1,1,1}$	1	$1-q$	$p_{1,1,1}$	$\frac{1}{2}$
$s_b=0$	$1-q$	$p_{0,0,1}$	1	q	$p_{0,0,1}$	1

Table 2.1: **Probabilities Summary.** Equilibrium price choice, signal probabilities, and success-probabilities in price-separating equilibria. Issuer's expected profit from charging, e.g., β_1^s is $(1-\beta_1^s)\cdot(q\cdot(qp_{1,1,1}+(1-q)p_{0,0,1})+(1-q)\cdot((1-q)p_{1,1,1}\frac{1}{2}+qp_{0,0,1}))=(q^2+(1-q)^2/2)p_{1,1,1}+2q(1-q)p_{0,0,1}$.

Step 1: Suppose first that the spread is separating and indicates $s_f=1$, so that $\nu=1$. The issuer has the choice between expected profits in separation, low riskless pooling, and high risky pooling. For given parameters q,C,N, the issuer will always choose the spread with maximal expected payoffs. The following two inequalities always holds when $C/N<R_1(q)$.

1. Pooling in $p_{1,\frac{1}{2},1}$ is always better than separation in $p_{1,1,1}$ and $p_{0,0,1}$ as

$$(1-\beta_1^s)\left((q^2+\frac{(1-q)^2}{2})p_{1,1,1}+2q(1-q)p_{0,0,1}\right)>(1-\beta_1^p)\,\alpha_1\,p_{1,\frac{1}{2},1}\tag{2.39}$$

2. Pooling in $p_{1,\frac{1}{2},1}$ is better than pooling in $p_{0,\frac{1}{2},1}$ if

$$(1-\beta_1^p)\,\alpha_{1,1}\,p_{1,\frac{1}{2},1}>p_{0,\frac{1}{2},1}\tag{2.40}$$

Suppose now that the spread triggers $\nu=0$. Again, we have to compare expected profits. All the inequalities hold if we restrict $C/N<R_1(q)$.

1. Pooling in $p_{1,\frac{1}{2},0}$ is better than separation in $p_{1,1,0}$ and $p_{0,0,0}$ if

$$(1-\beta_0^p)\left((1-q)+\frac{q}{2}\right)p_{1,\frac{1}{2},0}>(1-\beta_0^s)\left(\frac{3}{2}q(1-q)p_{1,1,0}+(q^2+(1-q)^2)p_{0,0,0}\right)\tag{2.41}$$

2. Pooling in $p_{1,\frac{1}{2},0}$ is better than pooling in $p_{0,\frac{1}{2},0}$ if

$$(1-\beta_0^p)\,\alpha_{0,0}\,p_{1,\frac{1}{2},0}>p_{0,\frac{1}{2},0}\tag{2.42}$$

for the purposes of exposition in the proof this notation is best; since the spreads β_ν^s as defined here are no equilibria, there should be no confusion. Details of the β's used in this proof are placed after the proof.

Thus if spreads are separating, irrespective of the spread-information inducing risky, high price-pooling is better than both price-separation and low price-pooling.

Step 2: We now show that the low-signal issuer will always mimic the high-signal issuer, and that defending separation is too costly. For the low-signal issuer it is profitable to mimic the high-signal issuer in β_1^p if

$$(1-\beta_1^p)\left((1-q)+\frac{q}{2}\right)\ p_{1,\frac{1}{2},1} > (1-\beta_0^p)\left((1-q)+\frac{q}{2}\right)\ p_{1,\frac{1}{2},0}. \tag{2.43}$$

Numerically the deviation profit is always higher, thus spread-separating in β_1^p, β_0^p cannot be an equilibrium.

Step 3: The high-signal issuer's defenses against mimicking have to be analyzed for any of the three candidate equilibrium spreads. Price-pooling inducing β_1^p, price-separating inducing β_1^s, and risk-less pooling inducing $\beta = 0$ would be defended by setting a higher βs. However, none of these defenses turn out to be feasible.

(a) *Defending Price-Separation.* The lowest spread $\tilde{\beta}$ for which the low-signal issuer will not mimic the price-separation inducing spread any longer, is given by

$$\begin{aligned}(1-\tilde{\beta})\left(\frac{3}{2}q(1-q)p_{1,1,1}+(q^2+(1-q)^2)p_{0,0,1}\right) &= \\ (1-\beta_0^p)\left((1-q)+\frac{q}{2}\right)\ p_{1,\frac{1}{2},0}&.\end{aligned} \tag{2.44}$$

Solving for $\tilde{\beta}$, numerically $\tilde{\beta}$ exceeds by far 10% (and thus lies outside the relevant parameter region). It also exceeds β_1^p, which brings us to the next case.

(b) *Defending Risky Price-Pooling.* If the high-signal issuer sets $\tilde{\beta} > \beta_1^p$ the low-signal issuer will no longer mimic if

$$(1-\tilde{\beta})\left((1-q)+\frac{q}{2}\right)\ p_{1,\frac{1}{2},1} = (1-\beta_0^p)\left((1-q)+\frac{q}{2}\right)\ p_{1,\frac{1}{2},0}. \tag{2.45}$$

Solving for $\tilde{\beta}$, numerically $\tilde{\beta}$ exceeds by far 10% (and thus lies outside the relevant parameter region).

(c) *Defending Riskless Price-Pooling.* If the high-signal issuer sets $\tilde{\beta} \in (0, \beta_1^s)$ the low-signal issuer will no longer mimic if

$$(1-\tilde{\beta})\; p_{0\frac{1}{2},1} = (1-\beta_0^p)\; \left((1-q)+\frac{q}{2}\right)\; p_{1,\frac{1}{2},0}. \tag{2.46}$$

Solving for $\tilde{\beta}$, numerically it exceeds by far 10% (and thus lies outside the relevant parameter region).

Thus, there is no spread-separating equilibrium.

Step 4: Consider now the spread-pooling equilibria. As usual, there are three candidate spreads: $\beta = 0$, $\beta_{\frac{1}{2}}^s$ and $\beta_{\frac{1}{2}}^p$. It turns out that profits under $\beta = 0$ are dominated by profits under the other two spreads. Moreover, $\beta_{\frac{1}{2}}^s$ is preferred by the high-signal issuer, $\beta_{\frac{1}{2}}^p$ from the low-signal issuer.

(a) *Low-Signal Issuer.* Price-separation is better than riskless price-pooling if

$$(1-\beta_{\frac{1}{2}}^s)\left(\frac{3}{2}q(1-q)p_{1,1,\frac{1}{2}} + ((1-q)^2+q^2)p_{0,0,\frac{1}{2}}\right) > p_{0,\frac{1}{2},\frac{1}{2}}. \tag{2.47}$$

Numerically, given $C/N < R_1(q)$, this inequality always holds. Risky price-pooling is better than risk-less price-pooling if

$$(1-\beta_{\frac{1}{2}}^p)\left((1-q)\;\frac{1+q}{2}+q\;\frac{2-q}{2}\right) p_{1,\frac{1}{2},\frac{1}{2}} > p_{0,\frac{1}{2},\frac{1}{2}}. \tag{2.48}$$

Numerically, given $C/N < R_1(q)$, this inequality also always holds. However, the high type prefers risky price-pooling to price-separation as

$$\begin{aligned} (1-\beta_{\frac{1}{2}}^p)\left((1-q)\frac{1+q}{2}+q\frac{2-q}{2}\right) p_{1,\frac{1}{2},\frac{1}{2}} > \\ (1-\beta_{\frac{1}{2}}^s)\left(\frac{3}{2}q(1-q)p_{1,1,\frac{1}{2}} + ((1-q)^2+q^2)p_{0,0,\frac{1}{2}}\right) \end{aligned} \tag{2.49}$$

holds numerically, given $C/N < R_1(q)$.

(b) *High-Signal Issuer.* Price-separation is better than risk-less price-pooling if

$$(1-\beta_{\frac{1}{2}}^s)\left((q^2+(1-q)^2/2)p_{1,1,\frac{1}{2}} + 2q(1-q)p_{0,0,\frac{1}{2}}\right) > p_{0,\frac{1}{2},\frac{1}{2}}. \tag{2.50}$$

Numerically, given $C/N < R_1(q)$, this inequality always holds. Risky price-pooling is better than risk-less price-pooling if

$$(1-\beta^p_{\frac{1}{2}})\left(q\frac{1+q}{2}+(1-q)\frac{2-q}{2}\right)p_{1,\frac{1}{2},\frac{1}{2}} > p_{0,\frac{1}{2},\frac{1}{2}}. \tag{2.51}$$

Numerically, given $C/N < R_1(q)$, this inequality always holds. However, price-separation is also almost always preferred to risky price-pooling as

$$\begin{aligned}(1-\beta^s_{\frac{1}{2}})\left((q^2+(1-q)^2/2)p_{1,1,\frac{1}{2}}+2q(1-q)p_{0,0,\frac{1}{2}}\right) >\\ (1-\beta^p_{\frac{1}{2}})\left(q\frac{1+q}{2}+(1-q)\frac{2-q}{2}\right)p_{1,\frac{1}{2},\frac{1}{2}}\end{aligned} \tag{2.52}$$

holds numerically, given $C/N < R_1(q)$.

Step 5: Thus there are two spread-equilibria that can be constructed to be PBEs. Conveniently, however, spread $\beta^p_{\frac{1}{2}}$ fails the Intuitive Criterion. To see this, define $\tilde{\beta}(q)$ to be the spread for given q that makes the low type not wanting to deviate from $\beta^p_{\frac{1}{2}}$, even if he was perceived to be the highest type. For simplicity, assume that at the deviation payoff spreads are set to be price-separating. Then $\tilde{\beta}(q)$ solves

$$\begin{aligned}(1-\tilde{\beta}(q))\left(\frac{3}{2}q(1-q)p_{1,1,1}+(q^2+(1-q)^2)p_{0,0,1}\right) =\\ (1-\beta^p_{\frac{1}{2}})\left((1-q)\frac{1+q}{2}+q\frac{2-q}{2}\right)p_{1,\frac{1}{2},\frac{1}{2}}.\end{aligned} \tag{2.53}$$

Numerically, for $q > 0.72$, $\tilde{\beta}(q)$ can be set to β^s_1, for smaller q, it has to be larger. However, numerically it also holds that for all q, $\tilde{\beta} < \hat{\beta}^s_1 = C/\mathsf{S}(1-\alpha_{0,1})/(\alpha_{0,1}p_{1,1,1}-p_{0,0,1})$, where $\hat{\beta}^s_1$ is the spread so that the low type bank is indifferent between choosing $p_{1,1,1}$ and $p_{0,0,1}$. (Recall that for higher spread the bank lowers the price to $\phi_{0,1}(p_{0,0,1})$). Consequently at every $\tilde{\beta}(q)$ the bank charges a separation price. Furthermore, numerically at for all q, the high type prefers to deviate to $\tilde{\beta}(q)$ if he is perceived to be the high type, whereas the low type prefers the current equilibrium. Hence there is a deviation that, in the best of all worlds for beliefs, is only profitable for the high type issuer and so the equilibrium $\beta^p_{\frac{1}{2}}$ fails the IC.

Consider now the price-separation-inducing spread and construct the same deviation $\tilde{\beta}(q)$ as above. It turns out, however that for every q and any for every $\tilde{\beta} < 10\%$,

$$(1-\tilde{\beta})\left(p_{1,1,1}\frac{3q(1-q)}{2} + p_{0,0,1}(q^2+(1-q)^2)\right) >$$
$$(1-\beta^s_{\frac{1}{2}})\left(p_{1,1,\frac{1}{2}}\frac{3q(1-q)}{2} + p_{0,0,\frac{1}{2}}((1-q)^2+q^2)\right). \tag{2.54}$$

Any $\tilde{\beta}$ satisfying this equation with equality could be taken as a benchmark for deviation-considerations. However, since there's no feasible $\tilde{\beta}$ that satisfies our restriction and equation (2.54) with equality, the out of equilibrium belief of low type deviation is IC-proof.

As a consequence of all this, the only IC-proof and issuer-efficient equilibrium is pooling in spreads $\beta^s_{\frac{1}{2}}$ which induce price-separation. $\square$

In the proof we used the following threshold values for spreads. They are computed in the same way as demonstrated in Subsection 2.4.1. Note that β^s_ν is *not* the same as the ones used in the next proof.

$$\beta^s_1 = \frac{1-\alpha_{1,1}}{\alpha_{1,1}p_{1,1,1} - p_{0,\frac{1}{2},1}}\frac{C}{\mathsf{S}}, \quad \beta^p_1 = \frac{1-\alpha_{1,0}}{\alpha_{1,0}p_{1,\frac{1}{2},1} - p_{0,0,1}}\frac{C}{\mathsf{S}}, \quad \beta^s_0 = \frac{1-\alpha_{0,1}}{\alpha_{0,1}p_{1,1,0} - p_{0,\frac{1}{2},0}}\frac{C}{\mathsf{S}},$$
$$\beta^p_0 = \frac{1-\alpha_{0,0}}{\alpha_{0,0}p_{1,\frac{1}{2},0} - p_{0,0,0}}\frac{C}{\mathsf{S}}, \quad \beta^s_{\frac{1}{2}} = \beta^s_\diamond, \quad \beta^p_{\frac{1}{2}} = \beta^p_\diamond.$$

Proof of Proposition 2.4

If the issuer signals his information, the bank's price choice carries no extra value. Thus in prices, μ is substituted with a diamond. For the bank's probability of a successful IPO, spreads do not carry information, thus in $\alpha_{j,\nu}$, $j = 0, 1$, spread information ν is substituted with a diamond. In Subsection 2.3.2, Equation (2.9) we have already described the spreads which induce banks to choose risky prices: (1) The high-signal bank chooses risky $p_{1,\diamond,1}$ with B_1 if it is offered at least β^s_1. (2) The low-signal bank chooses risky $p_{1,\diamond,0}$ if it is offered at least β_0^s, where

$$\beta_1^s = \frac{1-\alpha_{1,\diamond}}{\alpha_{1,\diamond}p_{1,\diamond,1} - p_{0,\diamond,1}}\frac{C}{\mathsf{S}}, \text{ and } \beta_0^s = \frac{1-\alpha_{0,\diamond}}{\alpha_{0,\diamond}p_{1,\diamond,0} - p_{0,\diamond,0}}\frac{C}{\mathsf{S}}. \tag{2.55}$$

First, we have to show that both types of issuer actually do want the respective bank to set those risky prices. (1) The high-signal issuer prefers the high-signal bank to set $p_{1,\diamond,1}$ and not $p_{0,\diamond,1}$ if its expected revenue is higher at the risky price, $\alpha_1(1-\beta)p_{1,\diamond,1}\mathsf{S} \geq p_{0,\diamond,1}\mathsf{S}$. (Note that $\beta = 0$ is sufficient for the bank to set the risk-free price.) Solving for β yields that the spread minimal separating has to satisfy

$$\beta_1^s \leq 1 - \frac{p_{0,\diamond,1}}{\alpha_1 p_{1,\diamond,1}} \quad \Leftrightarrow \quad C/N \leq \frac{(2q-1)^2}{2q} =: R_3(q). \tag{2.56}$$

Applying the same reasoning to the low-signal issuer, he prefers the low-signal bank to set risky $p_{1,\diamond,0}$ and not $p_{0,\diamond,0}$ if $\alpha_0(1-\beta)p_{1,\diamond,0}\mathsf{S} \geq p_{0,\diamond,0}\mathsf{S}$. Thus the separating threshold β_0^s has to satisfy[35]

$$\beta_0^s < 1 - \frac{p_{0,\diamond,0}}{\alpha_0 p_{1,\diamond,0}} \quad \Leftrightarrow \quad C/N \leq \frac{2q(2q-1)^2}{(1-q)^2}\, {p_{0,\diamond,0}}^2 =: R_4(q). \tag{2.57}$$

We restrict the analysis to the empirically relevant parameter space where spreads do not exceed 10%. Since we know $\beta_0^s > \beta^p > \beta_1^s$ we impose $\beta_0^s < 10\%$. This translates into any β has to be smaller than $(4q - 1 - 5q^2 + 2q^3)/5(1 - 2q + 2q^2) =: R_5(q)$. Numerically it is easy to check that $R_5 < \min\{R_1, R_2, R_3, R_4\}$, that is, requiring spreads not to exceed 10% is sufficient for all other restrictions to hold.

Second, we have to show that there is no profitable deviation for either issuer. (1) Consider the low-signal issuer. Notice that $\beta_1^s < \beta_0^s$ and $p_{1,\diamond,1} > p_{1,\diamond,0}$, i.e. the high-signal issuer's spread is lower and the offer price is higher so the low-signal issuer had the incentive to deviate if the low-signal bank sets $p_{1,\diamond,1}$ when being offered β_1^s. However, at β_1^s the high-signal bank is just indifferent between risky $p_{1,\diamond,1}$ and risk-free $p_{0,\diamond,1}$. Since the low-signal bank holds less favorable prospects about investors' valuations it will not set $p_{1,\diamond,1}$ and thus investors learn that the issuer's/bank's signal

[35] If any of these restrictions on C/N is satisfied strictly, the necessary spreads can be set lower.

is $s = 0$. But in this case β_0^s is the best choice for the low-signal issuer. (2) Consider now the high-signal issuer. Since $\beta_1^s < \beta_0^s$ and $p_{1,\diamond,1} > p_{1,\diamond,0}$ the high-signal issuer will never mimic the low-signal issuer.

Third, we have to check if pooling in spreads can be an equilibrium. If there is pooling in spreads banks set prices as in Section 2.3. Since we assume $C/N < R_5$ we also have $C/N < R_1$, but then the best to do is pooling in $\beta_{\frac{1}{2}}^p$ and banks setting $p_{1,\frac{1}{2},\frac{1}{2}}$. However, since $\beta_{\frac{1}{2}}^p > \beta_1^s$ and $p_{1,\frac{1}{2},\frac{1}{2}} < p_{1,\diamond,1}$ the high-signal issuer will want to deviate, and he is the only one who can do so profitably under the conditions set by the IC, so this cannot be an IC-proof equilibrium. □

Proof of Proposition 2.5

We will show that the bank earns non-negative profits at all three possible spread levels.

(a) With informed issuers, if $s_b = 0$, the spread is β_0^s and the low-signal bank sets $p_{1,\diamond,0}$ and incurs the risk of losing C. Instead of taking this risk, the bank may choose a risk-free price $p_{0,\diamond,0} > 0$. Being compensated for the risk means that the low-signal banks gets more than $\beta_0^s \mathsf{S} p_{0,\diamond,0} > 0$.

(b) Likewise, if $s_b = 1$, the issuer sets β_1^s and the bank sets $p_{1,\diamond,1}$. Instead, the bank could set price $p_{0,\diamond,1} > 0$ and realize risk-free profits. To make the bank set the risky price, the issuer has to pick a compensation which gives the bank at least $\beta_0^s \mathsf{S} p_{0,\diamond,1} > 0$.

(c) With uninformed issuers the spread is $\beta_\diamond^p$ and the high pooling price $p_{1,\frac{1}{2},\diamond}$ results. In this case expected profits are positive as long as we have

$$\frac{\alpha_0 + \alpha_1}{2}\, \beta^p p_{1,\frac{1}{2}} \mathsf{S} > (1 - \frac{\alpha_0 + \alpha_1}{2})\, C \quad \Leftrightarrow \quad \beta^p > \frac{2 - \alpha_0 - \alpha_1}{(\alpha_0 + \alpha_1) p_{1,\frac{1}{2}}}\, \frac{C}{\mathsf{S}}. \tag{2.58}$$

Numerical simulations show that holds true for all $q \in (.6, 1)$. □

Proof of Proposition 2.7

From Propositions 2.2 and 2.4 we know that an uninformed issuer always sets $\beta_\diamond^p$; an identically informed issuer with signal $s_b = 0$ sets β_0^s, if he has signal $s_b = 1$ he sets β_1^s.

Ex ante, the identically informed issuer gets either signal with equal probability. Thus for the claim to be true it must hold that

$$\frac{1}{2}\,\beta_1^s + \frac{1}{2}\,\beta_0^s > \beta_\diamond^p \quad\Leftrightarrow\quad \frac{1}{2}\,\frac{1-\alpha_1}{\alpha_1 p_{1,1} - p_{0,1}} + \frac{1}{2}\,\frac{1-\alpha_0}{\alpha_0 p_{1,0} - p_{0,0}} > \frac{1-\alpha_0}{\alpha_0 p_{1,\frac{1}{2}} - p_{0,0}}. \qquad (2.59)$$

Checking this numerically, the inequality holds if $C/N < R_1(q)$. By Proposition 2.6, the VC-backed issuer sets even lower spreads.

Instead of spread deviations suppose that an issuer abandons his commercial bank and seeks investment banking services from a third party, and let this be common knowledge. It is numerically straightforward to show that the high-signal issuer would not be interested in this move: The best to happen is that he would be perceived as a high-signal issuer. But then the highest expected payoff he could get from an independent bank is lower than what he gets from his commercial bank. The reason is that with a third party, there is a risk that the bank gets an unfavorable signal and charges the low price. Thus if the high-signal bank would not change, any change of banking-partner would be perceived as coming from a low-signal bank. It is straightforward to check numerically that the low-signal bank then would not want to deviate either. □

List of Restrictions:

$$\begin{aligned}
R_1 &= 2q(2q-1)^2(q-1)^2(1-q^2+q)/(4q-9q^2+19q^3-25q^4+17q^5-2q^6-1)\\
R_2 &= 2(q-1)(1-q-3q^2+2q^3)(2q-1)/3q(1-2q+2q^2)\\
R_3 &= (2q-1)^2/2q,\\
R_4 &= 2q(2q-1)^2(q-1)^2/(1-2q+2q^2)^2\\
R_5 &= (4q-1-5q^2+2q^3)/5(1-2q+2q^2)
\end{aligned}$$

Chapter 3

Working for Today or for Tomorrow: Incentives for Present-Biased Agents

3.1 Introduction

We examine self-control problems – modeled as time-inconsistent, present-biased preferences – in a multi-tasking environment. An agent must allocate effort between an incentivized and immediately rewarded activity and a private activity that pays out with some delay. Effort costs accrue immediately. As an example think of a situation in with an agent has to decide how much to work on her job and how much to care about her health. Effort on the job is assumed to pay out immediately (e.g. wage payment at the end of the month, piecework rate, job promotion, etc.) while time dedicated to care for one's health (free weekends, workouts, balanced diet instead of fast-food, etc.) pays out in the long-run only. Alternatively, one can think of a student's decision how much to study for a degree and how much to work for money while being a student.

The existing literature on present-biased preferences (O'Donoghue and Rabin (1999a, 1999b, 2001)) analyzes environments in which an agent must accomplish a single task only but has discretion *when* to do it. In contrast, we model situations in which agents must allocate effort *between* tasks and has to decide *how much* of each activity to accomplish – with the complication that some tasks pay out earlier than others.

We consider three types of agents. Time-consistent agents, sophisticated agents, and naive agents. Throughout the paper we abstract from complications of adverse selection. The principal is thus assumed to know the type of agent he is contracting with. To capture the basic effects we start out with the most simple set-up. The agent has to allocate effort between two tasks only. Effort allocated to the first task is incentivized by a principal by way of a linear contract and rewarded immediately. Effort devoted to the second task serves the agent's private benefit but pays out with some delay. Effort is taken to be perfectly observable and contractable. There is no risk involved. We consider a three-period setting. In a first period the principal offers a contract and the agent decides whether to accept or to refrain. In the next period the agent chooses effort levels. According to the incentive scheme, effort devoted to the principal's purposes pays out immediately, and all effort costs are borne in immediately. In the final period the agent's private benefit is realized.

We show that present-biased agents take decisions that do not maximize their long-run interest, irrespective of the intensity of incentives. Sophisticated agents are, however, never harmed by incentives relative to a situation without incentives as they always receive their reservation utility levels. With naive agents there are two effects. On the one hand, they wrongly belief to have a high reservation utility because they think they will not give in to a present bias in a situation without incentives. They thus only participate if they are paid high enough wages. On the other hand, they wrongly predict tomorrow's effort choices (naive agents think to act like time-consistent agents but in fact they will always give in to their present bias) which is exploited by the principal. We show that the second effect always dominates. Naive agents are thus harmed by incentives relative to the situation without incentives. Furthermore, we show that social welfare can *decrease* in the presence of incentives. With naive agents it may happen that their additional loss due to present-biased effort choices in the situation with incentives exceeds the principal's gain from offering the incentive contract. The model thus offers a new theoretical possibility of detrimental effects of incentives, complementary to existing arguments like the crowding out of intrinsic motivation by extrinsic rewards.

While the current paper is (to the best of our knowledge) the first analysis of present-biased preferences in a multi-tasking environment, there is an extensive literature that analyzes choice problems with time-inconsistent preferences in situations where agents have discretion *when* do complete a task. O'Donoghue and Rabin (1999a) derive a present-biased agent's decision with exogenously given levels for costs and rewards – one of those being immediate, the other delayed. O'Donoghue and Rabin (1999b) analyze a principal-agent setting. A procrastinating agent faces stochastic costs of completing a task. The principal offers an incentive contract to induce the agent to complete that task in time. If the principal knows the agent's cost distribution he can always achieve the first-best. With asymmetric information about costs the first-best is achieved with time-consistent agents only. With time-inconsistent agents incentives for timely completion and efficient delay in case of high costs must be traded-off. The second-best contract involves an increasing punishment for delay. O'Donoghue and Rabin (2001) show that providing a present-biased agent with additional choice options can harm those agents. Agents may refrain from completing an activity because they change to a better but never-to-completed alternative. DellaVigna and Malmendier (2003) analyze the health club industry and provide evidence for both, time-inconsistent behavior and naiveté. DellaVigna and Malmendier (2004) go a step further and derive the optimal contract design of firms if consumers have time-inconsistent preferences. They show that optimal contracts for naive agents have observed features in some industries like, among others, the health club and credit card industry – suggesting that people have self-control problems and that they are not fully aware of it.

The remainder of the paper is organized as follows. Section 3.2 presents the multi-tasking model with time-inconsistent agents. Section 3.3 presents our main results. Section 3.4 concludes.

3.2 The Model

In this section we lay out our model of time-inconsistent preferences in a multi-tasking environment. First, present-biased preferences and possible believes about future behavior are introduced. Second, a multi-tasking principal-agent environment is set up. In contrast to the seminal contribution by Holmström and Milgrom (1991), we add the complication that effort devoted to some tasks pays out immediately while effort devoted to other tasks pays out with delay. We assume that effort costs accrue immediately. Finally, we combine both approaches in a simple, three-period model with two tasks only.

3.2.1 Present-Biased Preferences and Beliefs

Let u_t be an agent's instantaneous utility in period t. In each period an agent does not only care about her instantaneous utility, but also about her discounted future instantaneous utilities. Let $U^t(u_t, ..., u_T)$ denote an agent's intertemporal preference from the point of view of period t. The standard model employed by economists is exponential discounting, that is $U^t(u_t, ..., u_T) = \sum_{\tau=t}^{T} \delta^\tau u_\tau$, where $\delta \in (0, 1]$ denotes the discount factor. In a parsimonious way, exponential discounting captures the fact that agents are impatient. In addition, it implies that agent's decisions are time-consistent. When considering trade-offs between two periods in time it does not matter when the agent is asked to take a decision. However, people tend to exhibit time-inconsistent preferences (Benzion, Rapoport, and Yagil (1989), Kirby (1997), Kirby and Herrnstein (1995)). By way of example: When being asked, most people will prefer to receive \$100 in 6 weeks over \$90 in 5 weeks from now. However, when being asked again for their preference 5 weeks from now, some people will reverse their decisions to wait for the higher payment and opt for the immediate payment of \$90. Such present-biased preferences have been modelled by Phelps and Pollack (1968) and later, among others, by Laibson (1997) and O'Donoghue and Rabin (1999a, 1999b, 2001).[36] We follow this

[36]See Frederick, Loewenstein, and O'Donoghue (2002) for a comprehensive overview of the literature.

literature and apply a two-parameter model that can capture present-biased preferences by a simple modification of exponential discounting.

Definition 3.1 (β, δ)*-preferences* are time-inconsistent preferences that are represented as follows: For all t, $U^t(u_t, u_{t+1}, ..., u_T) = \delta^t u_t + \beta \sum_{\tau=t+1}^{T} \delta^\tau u_\tau$, where $0 < \beta$, $\delta \leq 1$.

In this formulation, δ represents the long-run, time-consistent discount parameter. The parameter β represents the the bias for the present. If $\beta = 1$, (β, δ)-preferences coincide with standard exponential discounting. But if $\beta < 1$, an agent places more relative weight to period τ in period τ than she did in any other period prior to τ. Applied to the above example: Be $\delta = .95$ the weekly discount factor and $\beta = .9$ the present-bias. The agent then prefers \$100 in week 6 over \$90 in week 5 in every week before week 5 as $\$90 < .95 \cdot \$100$. But in week 5 her preference reverses as $\$90 > .9 \cdot .95 \cdot \$100$.

Most researchers have modeled time-inconsistent preferences by interpreting an agent at each point in time as a separate agent.[37] An agent thus consists of 'multiple selves', where each 'self' is choosing current behavior to maximize current preferences. The 'current self' knows that her 'future selves' control future behaviors and thus holds believes about her future selves. Strotz (1956) and Pollack (1968) applied two extreme assumptions and established the following labels:

Definition 3.2 (i) A *sophisticated agent* is fully aware of her future selves. Such an agent takes into account that future selves may exhibit time-inconsistent preferences. (ii) A *naive agent* thinks that future selves will take time-consistent decisions. Such an agent does not take into account that future selves in fact take present-biased decisions.

There is a long standing and lasting debate over whether people are naive or sophisticated. On the one hand, O'Donoghue and Rabin (1999a) report on self-commitment devices such as alcohol clinics, Christmas clubs, or fat farms, indicating that people are (at least partially) aware of their time-inconsistent behaviors. On the other hand,

[37]For an alternative approach see Gul and Pesendorfer (2001).

DellaVigna and Malmendier (2003, 2004) report on evidence from, among others, the health club and credit card industry that suggests that people are not (fully) aware of their future self-control problems. Apart from the two extreme assumptions, in principal, any degree of sophistication could be modeled. However, in this paper we follow the approach in O'Donoghue and Rabin (1999a) and analyze fully sophisticated and fully naive agents only.

3.2.2 Multi-Tasking with Immediate and Delayed Benefits

We analyze a situation in which an agent must allocate effort between different tasks. We further assume that all effort costs are immediate while some but not all tasks pay out with a delay only. Effort on the job is assumed to pay out immediately (wage payment) while time dedicated to care for, say, one's health pays out in the long-run only (absence of health problems). As another example, consider a student's decision to work a couple of hours per week for a firm or to fully concentrate on one's studies. While the wages from working for the firm are paid out immediately, higher wages that come along with good grades are realized only in the future.

In a classic paper Holmström and Milgrom (1991) derive optimal linear incentive contracts in a principal-agent setting with non-verifiable effort such that wages must condition on noisy signals.[38] Without loss of generality we employ the linear incentive model in this paper as well. However, the focus of our model very different. Holmström and Milgrom are interested in the implications on optimal incentive provision if performance measure are of diverging quality. They show that it can be optimal to refrain from providing explicit incentives if, for example, only one of two tasks can be measured, but some engagement in both tasks is desirable. They further analyze asset ownership and job design. In the current model we abstract from problems of measurement and risk allocation. We are interested in the implications for incentive provision if the incentivized task pays out immediate while the private activity pays out only

[38] The underlying assumption in their model is that the agent chooses effort levels continuously over the time interval $[0, 1]$ to control the drift vector of a Brownian motion. At each point in time the agent can observe his accumulated performance before acting. Holmström and Milgrom (1987) show that in such a setting the optimal incentive contract is indeed linear.

with delay. While existing models of time-inconsistent behavior analyze environments in which an agent must accomplish a single task but has discretion *when* to do it, our focus is on effort allocation *between* tasks.

3.2.3 Combining Present-Biased Preferences and Multi-Tasking

We now combine present-biased preferences with multi-tasking. To capture the basic effects we start out with the most simple set-up. The agent has to allocate effort between two tasks only. Effort allocated to the first task, e_1, is incentivized by a principal and rewarded by way of a linear incentive scheme (e.g. effort in the work place). The principal's benefit from e_1 is captured by $B(e_1)$, with $B'(e_1) > 0$ and $B''(e_1) < 0$. Function $B(e_1)$ is assumed to measure the principal's benefit in monetary terms. The reward is based on a signal μ that is produced by e_1. In this basic setting we abstract from issues of risk-allocation and assume that the signal is deterministic, i.e. $\mu(e_1) = e_1$. The incentive scheme can thus be written as $w(e_1) = \alpha e_1 + \gamma$. The principal's profit is then given by $\Pi = B(e_1) - \alpha e_1 - \gamma$. Effort devoted to the second task, e_2, serves the agent's private benefit $V(e_2)$, with $V'(e_2) > 0$ and $V''(e_2) < 0$ (e.g. caring for one's health). We assume that $V(e_2)$ accrues with one period delay only. Effort costs, $C(e_1, e_2)$, are however immediate. We assume $C_i(e_1, e_2) > 0$, $C_{ii}(e_1, e_2) > 0$, and $C_{ij}(e_1, e_2) > 0$, with $i, j \in \{1, 2\}$, subscripts denoting partial derivatives. We thus assume that effort levels are substitutes at the margin. This will be the case if, for example, effort is interpreted as measuring the time devoted to a certain activity. Both, $V(e_2)$ and $C(e_1, e_2)$ are assumed to represent the agent's benefit and cost in monetary terms.

We consider a setting with three periods only. In period $t = 0$ the principal offers a contract, i.e. values for α and γ, and the agent decides whether to accept or to refrain. In the next period, $t = 1$, the agent chooses effort levels. According to the incentive scheme, the effort devoted to the principal's purposes, e_1, pays out immediately. The agent also decides on effort devoted to the own private benefit. All effort costs are borne in immediately in $t = 1$. In the final period, $t = 2$, the agent's private benefit is realized. The timing of the game is summarized in Figure 3.1.

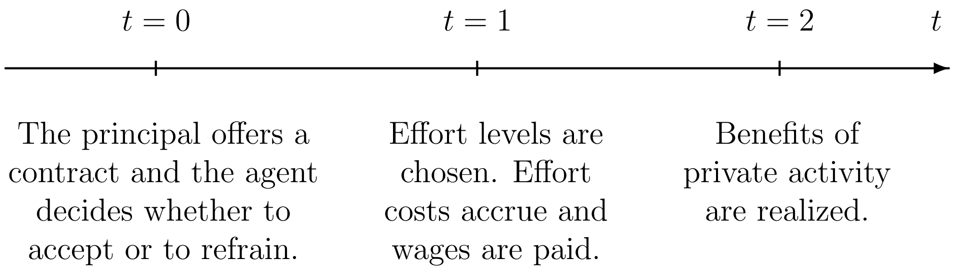

Figure 3.1: **The Timing of the Game.**

3.2.4 Benchmark: Incentives for Time-Consistent Agents

As a benchmark we first derive the optimal contract for time-consistent agents (TCs). Throughout the paper we will assume explicit functional forms in order to obtain closed form solutions. In Section 3.4 we discuss the generality of our results.

Assumption 3.1 $B(e_1) = b\ln(e_1)$, $V(e_2) = v\ln(e_2)$, and $C(e_1, e_2) = .05(e_1 + e_2)^2$.

In addition, but without loss of generality we normalize δ to unity. We first look at effort levels without incentives. In this case no action is taken in period $t = 0$. Without incentives effort devoted to the principals benefit, e_1, will always be set to zero. In period $t = 1$ a TC thus chooses e_2 to maximize her intertemporal utility which is given by

$$\max_{e_2} V(e_2) - C(e_2) = v\ln(e_2) - .05(e_2)^2. \tag{3.1}$$

Solving the first-order conditions yields $e_2^{\tilde{TC}} = \sqrt{10v}$, where superscript TC stands for 'time consistency' and the tilde indicates the situation without incentives. A TC then realizes an intertemporal (reservation) utility level of $\underline{U}^{TC} = V(e_2^{\tilde{TC}}) - C(e_2^{\tilde{TC}}) = .5v(\ln(10v) - 1)$. This is also a TC's welfare, defining an agent's welfare as follows.

Definition 3.3 The *welfare* of an agent is the sum of her instantaneous utility levels from a long-run perspective, i.e. $U^t(u_t, u_{t+1}, ..., u_T) = \sum_{\tau=t}^{T} \delta^\tau u_\tau$.

With TCs there is no difference between long-run and short-run perspective. The difference will however be important with time-inconsistent agents.

Consider now a principal who offers a linear incentive contract $w(e_1) = \alpha e_1 + \gamma$ in period $t = 0$. To determine the parameters of the contract the principal maximizes

$$\max_{\alpha,\gamma} B(e_1(\alpha)) - \alpha e_1(\alpha) - \gamma \tag{3.2}$$

subject to the participation constraint (PC)

$$\alpha e_1^{TC}(\alpha) + \gamma + V(e_2^{TC}(\alpha)) - C(e_1^{TC}(\alpha), e_2^{TC}(\alpha)) \geq \underline{U}^{TC}, \tag{3.3}$$

where $e_i^{TC}(\alpha)$ with $i \in \{1, 2\}$ are the effort levels that a TC chooses given incentive intensity α. In optimum (3.3) must hold with equality. Substitution of the PC yields

$$\max_{\alpha} B(e_1^{TC}(\alpha)) + V(e_2^{TC}(\alpha)) - C(e_1^{TC}(\alpha), e_2^{TC}(\alpha)) - \underline{U}^{TC}. \tag{3.4}$$

The principal thus maximizes the intertemporal social welfare by choice of incentive intensity α, where social welfare is defined as follows.

Definition 3.4 *Social welfare* is the sum the agent's welfare and the principal's profit.

Recall that we assumed $B(e_1)$, $V(e_2)$, and $C(e_1, e_2)$ to be measured in monetary terms. As the incentivized activity can be measured without error, the linear incentive scheme allows the principal to implement any level of e_1 at first-best costs.

If the TC accepts the contract she will choose effort levels in period $t = 1$ by maximizing

$$\max_{e_1,e_2} \alpha e_1 + \gamma + v \ln(e_2) - .05(e_1 + e_2)^2. \tag{3.5}$$

Solving the first-order conditions we get

$$e_1^{TC}(\alpha) = 10\alpha - \frac{v}{\alpha} \qquad \text{and} \qquad e_2^{TC}(\alpha) = \frac{v}{\alpha}. \tag{3.6}$$

Maximization of (3.4) then yields

$$\alpha^{TC} = \sqrt{(b+v)/10}, \tag{3.7}$$

such that effort levels are given by

$$e_1^{TC}(\alpha^{TC}) = \frac{\sqrt{10}\ b}{\sqrt{b+v}} \qquad \text{and} \qquad e_2^{TC}(\alpha^{TC}) = \frac{\sqrt{10}\ v}{\sqrt{b+v}}, \tag{3.8}$$

the first-best levels of e_1 and e_2 that maximize social welfare.

3.2.5 Incentives for Sophisticated Agents

The optimal incentive contract for sophisticated agents (sophisticates) is derived analogously. Absent incentives a sophisticate chooses e_2 in $t = 1$ to maximize

$$\max_{e_2} \beta V(e_2) - C(e_2) = \beta v \ln(e_2) - .05(e_2)^2, \tag{3.9}$$

which differs from (3.1) in the time-inconsistency parameter β only. Her optimal choice of e_2 is given by $e_2^{\tilde{P}B} = \sqrt{10\beta v}$, where superscript PB indicates 'present bias'. Absent incentives a sophisticate thus realizes a *welfare* level $\underline{U}^{PB} = V(t_2^{\tilde{P}B}) - C(t_2^{\tilde{P}B}) = .5v(\ln(10\beta v) - \beta)$. Notice that from a long-run perspective the discount factor between periods one and two is given by 1 and not by β. The difference of $\underline{U}^{TC}$ and $\underline{U}^{PB}$ is thus an agent's welfare loss in monetary terms due to time-inconsistent preferences. Subtracting $\underline{U}^{PB}$ from $\underline{U}^{TC}$ we get $.5v(\beta - 1 - \ln(\beta))$, which is positive whenever $\beta < 1$, and increasing with β decreasing.

The principal's objective function (3.2) is now subject to a sophisticate's PC

$$\alpha e_1^{PB}(\alpha) + \gamma + V(e_2^{PB}(\alpha)) - C(e_1^{PB}(\alpha), e_2^{PB}(\alpha)) \geq \underline{U}^{PB}. \tag{3.10}$$

Notice that the agent's benefit from e_2, V, is not discounted by β. A sophisticate decides in period $t = 0$ whether or not to accept the incentive contract. Both sides of (3.10) are thus discounted by β which therefore cancels. In optimum the PC must hold with equality. Substituting (3.10) into (3.2) yields

$$\max_{\alpha} B(e_1^{PB}(\alpha)) + V(e_2^{PB}(\alpha)) - C(e_1^{PB}(\alpha), e_2^{PB}(\alpha)) - \underline{U}^{PB}. \tag{3.11}$$

When choosing incentive intensity α the principal now has to take into account that the agent will give in to her present-biased preference when she decides on effort levels in period $t = 1$. If the contract is accepted a sophisticate maximizes

$$\max_{e_1, e_2} \alpha e_1 + \gamma + \beta v \ln(e_2) - .05(e_1 + e_2)^2. \tag{3.12}$$

Solving the first-order conditions gives effort choices

$$e_1^{PB}(\alpha) = 10\alpha - \frac{\beta v}{\alpha} \qquad \text{and} \qquad e_2^{PB}(\alpha) = \frac{\beta v}{\alpha}, \tag{3.13}$$

which differ from (3.6) only in the time-inconsistency parameter β. Maximizing now the principal's objective function (3.11) with respect to α yields

$$\alpha^S = \sqrt{(b + \beta v)/10}, \tag{3.14}$$

such that effort levels are given by

$$e_1^{TC}(\alpha^{TC}) = \frac{\sqrt{10}\, b}{\sqrt{b + \beta v}} \qquad \text{and} \qquad e_2^{TC}(\alpha^{TC}) = \frac{\sqrt{10}\, \beta v}{\sqrt{b + \beta v}}, \tag{3.15}$$

Again, for $\beta = 1$ both effort and incentive levels of TCs and sophisticates coincide. By comparison of (3.7) and (3.14) it can be seen that $\alpha^{TC} > \alpha^S$ whenever $\beta < 1$. Comparing (3.8) and (3.15) shows that $\beta < 1$ ensures $e_1^S > e_1^{TC}$ but $e_2^{TC} > e_2^S$. That is, given Assumption 3.1, TCs receive stronger incentives to work for the principal but work less hard than sophisticates. TCs choose a higher effort level in their private activities. This higher effort level increases effort costs such that it becomes relatively more expensive for the principal to compensate for the incentivized activity. Figures 3.2, 3.3, and 3.4 provide an illustration.

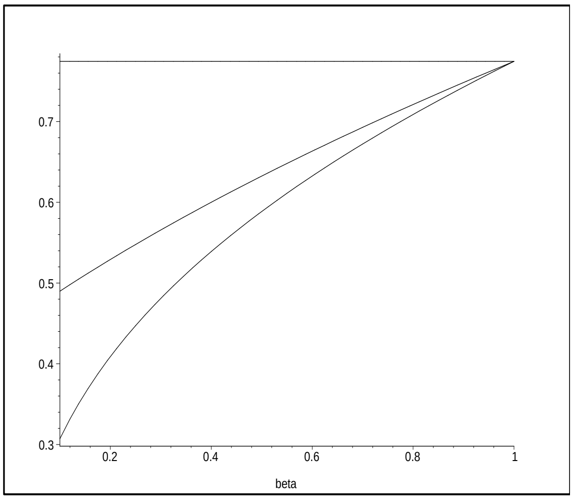

Figure 3.2: **Alpha levels.** The figure shows alpha levels as functions of β, given $b = 2$ and $v = 4$. The top straight line depicts incentive intensity with TCs. The curve in the middle is alpha for sophisticates, the lowest curve for naifs. For $\beta = 1$ incentive intensities coincide.

Sophisticates harm themselves due to time-inconsistent preferences. Their welfare is maximized at effort levels (3.8) which differ from (3.15) whenever $\beta < 1$. However, there is no additional loss caused by the introduction of explicit incentives for an immediately rewarded task. The participation constraint ensures that a sophisticate always receives her reservation welfare level of $\underline{U}^S$.

3.2.6 Incentives for Naive Agents

We now turn to naive agents (naifs). Absent incentives there is no difference between sophisticates and naifs. No action is taken in period $t = 0$, and it is only then that beliefs about preferences in period $t = 1$ can differ. In the case with incentives it appears - at first sight - unclear whether naifs or sophisticates will be better off. There are two opposing effects. On the one hand, in period $t = 0$ naifs wrongly believe that they will take a time-consistent, welfare maximizing decision in period $t = 1$. Hence, to ensure a naif's participation she must be given a *perceived* reservation utility of $\underline{U}^{TC}$ which exceeds her true reservation utility of $\underline{U}^{PB}$. On the other hand, a naif wrongly predicts her effort choices in $t = 1$ given incentive intensity α. She expects to act like a TC according to (3.6), but will indeed give in to her time-inconsistent preferences and

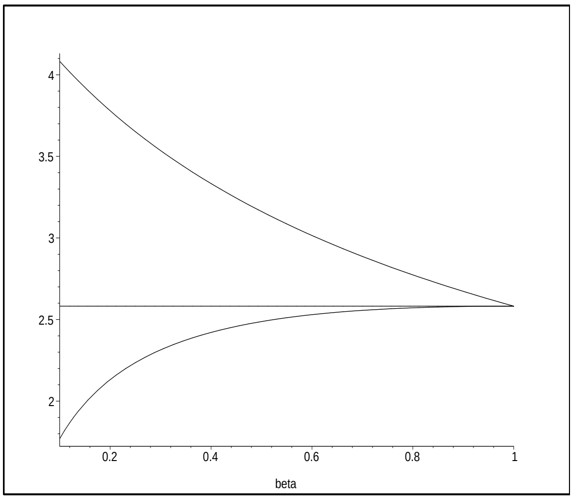

Figure 3.3: **Levels of e_1.** The figure depicts levels of e_1 as functions of β, given $b = 2$ and $v = 4$. The straight line in the middle is e_1^{TC}, the highest, decreasing line is e_1^S, and the lowest, increasing line depicts e_1^N. For $\beta = 1$ levels of e_1 coincide.

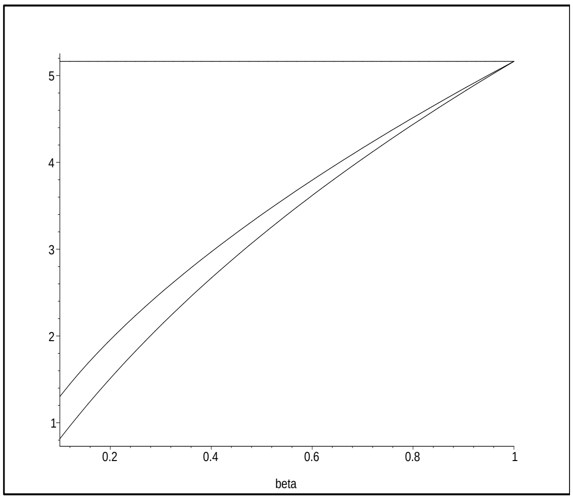

Figure 3.4: **Levels of e_2.** The figure depicts levels of e_2 as functions of β, given $b = 2$ and $v = 4$. The top, straight line is e_2^{TC}, the line in the middle is e_2^N, and the lowest, steepest line depicts e_2^S. For $\beta = 1$ levels of e_2 coincide.

act like a present-biased agent according to (3.13). Her *perceived* utility must not fall short of $\underline{U}^{TC}$, but as she will make different effort choices her *realized* utility level will fall short of $\underline{U}^{TC}$. In Section 3.3 we show that, given Assumption 3.1, it will always even fall short of $\underline{U}^{PB}$. That is, the effect due to wrong beliefs about future selves dominates and naifs are thus worse off than sophisticates. Since a sophisticate receives her reservation utility level which is identical to a naif's welfare without incentives, a naif is harmed by the principal's incentive contract.

In the following we derive the optimal incentive contract from a principal's point of view. The principal's objective function (3.2) is now subject to a naif's PC which is identical to a TC's PC, which is given in equation (3.3). Substituting (3.3) into (3.2) now yields

$$\max_{\alpha} B(e_1^{PB}(\alpha)) - \alpha e_1^{PB}(\alpha) - \underline{U}^{TC} + \alpha e_1^{TC}(\alpha) + V(e_2^{TC}(\alpha)) - C(e_1^{TC}(\alpha), e_2^{TC}(\alpha)) \quad (3.16)$$

Recall that the principal is assumed to know both an agent's time preference and her belief about future selves. The principal thus takes into account that the naif will give in to her time-inconsistent preferences in period $t = 1$ and choose effort levels according to equation (3.13). The first-order condition is now given by

$$(B'(e_1^{PB}) - \alpha)\frac{\partial e_1^{PB}}{\partial \alpha} - e_1^{PB} + e_1^{TC} + \underbrace{(\alpha - C_1)}_{=0}\frac{\partial e_1^{TC}}{\partial \alpha} + \underbrace{(V'(e_2^{TC}) - C_2)}_{=0}\frac{\partial e_2^{TC}}{\partial \alpha} = 0 \quad (3.17)$$

$$\Leftrightarrow \left(\frac{b}{10\alpha - \frac{\beta v}{\alpha}} - \alpha\right)\left(10 + \frac{\beta v}{\alpha^2}\right) - \left(10\alpha - \frac{\beta v}{\alpha}\right) + \left(10\alpha - \frac{v}{\alpha}\right) = 0. \quad (3.18)$$

Solving for α we get

$$\alpha^S = \frac{\sqrt{5}}{10}\sqrt{k + b - v(1-\beta)}, \quad (3.19)$$

with $k = \sqrt{\beta^2 v^2 + (2v^2 + 6bv)\beta + (v-b)^2}$. It is straightforward to show that α^N coincides with α^{TC} for $\beta = 1$. Given the optimized level of alpha naifs choose effort levels

$$e_1^S(\alpha^S) = \frac{\sqrt{5}\ (k + b - v(1+\beta))}{\sqrt{k + b - v(1-\beta)}} \qquad \text{and} \qquad e_2^S(\alpha^S) = \frac{\sqrt{5}\ (2\beta v)}{\sqrt{k + b - v(1-\beta)}}. \quad (3.20)$$

Figures 3.2, 3.3, and 3.4 illustrate the relative size of incentive and resulting effort levels for the three different types of agents.

3.3 Results

Since general functions only implicitly define the solutions to the maximization problems in Sections 3.2.4 to 3.2.6 we have assumed explicit functional forms. But even these simple functional forms do not always allow for closed form solutions. If necessary we will therefore stick to numerical examples to show the results of the paper. For completeness we first establish the following proposition.

Proposition 3.1 *A time-consistent agent's welfare is always higher than a sophisticated or naive agent's welfare.*

Proof: An agent's welfare is defined as the sum of her instantaneous utility levels setting $\beta = 1$. Without incentives only TCs maximize welfare. Both, a sophisticate's and a naif's objective functions at time of effort choice differ from their welfare maximizing objective functions. Their choices must thus be suboptimal. The optimal incentive contracts offer both TCs and sophisticates their reservation utility levels. By the first part of this proof the first exceeds the latter. A naif requests the reservation utility level of a TC. She is thus offered a value for α such that she received $\underline{U}^{TC}$ if she indeed chose like a TC. But in period $t = 1$ she gives in to her present-bias and deviates from her planned effort choice. Her realized utility level must thus lie below $\underline{U}^{TC}$. q.e.d.

The focus of the paper is the comparison of sophisticates and naifs. Naifs wrongly predict their future behaviors: In period $t = 0$ naifs think that they will behave like TCs in period $t = 1$. To accept the principal's incentive contract in $t = 0$ they must be given an perceived utility level that matches the reservation utility level of a TC. This 'commitment effect' works in favor of a naif. However, once in period $t = 1$ a naif deviates from her perceived effort choices and gives in to her present-bias. A naif is thus harmed by this second effect. In contrast, a sophisticate anticipates that she will act according to her present-biased preferences in $t = 1$ and thus requires an perceived utility level that matches the reservation utility of an time-inconsistent agent only. However, even though naifs receive a higher *perceived* utility level, they realize a lower *actual* utility level. We show that, given Assumption 3.1, this second effect always dominates. Naifs thus realize a lower welfare than sophisticates, and this welfare is even lower than the welfare level naifs realize without incentives. This is summarized in the following proposition.

Proposition 3.2 *Given Assumption 3.1, in the presence of incentives a naive agent's welfare can be lower than a sophisticated agent's welfare. A naive agent's welfare can thus be reduced by accepting the incentive contract.*

Proof: The existence of a non-empty parameter space for which the result holds true is shown by numerical example. Figure 3.5 plots welfare levels as functions of β, given $b = 2$ and $v = 4$. The top, straight line depicts a TC's welfare. The curve below

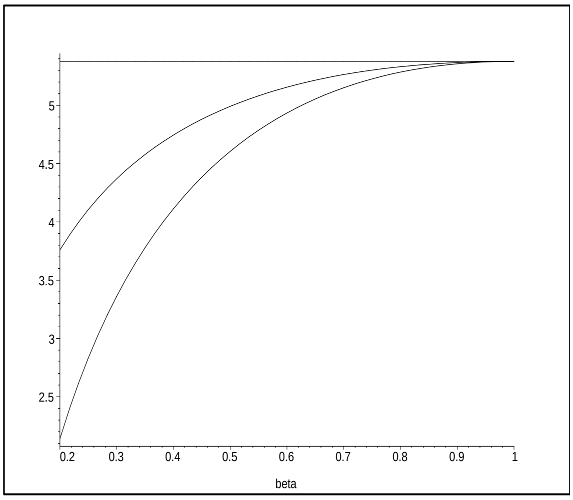

Figure 3.5: **Agents' Welfare Levels with Incentives.** The figure depicts welfare levels as a functions of β, given $v = 4$. The top, straight line depicts a TC's welfare. The curve below depicts a sophisticate's welfare, the lowest curve depicts a naif's welfare. For $\beta = 1$ welfare levels coincide.

depicts a sophisticate's welfare, the lowest curve depicts a naif's welfare. For $\beta = 1$ welfare levels coincide. q.e.d.

From Proposition 3.1 we already know that sophisticates and naifs are always worse off than TCs. We now show that, given Assumption 3.1, even though social welfare realized with sophisticates and naifs is always below social welfare realized with TCs, the principal's profit can be higher if the agent is naive; it can be lower with sophisticates.

Proposition 3.3 *Given Assumption 3.1, the principal's profit from contracting with naive agents can be higher than profit from contracting with TCs, even though social welfare is always lower if agents have present-biased preferences. With sophisticates the principal's profit can be lower than with TCs.*

Proof: The existence of a non-empty parameter space for which the result holds true is shown by numerical example. Figure 3.6 plots levels of social welfare against β, given $b = 6$ and $v = 4$. The top straight line is social welfare with TCs. The curve below depicts social welfare with naifs, and the lowest curve with sophisticates. Figure 3.7 depicts profit levels. The straight line in the middle are profits in case of a TC. The top, decreasing line depicts the principal's profits with naifs. Profits with sophisticates

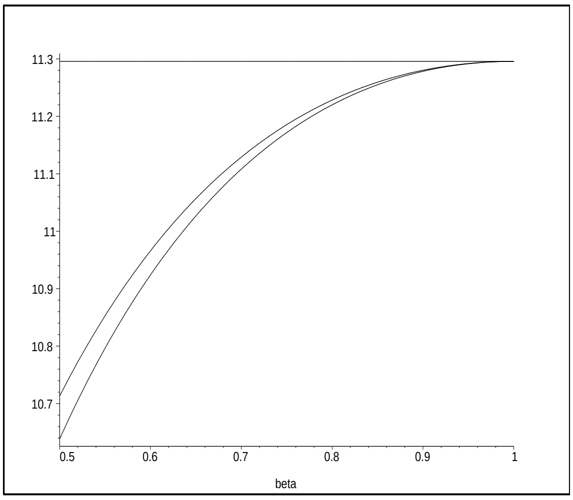

Figure 3.6: **Social Welfare Levels.** The figure depicts social welfare as function of β for $b = 6$ and $v = 4$. The top, straight line depicts social welfare with TCs, the curve in the middle with naifs, the lowest with sophisticates. For $\beta = 1$ values coincide.

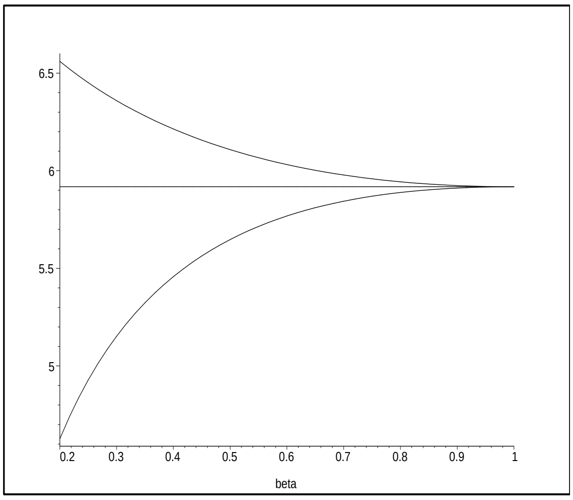

Figure 3.7: **Profit Levels.** The figure depicts profit levels as function of β, given $b = 6$ and $v = 4$. The straight line depicts profits with TCs, the top, decreasing curve with naifs, the lowest curve with sophisticates. For $\beta = 1$ values coincide.

are the lowest, increasing line. q.e.d.

In the following we are interested in the change of social welfare when the principal offers incentive contracts as compared to the situation without incentives. From Proposition 3.2 we know that naifs are harmed by incentives. The principal, on the contrary, gains when offering incentive contracts. From Proposition 3.3 we know that his profit when contracting with naifs can increase, the more severe the time-inconsistency problem gets, i.e. the lower β. Furthermore, in the following we show that the agent's loss can exceed the principal's gain. That is, social welfare may *decrease* if the principal provides incentives relative to the situation without incentives. With TCs or sophisticates this can never happen. Those agents always receive their reservation welfare levels and the principal extracts the complete surplus from the additional, efficient activity. This finding is summarized in the following proposition.

Proposition 3.4 *Given Assumption 3.1, with naive agents social welfare may decrease if the principal offers incentive contracts as compared to the situation without incentives.*

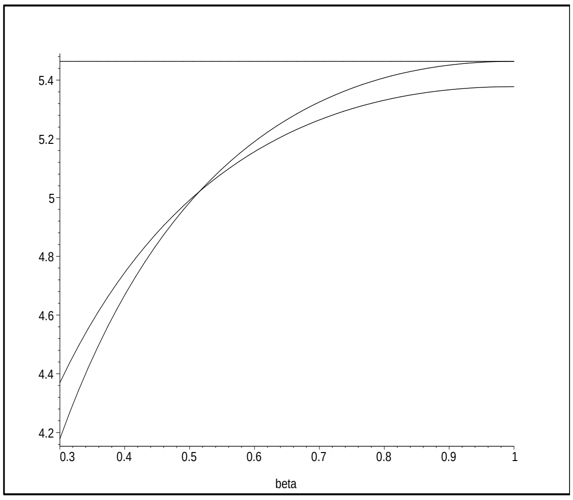

Figure 3.8: **Welfare Comparison.** The figure shows welfare levels as function of β, given $b = 2$ and $v = 4$. The top straight line depicts social welfare with TCs. The curve that coincides with social welfare with TCs at $\beta = 1$ is social welfare with naif agents. The curve that lies below at $\beta = 1$ is the welfare of a time-inconsistent agent without incentives. Social surplus with a naif decreases faster than a naif's welfare without incentives and eventually falls short of it as β decreases.

Proof: The existence of the effect is shown by numerical example. Figure 3.8 plots levels of social welfare against β, given $b = 2$ and $v = 4$. The top straight line depicts social welfare with TCs and is included as benchmark only. The curve that coincides with social welfare with TCs at $\beta = 1$ is social welfare with time-inconsistent agents. The curve that lies below at $\beta = 1$ is the welfare of a time-inconsistent agent without incentives. Given the parameter, social surplus with a naif decreases faster than a naif's welfare without incentives and eventually falls short of a naif's welfare without incentives. Social welfare is then lower if the principal offers an incentive contract as compared to the situation without incentives. With a sophisticate this cannot happen, even though both a sophisticate's welfare without incentives and social welfare with incentives coincide with the respective curves for a naif. With naifs social welfare falls short of a time-inconsistent agent's welfare without incentives exactly at the value of where the principal's profit with sophisticates falls negative. For such low values of γ the principal would thus not offer not offer an incentive contract to a sophisticate. With naifs her rises as γ decreases. q.e.d.

Proposition 3.4 shows that incentives can have detrimental effects. From a very different perspective, beginning with Titmuss (1970), there exists a literature discussing negative effects of incentives. The main argument is ‘motivation crowding out’. According to this theory extrinsic rewards can be harmful because they may destroy intrinsic motivation. See Frey and Jegen (2001) for an overview of both ‘crowing theory’ and empirical evidence. For recent experimental evidence see Gneezy and Rustichini (2000a, 2000b) and Fehr and Gächter (2002). In the context of our model the cause of the detrimental effect of incentives is very different. We show that naifs are harmed by incentives because the presence of incentives increases the mistake they make due to their present-biased preferences. By definition of naiveté, the agent does not anticipate this behavior and thus does not get compensated for this mistake. Furthermore, we have shown that the principal’s profit from providing incentives can be smaller than the loss that accrues to a naive agent. In this case, social welfare is reduced by the presence of incentives.

3.4 Conclusion

In this paper we have analyzed self-control problems in a multi-tasking environment. While the existing literature analyzed environments in which a present-biased agent must decide *when* to accomplish a single task, in this model we look at situations in which an agent must allocate effort *between* multiple tasks and decide *how much* effort to exert. We furthermore assumed that effort devoted to different activities pays out at different points in time.

More specifically, a principal offers a linear incentive contract for an immediately rewarded task. Agents must allocate effort between this task and a private activity that pays out only tomorrow. Such an activity could be, for example, caring for one’s health or continuing to go to school. Effort costs accrue immediately. There are three different types of agents. Time-consistent agents, sophisticated agents, and naive agents. Throughout the paper we assumed away complications of adverse selection. It was thus assumed that the principal knows the type of agent he is contracting with. We find that present-biased agents take decisions that do not maximize their long-run welfare, irrespective of the intensity of incentives. Sophisticated agents are never harmed by incentives relative to the case without incentives as they always receive their reservation utility levels. However, naive agents can be worse off in the presence of incentive contracts as compared to the case without incentives. On the one hand, they wrongly expect a high reservation utility and participate only if they are paid high enough wages. On the other hand, they wrongly predict tomorrow's effort choices which is exploited by the principal. Furthermore, even though agents with time-inconsistent preferences are always harmed by their present bias, the principal's profit with naifs can increase as the present-bias becomes more severe. Finally we show that social welfare can *decrease* in the presence of incentives even though the principal offers an efficient additional production opportunity. With naive agents it can happen that the additional welfare loss due to present-biased effort choices in the presence of incentive contracts exceeds the principal's gain from offering the incentive contract. The model thus offers a new theoretical possibility of detrimental effects of incentives, complementary to existing arguments like the crowding out of intrinsic motivation by extrinsic rewards.

The results of the paper were shown assuming explicit utility functions. The exact conditions under which the effects highlighted in this paper hold true remain to be identified in future research activity. However, their existence could be established.

An possible extension of the model will be to drop the assumption that the principal knows the type of agents the is contracting with and analyze possible screening contracts. Another possible extension will be the analysis of a setting with immediate private benefits and delayed wage payment. There are plenty natural situations imaginable where this constellation is of relevance. While continuing to work on one's Ph.D. thesis after 5 p.m. pays out only with delay, the private benefit of a relaxed evening however accrues immediately.

Chapter 4

Inequity Aversion and Moral Hazard with Multiple Agents*

4.1 Introduction

We analyze how inequity aversion (Fehr and Schmidt (1999), Bolton and Ockenfels (2000)) interacts with incentive provision in an otherwise standard moral hazard model with multiple agents.[39] The theory of inequity aversion assumes that some but not all agents suffer a utility loss if their own material payoffs differ from the payoffs of other agents in their reference groups. The approach can explain a large variety of seemingly diverging experimental findings that often conflict with the standard assumption of pure selfishness.[40] This paper goes a step further and applies the theory of inequity aversion to the theory of incentives. If agents do not simply maximize their own material payoffs but also care for other agents' payoffs they will respond differently to incentives than predicted under the assumption of pure selfishness. Incorporating social preferences into the theory of incentives – thereby either exploiting them or paying tribute to an additional constraint – may help to understand why real world contracts often differ from those contracts found optimal by the standard theory.

*The chapter is based on joint work with Ferdinand von Siemens from the University of Munich.

[39] See Grossman and Hart (1983) and Mookherjee (1984)

[40] For an overview of the literature see, for example, Fehr and Schmidt (2003) and Camerer (2003).

In a classic contribution to the theory of incentives Holmström and Milgrom (1991, p. 24) state that "it remains a puzzle for this theory that employment contracts so often specify fixed wages and more generally that incentives within firms appear to be so muted, especially compared to those of the market." The authors offer an explanation for the paucity of incentives based on the assumption that agents conduct multiple tasks, and that tasks are measured with varying degrees of precision.

We offer an alternative, behavioral explanation to account for the observation that incentives offered to employees within firms are generally 'low-powered' compared to 'high-powered' incentives offered to independent contractors. We assume that within firms social comparisons are pronounced whereas in the marketplace they are negligible.[41] We further assume that an agent suffers a utility loss if another agent conducting a similar task within the same firm receives a higher wage. We find that behindness aversion (suffering only when being worse off) unambiguously increases agency costs of providing incentives. As a consequence, behindness aversion may render equitable flat wage contracts optimal – even though incentive contracts are optimal with selfish agents. Hence, within firms where social comparisons are significant we find 'low powered' flat wage contracts to be optimal, whereas 'high powered' incentive contracts will be given to 'unrelated agents' in the marketplace.

Furthermore, we argue that our analysis can contribute to the question of the optimal size of a firm. Suppose the principal can set up different firms, but setting up a firm involves fixed costs. The principal now faces a trade-off. On the one hand, 'integration' of several agents within a single firm causes social comparisons and, as shown in this paper, increased agency costs of providing incentives. On the other hand, 'separation' of agents into different firms involves additional fixed costs. The solution to this trade-off defines, in the context of this model, the optimal degree of integration.

More specifically, in this paper we derive optimal moral hazard contracts assuming risk- and inequity averse agents that constitute each other's reference group. The agents however do not compare themselves to the principal. Agents carry out identical tasks and regard it as unfair if their wage payments differ. We further assume

[41]See, for example, Bewley (1999) for supporting evidence.

that the principal is both risk neutral and selfish. To keep the analysis tractable we consider the most simple set-up with two agents, two effort levels and two possible output realizations; to receive closed form solutions we assume an explicit utility function and a linear inequity term as in Fehr and Schmidt (1999). In the appendix we however show that our results hold true (1) for any concave utility function and (2) irrespective of the functional form of disutility from inequity. Effort is taken to be non-contractible such that incentive compatible wages must condition on stochastic output realizations. Hence, agents suffer if output realizations and thus wages differ. We show that behindness aversion among agents unambiguously increases agency costs of providing incentives. This also holds true if agents, in addition, suffer from being better off unless they account for effort costs in their comparisons.

The intuition behind this finding can be seen as follows. Inequity aversion effects an utility loss if output realizations diverge. The resulting, reduced utility levels could be implemented without inequity aversion as well, simply by lowering the wages. Since these lower utility levels were not optimal without inequity aversion, they cannot be optimal now.

Increased agency costs can undermine efficiency in two ways. First, equitable flat wage contracts may become optimal even though incentive contracts are optimal with selfish agents. Second, to avoid social comparisons the principal may employ one agent only, thereby forgoing the efficient effort provision of the other agent. This second effect of inequity aversion is qualitatively different from the impact of risk aversion on optimal contracts. The principal can respond to high degrees of risk aversion only by waiving incentives and offering flat wages, whereas with inequity aversion – or more generally with social preferences – he has an additional instrument at hand if he can control an agent's reference group. It is possible to eliminate inequity and still provide incentives to at least one agent. We call this the 'reference group effect'. Third, endowing the principal with the option to set up a second firm at a fixed cost allows to analyze whether 'integration' or 'separation' is optimal.

Further results are derived. Since optimal wages condition on the output realization of the respective other agent as well, the sufficient statistics result due to Holmström

(1979) does not apply. We find that inequity aversion renders team contracts optimal even if output is uncorrelated. Analyzing the interaction between risk and inequity aversion, we find that the additional agency costs due to inequity aversion are higher, the higher the degree of risk aversion. With risk neutral agents inequity aversion does not impact equilibrium agency costs as long as no limited liability constraint binds. Finally, labor contacts often encompass a clause prohibiting employees to communicate their salary. At first sight, inequity aversion could serve as an explanation for this observation. We however show that secrecy of salaries only further increases the additional agency costs due to inequity aversion.

Related Literature

Itoh (2003) and Demougin and Fluet (2003) are most related to our paper. Itoh (2003) analyzes how inequity aversion among risk neutral agents changes optimal incentive contracts, assuming limited liability to be the source of moral hazard. In contrast to our results, Itoh finds that inequity aversion can never harm the principal. With risk neutrality the principal can always choose a fully equitable contract out of the set of contracts that are optimal without inequity aversion. Moreover, inequity aversion can even increase the principal's profit. With limited liability the principal may be forced to pay the agents rents to provide incentives because there is a lower bound on agents' wage payments. However, inequity aversion enables the principal to punish an agent harsher than paying the lowest possible wage level, simply by paying other agents more, thereby reducing agents' rents. Demougin and Fluet (2003) also analyze a two agents moral hazard problem assuming risk neutrality and limited liability. They compare group and individual bonus schemes for behindness-averse agents and derive conditions under which either scheme implements a given effort level at least costs.

Inequity aversion between multiple agents is also analyzed by Rey Biel (2003) and Neilson and Stowe (2003). Rey Biel (2003) analyzes a setting with two inequity averse agents and a principal in which agents' effort choices deterministically translate into output. He exogenously assumes the participation constraint to be slack and finds that the principal can always exploit inequity aversion to extract more rents from his agents. Neilson and Stowe (2003) restrict their analysis to linear piece-rate contracts and identify the conditions under which other-regarding preferences lead workers to exert more or less effort than selfish agents, and whether the optimal piece rate is higher or lower for inequity averse agents.

Englmaier and Wambach (2003) and Dur and Glaser (2004) consider comparisons between agents and principal. Englmaier and Wambach (2003) find that the sufficient statistics result does not apply and that inequity aversion causes a strong tendency towards linear sharing rules. Dur and Glaser (2004) show that inequity aversion can be a reason for high incentives, even for profit sharing, as this reduces inequity.

In Bartling and von Siemens (2004) we analyze how incentive provision in team production is affected if agents are inequity averse. In contrast to the classic result by Holmström (1982) we find that efficient effort choices can be implemented by simple budget-balancing sharing rules if agents are sufficiently inequity averse. Conditions for efficiency become less restrictive the smaller the team. This fits common observation that small teams often work well whereas larger ones suffer from free-riding.

The remainder of the paper is organized as follows. Section 4.2 presents the basic model. In Section 4.3 we derive the optimal incentive contracts for inequity averse agents. Section 4.4 presents our main results. Section 4.5 explores the implications of our results for the optimal firm size. Section 4.6 analyzes the case with secret salaries. In Section 4.7 we discuss comparison of rents, disutility from being better off, and status preferences. Section 4.8 concludes. In the Appendix we discuss the generality of our results.

4.2 The Model

4.2.1 Projects, Effort, and Probabilities

Suppose a principal can employ two risk averse agents. If employed, each agent manages a project with stochastic output $x \in \{x_l, x_h\}$, where $x_h > x_l$ and $\Delta x := x_h - x_l$. Each agent faces a binary effort choice. He either exerts effort, $e = 1$, or he shirks, $e = 0$.

Effort costs are denoted by $\psi(1) = \psi > 0$ while shirking is assumed to be costless, $\psi(0) = 0$. If an agent exerts effort, the output of his project is x_h with probability π and x_l with probability $1 - \pi$, where $\pi \in]0, 1[$. If an agent shirks, the output of his project is always x_l. Effort is assumed not to be contractible. The agents' projects are independent, their production outcomes are uncorrelated.

4.2.2 Preferences: Risk- and Inequity Aversion

We depart from the standard literature by assuming that agents are inequity averse in the sense of Fehr and Schmidt (1999).[42] We assume that an agent's utility is additively separable in the following three components. First, each agent enjoys utility $u(w)$ from his wage payment w by the principal. To derive explicit results we assume this utility function to take on the specific form[43]

$$u(w) = (-1 + \sqrt{1 + 2rw})/r. \tag{4.1}$$

This function is strictly increasing and convex for all $w > -1/2r$. Thus, the agent is risk averse with respect to his income. The corresponding inverse function $h(x) := u^{-1}(x) = x + rx^2/2$ is well defined for all $x > -1/r$. For small w, r can be considered as the agent's approximated degree of absolute risk aversion. This approximation is correct at a zero wage: $-u''(w)/u'(w)|_{w=0} = r$. Second, an agent incurs effort costs ψ if he works; shirking is costless. Finally, an agent suffers from inequity. We assume an agent's reference group to be confined to the other agent, thus the agents do not compare themselves to the principal. Agents carry out an identical task and regard it as unfair if wage payments differ. Since the principal conducts a different 'task' his payoff is not taken to be a point of reference. The identification of an agent's relevant reference group will, however, ultimately be an empirical question.

In the body of the paper we restrict attention to 'behindness aversion'. Whenever

[42] See Bolton and Ockenfels (2000) for a related formulation of inequity aversion.

[43] In the appendix we show that our results neither hinge upon this explicit utility function nor on the assumed linear formulation of inequity aversion by Fehr and Schmidt (1999). The chosen functional forms however allow to derive closed from solutions.

an agent receives a lower payoff than the other agent he suffers a utility loss, but agents do not suffer if they are better off than the other agent. More formally, suppose agent $i \in \{1, 2\}$ receives wage w_i, whereas agent $j \neq i$ receives wage w_j. Agent $i's$ utility function can then be written as

$$v_i(w_i, w_j) = u(w_i) - \psi(e) - \alpha \cdot \max[u(w_j) - u(w_i), 0]. \tag{4.2}$$

The parameter $\alpha \geq 0$ is a measure of behindness aversion. The higher α the more an agent suffers from inequity. Notice that the above formulation does *not* imply that agents compare utilities interpersonally, but rather that agent i suffers from the inequity between the utility he obtains from wage w_i and the utility he would enjoy when receiving the higher wage w_j himself. Both agents maximize expected utility.

Despite the evident experimental evidence on inequity aversion it is still an open question what *exactly* people compare; whether they focus, for example, on wage payments or utility from wage payments, and whether they account for differences in effort costs or not.[44] In this paper, we assume that agents compare utility levels as this renders the principal's maximization problem well behaved.[45] To avoid tedious case distinctions we neglect the possibility that agents account for effort costs in their comparisons. In Section 4.7.1 we however show that accounting for effort costs in the inequity term does not conflict with but rather reinforces the qualitative results of this paper. In Section 4.7.2 we show that introducing suffering from being better off, again, only reinforces our qualitative results *unless* agents account for effort costs in their comparisons.

The principal is both risk-neutral and unaffected by inequity concerns. He maximizes expected output minus expected wage payments.

[44]For a more detailed discussion of inequity aversion see Fehr and Schmidt (1999, 2003).

[45]Otherwise constraints are not linear, the maximization problem not concave, and the solution not straightforwardly characterized by first-order conditions.

4.3 Contracts

We focus on symmetric contracting such that the principal offers identical contracts when employing both agents. The principal has three options. He can either employ both agents and implement effort or shirking, or he can decide to employ one agent only to avoid social comparisons.[46] In the following section we derive optimal contracts implementing these effort choices.

4.3.1 Benchmark: The Single Agent Case

The principal can avoid social comparisons by employing one agent only. Recall that we have confined an agent's reference group to the respective other agent working with the same principal. With a single agent inequity aversion is thus irrelevant. The optimal contract for the employed agent (incentive or flat wage contract) then depends on the standard parameters of the model via the participation and incentive constraint. Suppose first the principal wants to implement high effort. Since effort is not verifiable wages must condition on stochastic output realizations and the classic risk-incentive trade-off arises. Define w_i as the agent's wage if his output is $i \in \{h, l\}$, and define $u_i := u(w_i)$. To render the principal's maximization problem concave, we rewrite the principal's objective function and the constraints in terms of u_h and u_l. An agents outside option is normalized to zero. The resulting first-order conditions then yield

$$u_h^* = \frac{\psi}{\pi} \qquad \text{and} \qquad u_l^* = 0 \tag{4.3}$$

as the optimal contract, and profit can be written as

$$P_1^i = \pi x_h + (1-\pi)x_l - \left[h(\psi) + \frac{r\psi^2(1-\pi)}{2\pi}\right] \tag{4.4}$$

[46] In principle, he could also offer a 'hybrid contract': an incentive contract to one agent and a 'non-incentive contract' to the other agent. Note that due to inequity aversion such a 'non-incentive contract' would not be a flat wage contract. It can be shown that considering the 'hybrid contract' would not change the qualitative results of this paper.

where superscript i denotes 'incentive contract' and the subscript shows the number of agents employed. Define

$$RAC := \frac{r\psi^2(1-\pi)}{2\pi}$$

as the 'risk-agency-costs' that have to be payed on top of the first-best cost of effort implementation $h(\psi)$ due to risk aversion.

Suppose now the principal offers a flat wage contract. The agent then never exerts effort and the participation constraint is satisfied at flat wage $w^f = 0$. The principal's profit in this case is $P_1^f = x_l$. The difference in expected profit from implementing effort as compared to paying a flat wage is given by

$$B := \pi\Delta x - h(\psi) - RAC.$$

Thus, it is optimal for the principal to implement high effort if and only if

$$B \geq 0 \quad \Leftrightarrow \quad \pi\Delta x \geq h(\psi) + RAC. \tag{4.5}$$

The principal offers an incentive contract whenever the expected output increase is sufficiently large relative to the first best cost of implementing effort and the RAC. The condition is more likely to be met if effort cost and risk aversion are small and the information content of the project outcomes is high. Exerting effort is efficient if $\pi\Delta x \geq h(\psi)$ but risk aversion leads to a trade-off between insurance and efficiency and causes additional RAC. This leads to inefficient effort choices if $h(\psi) + RAC \geq \pi\Delta x \geq h(\psi)$. If $\pi\Delta x \geq h(\psi) + RAC$ the efficient effort level is implemented but risk aversion reduces the principal's expected profit.

In the next section we show that inequity aversion amplifies these effects. Inequity aversion causes additional agency costs which unambiguously rise as the level of inequity aversion rises. This further reduces the principal's expected profit, and it can lead to additional inefficiencies. Throughout the paper we therefore assume incentive condition (4.5) to be fulfilled. $B < 0$ is the uninteresting case since flat wage contracts would then always be optimal – even without inequity aversion.

4.3.2 The Two Agents Case

In this section we consider the two agents case. Both agents work within the same firm and we thus assume that they compare their wage levels. An agent suffers a utility loss in case he is behind. In contrast, we assume that an agent would not compare his wage to the wage of an agent with whom he only interacts in the market, i.e. an agent that works for another principal.

With incentive contracts inequity arises naturally as output is stochastic and incentive compatible wages must condition on output realizations. At first glance the effect of inequity aversion on agency costs is ambiguous. Behindness aversion increases incentives because exerting effort reduces the probability of being behind. At the same time agents anticipate that even if they exert high effort with positive probability they will be behind. Ex ante agents have to be compensated for this expected utility loss to ensure participation.

We show that the positive effect on incentives is always dominated by the negative effect on participation and, therefore, behindness aversion unambiguously *increases* the agency costs of providing incentives. The intuition can be seen as follows. Without inequity aversion the second-best optimal incentive contract assigns wage levels to each possible output realization such that both IC and PC are fulfilled and binding. For some output realizations (i.e. agent one is successful, agent two is not) the contract assigns diverging wage levels to the agents (agent one receives a higher wage than agent two, assuming the monotone likelihood ration to hold). If now inequity aversion is considered, the utility of agents receiving less than others (agent two) is reduced by the amount of suffering from being behind. However, this lower utility level could have been achieved without inequity aversion as well – simply by lowering the respective wage level, which reduces the principal's cost. As this was not optimal without inequity aversion it cannot be optimal now.

In the appendix we show that this intuition holds generally. Assuming only concavity of the utility function we show that inequity aversion renders it weakly more expensive to implement each possible effort level. However, our arguing does not hold when there is limited liability. With limited liability the lowest possible wage payment and thus the lowest possible utility level for an agent is bounded from below. To provide incentives the principle may thus be forced to leave the agents rents. In this case inequity aversion provides the principal with the possibility to reduce the lowest possible utility level. An agent can now not only be punished by paying out the lowest wage level but in addition by paying other agents a higher wage. The lowest possible utility level for an agent can thus be reduced without violating the limited liability constraint. This in turn enables the principal to reduce the agents' rents.

Suppose first the principal does not want to implement effort. He then offers two flat wage contracts. Since there is never inequity, inequity aversion is irrelevant and the principal's profit is simply

$$P_2^f = 2 \cdot P_1^f = 2\, x_l. \tag{4.6}$$

Suppose now the principal wants to implement effort. We show that the principal's expected profit is not just twice the expected profit in the single agent case but $P_2^i \leq 2\, P_1^i$. As both agents are symmetric, we assume that optimal wages are symmetric in the sense that they condition on the output realizations of both projects but not on the identity of the agent. Denote by w_{ij} the wage of an agent with output i if the other agent's output is j. Define $u_{ij} := u(w_{ij})$ as an agent's utility from wage w_{ij}. As there are four possible states of the world, a contract determines four wage levels: w_{ll}, w_{hh}, w_{lh}, and w_{hl}, where h stands for high and l for low output. To render the principal's maximization problem concave with linear constraints, we rewrite the principal's objective function and the constraints in terms of u_{hh}, u_{hl}, u_{lh}, and u_{ll}.

Recall that the maximum functions in the agents' utility functions in (4.2) create potential kinks. At these points, the utility functions and thus the PC and IC are not differentiable, potentially rendering it impossible to characterize optimal contracts by first-order conditions. However, the following lemma allows to avoid this problem.

Lemma 4.1 *The optimal incentive compatible contract for two inequity averse agents satisfies* $u^*_{hl} \geq u^*_{lh}$.

Proof: Suppose this was not the case, that is $u_{hl} < u_{lh}$ at the optimum. Then the IC and PC are given by

$$\begin{aligned}
&(\text{I C'}) \qquad && \pi^2 u_{hh} + \pi(1-\pi)[u_{hl} - \alpha(u_{lh} - u_{hl})] - \pi^2 u_{lh} - \pi(1-\pi)u_{ll} - \psi \ \geq 0 \\
&(\text{PC'}) \qquad && \pi^2 u_{hh} + \pi(1-\pi)[u_{hl} - \alpha(u_{lh} - u_{hl})] + \pi(1-\pi)u_{lh} - (1-\pi)^2 u_{ll} - \psi \ \geq 0
\end{aligned}$$

Consider changes $du_{lh} < 0$ and $du_{hl} = -du_{lh}(1-\alpha)/(1+\alpha)$. This leaves (PC') unaffected but improves (IC'). The principal's profit increases by

$$dP_2^i = 2\pi(1-\pi)\left[h'(u_{hl})\frac{1-\alpha}{1+\alpha} - h'(u_{lh})\right] du_{lh},$$

which is strictly larger than zero as $(1-\alpha)/(1+\alpha) \leq 1$, $du_{lh} < 0$, $u_{hl} < u_{lh}$, and $h''(u) > 0$. q.e.d.

Notice that $du_{lh} < 0$ has a twofold effect on (PC'). On the one hand, this decreases the agent's utility if his own project fails whereas the other agent's project is successful. On the other hand, unfavorable inequity decreases if the agent himself is successful whereas the other agent is unfortunate. In the latter case the agent's utility increases. If the inequity reducing effect dominates, $\alpha > 1$, the principal may decrease both u_{hl} and u_{lh} while keeping (PC') unaffected and not impairing (IC'). In either case, the principal can increase his expected profit without violating a constraint, and $u_{hl} < u_{lh}$ cannot be optimal.

By Lemma 4.1 we can introduce an additional constraint, $u_{hl} - u_{lh} \geq 0$, without restricting the attainable maximum. We call this constraint the *Order Constraint* (OC). The maximum functions in the agents' utility functions are thus removed and the principal maximizes

$$P_2^i = 2\left[x_l + \pi^2\left[\Delta x - h(u_{hh})\right] + \pi(1-\pi)[\Delta x - h(u_{lh}) - h(u_{hl})] - (1-\pi)^2 h(u_{ll})\right] \tag{4.7}$$

with respect to u_{hh}, u_{hl}, u_{lh}, and u_{ll}, where

$$
\begin{aligned}
&\text{(IC)} \qquad & \pi^2 u_{hh} + \pi(1-\pi)u_{hl} - \pi^2[u_{lh} - \alpha(u_{hl} - u_{lh})] - \pi(1-\pi)u_{ll} - \psi \ &\geq 0 \\
&\text{(PC)} \qquad & \pi^2 u_{hh} + \pi(1-\pi)u_{hl} + \pi(1-\pi)[u_{lh} - \alpha(u_{hl} - u_{lh})] + (1-\pi)^2 u_{ll} - \psi \ &\geq 0 \\
&\text{(OC)} \qquad & u_{hl} - u_{lh} \ &\geq 0.
\end{aligned}
$$

are the constraints characterizing the principal's choice set.[47]

We begin by assuming that the OC is not, whereas the IC and PC are binding. Solving the resulting first-order conditions then yields

$$u_{hh}^* = \frac{\psi}{\pi} + \frac{\alpha(1-\pi)(1+\alpha(\pi+r\psi))}{rk} \tag{4.8}$$

$$u_{hl}^* = \frac{\psi}{\pi} - \frac{\alpha\pi(1+\alpha(\pi+r\psi))}{rk} \tag{4.9}$$

$$u_{lh}^* = \frac{\alpha(1-\pi(1+\alpha\pi)+(1+\alpha)r\psi)}{rk} \tag{4.10}$$

$$u_{ll}^* = -\frac{\alpha(\pi(1+\alpha(1+\pi))-r\psi)}{rk} \tag{4.11}$$

where $k = 1+\alpha\pi(2+\alpha(1+\pi))$. The Lagrange multipliers for the PC and IC are given by $\mu = 2(1+r\psi[1+\alpha(1+\pi)+2\pi\alpha^2]+2\pi\alpha(1+\pi\alpha))/k$ and $\lambda = 2(1-\pi)(r\psi[1+\pi\alpha+2\pi\alpha^2]+\pi\alpha(1+2\pi\alpha))/\pi k$. Since these are strictly positive, both the PC and IC are indeed binding as initially assumed. We also have to check whether it holds true that the OC is slack, i.e. whether we have $u_{hl}^* \geq u_{lh}^*$. The difference is given by $(r\psi+\alpha\pi(r\pi-1))/r\pi k$. Thus, the OC is indeed slack and the solutions (4.8) - (4.11) are valid if and only if either $r\psi \geq 1$ or $r\psi < 1$ and $\alpha < \tilde{\alpha}$ where

$$\tilde{\alpha} := r\psi/(\pi(1-r\psi)). \tag{4.12}$$

[47] We do not consider dominant strategy implementation in this paper, i.e. we only look at contracts such that the constraints are satisfied for one agent given that the other agent behaves as expected. Even though both agents participating and exerting effort then forms a Nash equilibrium it is possibly not unique.

Finally, since $h(u)$ is defined for $u \geq -1/r$ only, we have to verify that this always holds. Algebraic manipulations show that $u^*_{lh} \geq u^*_{ll}$. The solution is thus valid if $u^*_{ll} \geq -1/r$, or $ru^*_{ll} \geq -1$. This condition holds with equality if $r = \tilde{r} = -(1+\alpha\pi)/(\alpha\psi)$. Differentiating ru^*_{ll} with respect to r yields $\alpha\psi/k \geq 0$. Hence, ru^*_{ll} rises in r. Since we must have $r > 0$, r always exceeds $\tilde{r}$, and u^*_{ll} never falls short of $-1/r$.

Suppose now that all the constraints PC, IC, and OC are binding. The binding OC forces the principal to set $u_{lh} = u_{hl}$. This restriction on the contract design eliminates inequity but comes at a cost. Solving the corresponding first-order conditions we get

$$u^*_{hh} = \frac{\psi}{\pi} + \frac{(1-\pi)\psi}{\pi} \tag{4.13}$$

$$u^*_{hl} = u^*_{lh} = \frac{\psi}{\pi} - \psi = \frac{(1-\pi)\psi}{\pi} \tag{4.14}$$

$$u^*_{ll} = -\psi. \tag{4.15}$$

The Lagrange multipliers of the PC and IC are $\mu = 2r\psi + 2$ and $\lambda = 4r\psi(1-\pi)/\pi$. Since both are strictly positive, the PC and IC are indeed binding as initially assumed. As $h(u)$ is defined for $u \geq -1/r$ only, the above solution is valid only if $r\psi < 1$.

The overall optimal solution depends on whether the OC is binding or not, which in turn depends on r,ψ and α. This is summarized in the following proposition.

Proposition 4.1 (Optimal Contracts For Inequity-Averse Agents)

i) Suppose $r\psi \geq 1$. The optimal incentive compatible contract for two inequity averse agents is given by (4.8) - (4.11).

ii) Suppose $r\psi < 1$. If $\alpha < \tilde{\alpha}$, the optimal incentive compatible contract for two inequity averse agents is given by (4.8) - (4.11). If $\alpha \geq \tilde{\alpha}$, it is given by (4.13) - (4.15).

Proof: There are two cases. First, suppose $r\psi \geq 1$. Then solution (4.13) - (4.15) is not valid as $u^*_{ll} < -1/r$, whereas solution (4.8) - (4.11) is valid for all α as we always get $(r\psi + \alpha\pi(r\pi - 1))/r\pi k > 0$. Second, suppose $r\psi < 1$. Then for all $\alpha < \tilde{\alpha}$ both extreme points are candidates for the overall solution, but (4.8) - (4.11) dominates as the maximum is not restricted by the OC. For all $\alpha \geq \tilde{\alpha}$, only solution (4.13) - (4.15) is valid. q.e.d.

4.4 Results

4.4.1 Inequity Aversion Renders Team Contracts Optimal

Since we assume output to be uncorrelated an agent's output realization does not contain information about the other agent's effort choice. According to the classic result by Holmström (1979) optimal wages should only condition on sufficient statistics for effort choices. In our model wages should thus only condition on the own output realization. Nonetheless, since agents compare the utility levels from their wages optimal contracts also condition on the other agent's output realization in order to reduce inequity. Therefore, the sufficient statistics result does not apply.[48] Define a *team contract* as a compensation scheme such that an agent's wage depends positively on the other agent's success. Thus, in a team contract we have $w_{hh} > w_{hl}$ and $w_{lh} > w_{ll}$. As summarized in the following proposition inequity aversion renders team contracts optimal.

Proposition 4.2 (Team Contracts)
The sufficient statistics result does not apply: Inequity aversion renders team contracts optimal even if output is uncorrelated.

Proof: Comparison of the relevant utility levels in Proposition 4.1 yields $u^*_{hh} - u^*_{hl} = \alpha(1 + \alpha(\pi + r\psi))/rk \geq 0$ and $u^*_{lh} - u^*_{ll} = \alpha(1 + \alpha(\pi + r\psi))/rk \geq 0$. q.e.d.

Since output is stochastic, agents obtain different output realizations with positive probability even though both agents exert high effort. The unfortunate agent then

[48] In the context of interdependent preferences this result naturally arises. It was first shown in Englmaier and Wambach (2003).

suffers from obtaining a lower wage than the fortunate agent. The optimal contract accounts for this effect and adjusts wage levels accordingly.

4.4.2 Inequity Aversion Causes Additional Agency Costs

In the benchmark case of a single agent inequity aversion is irrelevant and does not influence the principal's profit. This is also the case with flat wage contracts for two agents as there is never inequity. However, with incentive contracts for two inequity averse agents additional agency costs arise. Suppose $r\psi \geq 1$ or $r\psi < 1$ but $\alpha < \tilde{\alpha}$ such that the optimal contract is characterized by (4.8) - (4.11). Substituting optimal utility levels, the principal's maximum profit is then given by

$$P_2^i = 2\,P_1^i - IAC, \tag{4.16}$$

where

$$IAC := \frac{\alpha(1-\pi)(2r\psi + \pi\alpha(r\psi - 1) + \alpha r^2\psi^2)}{rk}. \tag{4.17}$$

denotes the 'inequity agency costs', the additional agency cost due to inequity aversion. Inequity aversion has a negative effect on the principal's maximum profit as the above solution is only valid if either $r\psi \geq 1$ or $\alpha \leq \tilde{\alpha}$ holds, and this ensures that IAC are positive. Equivalently, suppose $r\psi < 1$ and $\alpha < \tilde{\alpha}$ such that the optimal contract is characterized by (4.13) - (4.15). Substituting optimal utility levels, the principal's maximum profit is then given by

$$P_2^i = 2\,P_1^i - IAC, \tag{4.18}$$

where

$$IAC := \frac{r\psi^2(1-\pi)}{\pi} \tag{4.19}$$

denotes the 'inequity agency costs' in this case. Again, the principal's profit with two hard working agents is strictly less than twice the profit with only one hard working agent as the IAC are always positive. Note that in the latter case the IAC do not depend on α as the above solution is subject to the OC binding and inequity is completely eliminated. However, inequity aversion reduces the principal's profit as it forces him to set $u^*_{hl} = u^*_{lh}$ via the binding OC. We can now derive the following result.

Proposition 4.3 (Additional Agency Costs)
Inequity aversion among agents causes additional agency costs of implementing effort. These agency costs weakly increase and converge as the level of inequity aversion rises.

Proof: Suppose $r\psi \geq 1$. The IAC are then given by (4.17) and $r\psi \geq 1$ ensures (4.17) > 0. Differentiating (4.17) with respect to α yields

$$\frac{\partial\, IAC}{\partial \alpha} = \frac{2(1-\pi)(1+\alpha(r\psi+\pi))(r\psi(1+\alpha\pi)-\alpha\pi)}{rk^2}, \tag{4.20}$$

which is strictly positive as $r\psi \geq 1$. The limit of IAC is given by

$$\lim_{\alpha\to\infty} IAC = \frac{1-\pi}{1+\pi}\left[r\psi(2\pi+r\psi)-1\right], \tag{4.21}$$

where $r\psi \geq 1$ again ensures the expression to be positive. Suppose now $r\psi < 1$. In case $\alpha \leq \tilde{\alpha}$ the above arguments on sign of IAC and their derivative w.r.t. α apply. In case $\alpha > \tilde{\alpha}$ the IAC are given by (4.19) which is positive as we have $r > 0$, $\psi > 0$, and $\pi \in]0,1[$, does not change in α, and is thus equal to the limit as $\alpha \to \infty$. q.e.d.

Proposition 4.3 proves that the negative effect of inequity aversion on the PC always dominates the positive effect on the IC. The negative effect of inequity is however bounded because the principal can always equate w_{hl} and w_{lh} if α becomes too large. The optimal contract then remains unchanged as α further increases. The intuition for the dominance of the effect on the PC can best be seen when approaching the problem from a different angle. Inequity aversion effects a utility loss in certain states of the world. If the resulting reduced utility level were second-best optimal, then they could be realized without inequity aversion as well – simply by lowering wage payments. As lower utility levels are not second-best optimal without inequity aversion, they cannot be optimal now. In the appendix we show that this intuition straightforwardly generalizes to less restrictive settings.

As an alternative intuition for the result consider the following. Suppose the OC is binding. To eliminate suffering from inequity aversion utility levels in case of diverging output realizations are equated. This clearly impairs incentives to exert effort. Hence, in cases with identical output realizations wage payments must become more extreme. Agents then have to bear more risk for which they must be compensated. The same reasoning holds true if the OC is not binding. In addition to the increased risk, agents then also have to be compensated for the inequity they bear despite the wage compression in case output realizations diverge. This leads to the next proposition.

Proposition 4.4 (Complementarity)
The more risk averse the agents, the higher the additional agency costs due to inequity aversion.

Proof: In case the OC does not bind the IAC are given by (4.17). Differentiating (4.17) with respect to r yields

$$\frac{\partial IAC}{\partial r} = \frac{(1-\pi)\alpha^2(\pi + r^2\psi^2)}{rk^2} \tag{4.22}$$

which is unambiguously positive. In case the OC binds and the IAC are given by (4.19) the respective partial derivative is clearly positive. q.e.d.

Since contracts that account for inequity aversion lead to more risk bearing, the higher the degree of risk aversion, the higher the additional agency costs caused by inequity aversion. Risk aversion and inequity aversion thus have complementary effects.

Consider the extreme case of risk neutral agents, i.e. $u(w) = w$. The principal's expected incentive compatible wage payment per agent is then $\psi + \pi(1-\pi)(\alpha+\beta)(w_{hl} - w_{lh})$, the sum of the first-best costs of implementing effort and compensation for inequity bearing. Notice that in the context of this model a limited liability constraint will never bind as we have normalized the success probability when shirking to zero. There is thus no rent that has to be given to the agent, i.e. the PC is binding. Since inequity aversion has an unambiguously negative effect on the PC, any amount of inequity decreases the principal's expected profit. A possible positive effect of inequity aversion on incentive provision cannot be realized since incentives can be provided at first-best costs already. With risk neutral agents (and no limited liability constraint binding) a large set of optimal contracts can implement efficient effort choices at first-best costs. With inequity aversion only a subset of these optimal contracts remains optimal, namely those contracts with $w_{hl} = w_{lh}$. The remaining subset of optimal contracts is however non-empty. For example, the contract with $w_{hh} = \psi/\pi^2$ and $w_{hl} = w_{lh} = w_{ll} = 0$ is always possible. It provides incentives at first-best costs and eliminates all inequity. We summarize our findings in the following proposition.

Proposition 4.5 (Risk Neutrality)
With risk-neutral agents and no limited liability constraint binding, inequity aversion reduces the set of optimal contracts but does not impact the equilibrium outcome.

Itoh (2003) also analyzes a moral hazard setting with risk-neutral agents but assumes limited liability constraints to bind. In this case agents receive a rent. Inequity aversion provides the principal with the possibility to reduce an agent's utility below the level that arises from paying the lowest possible wage level, simply by paying other agents more. Inequity aversion can thus reduce the principal's rent payments in case of effort implementation, and inequity aversion can then have an impact on the equilibrium outcome.

4.4.3 Inequity Aversion and Efficiency

In this section we derive the conditions under which inequity aversion causes an efficiency loss similar to the efficiency loss that arises if risk aversion renders flat wage contracts optimal. There are however two qualitative differences between risk agency costs, RAC, and inequity agency costs, IAC. First, the RAC are unbounded. Therefore, an efficiency loss due to underprovision of effort always occurs if only risk aversion is sufficiently large. In contrast, the IAC are bounded. It can be that no inefficiency arises even if the degree of inequity aversion goes to infinity. The reason is that the principal can always equate wage levels in case of diverging output realizations, thereby eliminating inequity while still providing incentives. It is however not possible to provide incentives and eliminate agents' risk. Second, if the RAC are large the principal can only offer flat wage contracts to avoid the agents' risk exposure. In contrast, there are two means by which inequity can be avoided. As with risk aversion, the principal can either offer flat wage contracts – thereby forgoing profits from effort implementation. We call this case 'underprovision of effort'. Or he can employ a single agent only. Then there is no reference group and thus no social comparisons and no suffering from inequity. The principal will then provide incentives to a single agent – thereby forgoing the profit from employing the second agent. We call this the 'reference group effect'. In the following we identify the conditions under which either case arises.

Underprovision of Effort

Two conditions have to be met such that inequity aversion renders flat wage contracts more profitable than incentive contracts. First, the expected profit from two flat wage contracts must exceed expected profits from a single incentive contract. This condition ensures that offering two flat wage contracts is the best alternative to offering two incentive contracts. Second, for sufficiently high levels of α the IAC must exceed the difference in expected profits from two incentive contracts (without inequity aversion) and two flat wage contracts. With flat wage contracts wages never diverge and inequity aversion is irrelevant. This is summarized in the following proposition.

Proposition 4.6 (Underprovision of Effort)
If and only if $x_l \geq B$ and $2\,B < \lim_{\alpha\to\infty} IAC$, there exists a threshold level of inequity aversion $\hat{\alpha}$ such that for all $\alpha \geq \hat{\alpha}$ flat wage contracts maximize the principal's expected profit, even though incentive contracts are profit maximizing with selfish or unrelated agents.

Proof: The first condition ensures that expected profit from two flat wage contracts exceed expected profits from a single incentive contract. Formally, $P_2^f = 2x_l \geq x_l + B = P_1^i \Leftrightarrow x_l \geq B$. Consider now the second condition. P_2^i denotes the principal's expected profit when offering two incentive contracts. If $\alpha = 0$ we have $P_2^i(\alpha = 0) = 2\,P_1^i$. By assumption, $2\,P_1^i - P_2^f = 2\,B > 0$. Without inequity aversion the principal thus employs both agents and implements high effort. By Proposition 4.3, P_2^i decreases in α and converges to

$$\lim_{\alpha\to\infty} P_2^i(\alpha) = 2\,P_1^i - \lim_{\alpha\to\infty} IAC. \tag{4.23}$$

We thus have $P_2^f > \lim_{\alpha\to\infty} P_2^i$ if and only if

$$2\,B < \lim_{\alpha\to\infty} IAC. \tag{4.24}$$

From (4.17) and (4.19) we know that $\lim_{\alpha\to\infty} IAC > 0$. The parameter space for which (4.24) holds is thus non-empty. If $2\,B$, the gain of providing incentives to two agents, falls short of $\lim_{\alpha\to\infty} IAC$, the limit of the inequity agency costs of providing incentives, there exists a unique threshold level of inequity aversion $\hat{\alpha}$ such that for $\alpha < \hat{\alpha}$ two incentive contracts, and for $\alpha \geq \hat{\alpha}$ two flat wage contracts maximize the principal's expected profit. Existence and uniqueness of threshold $\hat{\alpha}$ is ensured since P_2^i is continuous and strictly decreasing in α. q.e.d.

The left panel of Figure 4.1 provides an illustration of Proposition 4.6. Without inequity aversion, $\alpha = 0$, expected profits from two incentive contracts exceed expected profits from both two flat wage contracts and a single incentive contract. Condition $x_l \geq B$ ensures that the principal's best alternative to offering two incentive contracts is offering two flat wage contracts. As α increases, the IAC increase and reduce the principal's expected profit from two incentive contracts. At $\hat{\alpha}$ the IAC equal the difference in expected profits between two incentive and two flat wage contracts, $2B$. Therefore, for levels of inequity aversion exceeding $\hat{\alpha}$, two flat wage contracts maximize the principal's expected profit.

Proposition 4.6 is the central finding our this paper: inequity aversion can render flat wage contracts optimal even though incentive contracts are optimal with selfish agents. We interpret this as an explanation for the observed 'low powered' incentives within

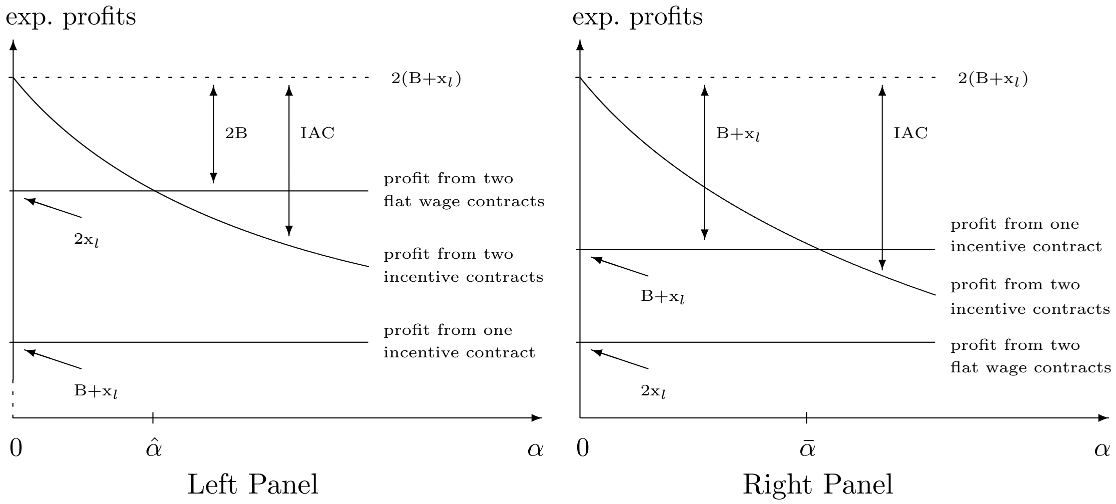

Figure 4.1: **Underprovision of Effort and the Reference Group Effect.**
Left Panel: Expected profit levels for $B < x_l$. In this case expected profits from two flat wage contracts exceed profits from one incentive contract. If the additional agency costs due to inequity aversion, IAC, exceed the difference in expected profits between flat wage and incentive contracts, $2B$, as α increases, then there exists a threshold level $\hat{\alpha}$ such that two flat wage contracts maximize the principal's expected profit for $\alpha \geq \hat{\alpha}$.
Right Panel: Expected profit levels for $x_l < B$. In this case expected profits from one incentive contract exceed profits from two flat wage contracts. If the additional agency costs due to inequity aversion, IAC, exceed the expected profit from an additional incentive contract absent inequity aversion, $B + x_l$, as α increases, then there exists a threshold level $\bar{\alpha}$ such that a single incentive contract maximizes the principal's expected profit for $\alpha \geq \bar{\alpha}$.

firms – as compared to 'high powered' incentives in the market. This interpretation hinges upon the assumption that agents compare their wage payments within firms but not within the market. Although the determinants of an agents reference group will ultimately be an empirical question, co-workers within a firm are a natural candidate for a reference group. However, crucial to our analysis is that there are two agents who compare their wages and dislike inequity. Our results – though not our interpretation – would hold if we assumed two principals, each of them offering an incentive contract to a single agent, and these two agents comparing wages.

The Reference Group Effect

Suppose now that the principal can influence an agent's reference group. Two conditions have to be met such that inequity aversion renders it more profitable for the principal to offer an incentive contract to a single agent than offering incentive contracts to two agents. When two inequity averse agents work for the principal they compare their wage levels and suffer from inequity. In contrast, with a single agent no comparisons take place, and thus no IAC arise. In Section 4.5 we further explore this 'reference group' or 'firm size effect' in a slightly enriched setting; for completeness we now derive the conditions that have to be met in this basic set-up. First, the expected profit from a single incentive contract must exceed expected profits from two flat wage contracts. This condition ensures that offering a single incentive contract is the best alternative to offering two incentive contracts. In contrast to the previous section, here it must hold that $x_l < B$. Second, for sufficiently high levels of α the IAC must exceed the difference in expected profits from offering two incentive contracts (without inequity aversion) and expected profits from offering a single incentive contract. With a single incentive contract inequity aversion is irrelevant as there is no reference group. This is summarized in the following proposition.

Proposition 4.7 (Reference Group Effect)
If and only if $x_l < B$ and $B + x_l < \lim_{\alpha\to\infty} IAC$, there exists a threshold level of inequity aversion $\bar{\alpha}$ such that for $\alpha > \bar{\alpha}$ the principal employs a single agents only to avoid social comparisons, even though employing both agents maximizes the principal's expected profit without inequity aversion.

Proof: As before, $P_2^i(\alpha = 0) = 2\,P_1^i$ such that without inequity aversion it maximizes the principal's expected profit to employ both agents and implement high effort. However, P_2^i decreases as α rises, and it may eventually fall short of P_1^i. As $P_1^i = P_1^f + B$ and $P_1^f = x_l$, it holds that $\lim_{\alpha\to\infty} P_2^i < P_1^i$ if and only if

$$B + x_l < \lim_{\alpha\to\infty} IAC. \tag{4.25}$$

From (4.17) and (4.19) we know that $\lim_{\alpha\to\infty} IAC > 0$, so the parameter space for which (4.25) holds true is non-empty. Whenever the base output, x_l, and the benefit from giving incentives, B, are sufficiently small, there exists a unique level of inequity aversion $\bar{\alpha}$ such that for $\alpha < \bar{\alpha}$ two incentive contracts, whereas for $\alpha \geq \bar{\alpha}$ a single incentive contract maximizes the principal's expected profit. Existence and uniqueness of threshold $\bar{\alpha}$ is ensured since P_2^i is continuous and strictly decreasing in α. q.e.d.

The right panel of Figure 4.1 provides an illustration of Proposition 4.7. Without inequity aversion, $\alpha = 0$, expected profits from two incentive contracts exceed expected profits from both two flat wage contracts and a single incentive contract. Condition $x_l < B$ ensures that the principal's best alternative to offering two incentive contracts is offering a single incentive contract. As α increases, the IAC increase and reduce the principal's expected profit from two incentive contracts – but not the expected profit from a single incentive contract as in this case no social comparisons take place. At $\bar{\alpha}$ the IAC equal the expected profit from an additional incentive contract without inequity aversion, $x_l + B$. Therefore, for levels of inequity aversion exceeding $\bar{\alpha}$, a single incentive contract maximizes the principal's expected profit.

In case neither $x_l \geq B$ *and* $2\,B < \lim_{\alpha\to\infty} IAC$, the conditions stated in Proposition 4.6, nor $x_l < B$ *and* $B + x_l < \lim_{\alpha\to\infty} IAC$, the conditions stated in Proposition 4.7, there is no inefficiency caused by the additional agency cost due to inequity aversion – even if the degree of inequity aversion goes to infinity. The principal is nevertheless harmed by inequity aversion since his expected profit is reduced by the amount of the IAC. In contrast, the RAC will always lead to an inefficiency if only the degree of risk aversion becomes sufficiently large.

The effect of inequity aversion in the case with 'underprovision of effort' is qualitatively similar to the effect of risk aversion. Providing incentives becomes more expensive as either aversion becomes more pronounced, and this may render flat wage contracts optimal for the principal. However, the 'firm size effect' is qualitatively different from the inefficiency that can arise due to risk aversion. The principal can respond to risk aversion only by adopting an agent's contract, whereas with inequity aversion – or more generally with social preferences – he has an additional instrument at hand as he can control the agents' reference groups. Incorporating this finding into richer models with, for example, heterogeneous agents with respect to the degree of inequity aversion or productivity, or allowing for multi-tasking will yield deeper insights into the determinants of real world wage contracts, the optimal design of institutions, and the boundary of the firm. In the following section, while keeping the assumption of homogeneous agents, we enrich the model by allowing the principal to separate the agents into different firm at a fixed cost. We will argue that the interaction between inequity aversion and moral hazard can contribute to the old question of the nature and size of the firm.

4.5 The Nature and Size of the Firm

The 'property rights approach' of the theory of the firm – pioneered by Grossman and Hart (1986) and Hart and Moore (1990) – defines a firm as the physical assets it consists of. In contrast to the 'transaction cost approach' of the theory of the firm (Coase (1937), Williamson (1975, 1985)), the property rights approach can explain both, advantages and disadvantages (better incentives to invest for one, but worse incentives for the other party) of 'integration' within a unified framework. An optimal degree of integration, that is, an optimal firm size can thus be determined. In this section, we propose a new approach. We focus on one characteristic that distinguishes the firm from the market. The firm is seen as an economic entity within which social comparisons matter – in contrast to the market in which they are negligible.

In this section we enrich our model by endowing the principal with the option to separate the agents by setting up an additional firm. We assume that agents compare payoffs only with agents that work within the same firm but not with agents that work in distinct firms.[49] Additional agency costs due to inequity aversion can thus be avoided by separating agents into different firms. If agents can be separated, that is, if social comparisons can be prevented at not cost, the purpose of this paper dissolves. The principal would then always separate the agents. However, we further assume that setting up a firm involves the expense of fixed costs, denoted by F. These fixed costs are taken to be sufficiently low such that the principal realizes positive profits when offering an incentive contract to a single agent, that is $F < P_1^i$. Alternatively,

[49]See, again, Bewley (1999) for supporting evidence.

complementarities in production could be assumed such that, absent inequity aversion, it is advantageous to have agents work together.

The principal now faces a trade off. On the one hand, employing two agents within a single firm economizes on fixed costs (or enables the principal to realize complementarities in production). On the other hand, integrating the agents within a single firm provokes social comparisons that increase agency costs of providing incentives. The solution to this trade-off thus defines the optimal size of the firm, whether there is 'integration' of both agents within a single firm or 'separation' of the agents into two distinct firms. If the firm is integrated, we can have both, incentive and flat wage contracts. In case of separation, the principal will always offer incentive contracts. The following proposition identifies the conditions under which either regime is optimal.

Proposition 4.8 (Optimal Firm Size)

i) If and only if $F \leq \min[\lim_{\alpha\to\infty} IAC, 2B]$ then there exists a threshold level of inequity aversion $\hat{\hat{\alpha}}$ such that for $\alpha \geq \hat{\hat{\alpha}}$ separation is optimal: The principal bears fixed costs F twice to set up two distinct firms, and she offers in each firm a single incentive contract. For $\alpha < \hat{\hat{\alpha}}$ integration is optimal: The principal sets up a single firm and offers two incentive contracts.

ii) If $F > 2B$ then there is always integration, irrespective of the degree of inequity aversion α. If, in addition, $\lim_{\alpha\to\infty} IAC \leq 2B$ then integration with incentive contracts is optimal for all α. If, in addition, $\lim_{\alpha\to\infty} IAC > 2B$ then there exists a threshold level of inequity aversion $\bar{\bar{\alpha}}$ such that for $\alpha \geq \bar{\bar{\alpha}}$ integration with flat wage contracts is optimal, whereas for $\alpha < \bar{\bar{\alpha}}$ integration with incentive contracts is optimal.

Proof: i) If $F \leq 2B$ then it is always better to offer incentive contracts in two separated firms than to offer two flat wage contracts within a singe firm. Recall that the gain of providing incentives is given by B per agent, while the cost of setting up a second firm is F. Notice that in both cases the degree of inequity aversion is irrelevant. The best alternative to offering two incentive contracts within a single firm is thus separating the agents in two firms but still offering incentive contracts. Integrating two agents with incentive contracts saves on fixed costs but provokes social comparisons, that is additional agency costs IAC. If the latter exceed the first, $F \leq IAC$, then separation becomes optimal. From Proposition 4.3 we know that the IAC rise with the degree of inequity aversion α; at $\alpha = 0$ we have $IAC = 0$. If $F \leq \lim_{\alpha\to\infty} IAC$ there must thus exist a threshold level $\hat{\hat{\alpha}}$ such that for $\alpha < \hat{\hat{\alpha}}$ we have $F > IAC$, i.e. integration, and for $\alpha \geq \hat{\hat{\alpha}}$ we have $F \leq IAC$, i.e. separation.

ii) If $F > 2B$ then, by the above arguing, the best alternative to offering two incentive contracts within a single firm is offering two flat wage contracts within a single firm. Notice that it is never optimal to offer flat wage contracts and to separate the agents. Even if the IAC become very large there is thus never separation. Absent inequity aversion the profit difference between the two regimes is $2B$. If $\lim_{\alpha\to\infty} IAC \leq 2B$ there will thus always be integration with incentive contracts. If however $\lim_{\alpha\to\infty} IAC > 2B$ then, by the above arguing, there exists a threshold $\bar{\bar{\alpha}}$ such that for $\alpha < \bar{\bar{\alpha}}$ integration with incentive contracts is still optimal but for $\alpha \geq \bar{\bar{\alpha}}$ integration with flat wage contracts becomes optimal. q.e.d.

Figure 4.2 offers an illustration of Proposition 4.8. In all cases, at $\alpha = 0$ the expected profits from two integrated incentive contracts exceeds the expected profit either from offering separated incentive contracts or offering two integrated flat wage contracts. Notice that separated flat wage contracts can never be optimal. The left panel of Figure 4.2 shows expected profit levels in case $F < 2B$. This condition ensures that expected profits from two separated incentive contracts exceed profits from two integrated flat wage contracts. If the additional agency costs due to inequity aversion, IAC, exceed the cost of separation, F, as α goes to infinity, then there exists a threshold level $\hat{\hat{\alpha}}$ such that expected profits from two separated incentive contracts exceed expected profits from two integrated incentive contracts for $\alpha \geq \hat{\hat{\alpha}}$.

The right panel of Figure 4.2 shows expected profit levels in case $F > 2B$. This condition ensures that expected profits from two integrated flat wage contracts exceed profits from two separated incentive contracts. If the IAC exceed the expected profit

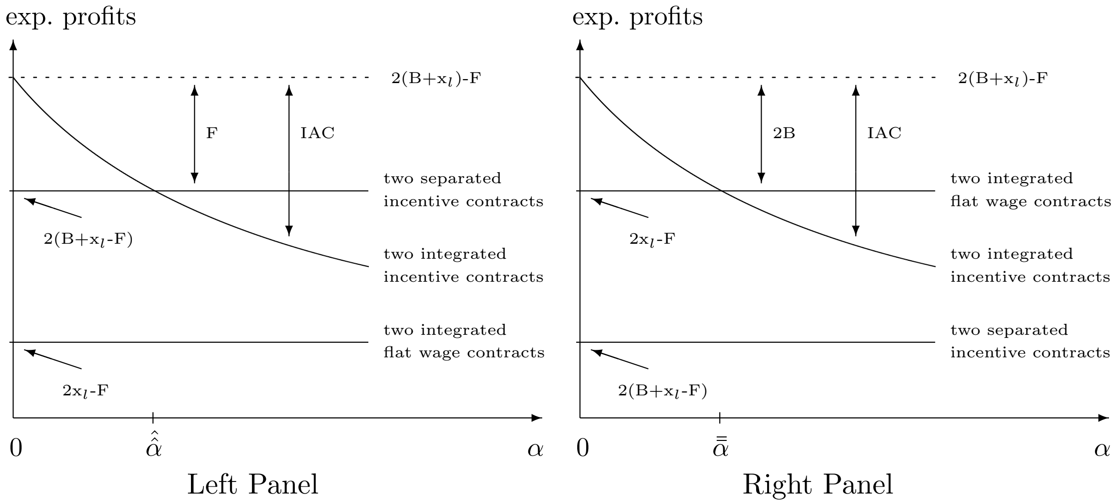

Figure 4.2: **The Optimal Firm Size.**
Left Panel: Expected profit levels with $F < 2B$. In this case expected profits from two separated incentive contracts exceed profits from two integrated flat wage contracts. If the additional agency costs due to inequity aversion, IAC, exceed the cost of separation, F, as α increases, then there exists a threshold level $\hat{\hat{\alpha}}$ such that separated incentive contracts become optimal for $\alpha \geq \hat{\hat{\alpha}}$.
Right Panel: Expected profit levels with $F > 2B$. In this case expected profits from two integrated flat wage contracts exceed profits from two separated incentive contracts. If the IAC exceed the expected profit difference between integrated incentive contracts absent inequity aversion and integrated flat wage contracts, $2B$, as α increases, then there exists a threshold level $\bar{\bar{\alpha}}$ such that integrated flat wage contracts become optimal for $\alpha \geq \bar{\bar{\alpha}}$.

difference between two integrated incentive contracts absent inequity aversion and two integrated flat wage contracts, $2B$, as α goes to infinity, then there exists a threshold level $\bar{\bar{\alpha}}$ such that expected profits with two integrated flat wage contracts exceed expected profits with two integrated incentive contracts for $\alpha \geq \bar{\bar{\alpha}}$.

In this section we have argued that social comparisons can contribute to the old question of the the optimal degree of integration. We do not claim that social comparisons can fully explain the size of the firm nor do we claim that they are the *main* determinant. However, many situation are imaginable where an employer – being indifferent otherwise – wants to separate employees to prevent social comparisons. Even though throughout the paper we did not model heterogeneity in agents' productivity, consider the observation that many firms outsource activities very often (but certainly not exclusively) at the extreme ends of the productivity scale. There are often external consultants that are, in comparison to customary wage levels within the firm, relatively well payed. By the same token, employees of external cleaning companies earn relatively little. Outsourcing of these activities may thus – at least partly – be explained by the intent to maintain a balanced wage structure within the 'core of the firm.'

An employer may not necessarily separate employees into different firms but, in case this sufficiently cuts down social comparisons, into different, say, departments of a firm. In this case our model can contribute to the literature on the internal organization of the firm. Consistent with our arguing is also the observation that within firms (or any other organization) there are often many small rungs in the job ladder, all distinguished by differentiated job titles (junior analyst, senior analyst, junior consultant, senior consultant, etc.). If employees tend to compare only to other employees on same rung of the job ladder and accept that, for example, employees 'above them' may earn more, then 'separating' agents into different 'job categories' may be explained by the employer's intent to cut down social comparisons. The determinants of employees' relevant reference groups – be it a firm, a department, a job category, or some other attribute – will ultimately be an empirical question. However, we claim that co-workers within the same firm are a natural candidate.

4.6 Secrecy of Salaries

The central result of the paper states that inequity aversion among agents increases agency costs. At first sight our results could serve as an explanation for the fact that many labor contracts impose a clause that prohibits employees from communicating their salaries to their colleagues. If – by way of secret salaries – social comparisons can be prevented, the increase in agency costs can be prevented as well. In this section we show that this is not necessarily the case.

Suppose agents can be separated such that the other agent's output realization is not observable. Suppose further that wages do not get communicated because labor contracts prohibit this but that the contracts themselves are common knowledge. We maintain the assumption that the agents' reference group is the respective other agent that is employed with the same principal. (If agents can be separated in a way such that they do not compare themselves any longer the IAC can trivially be avoided.) We now derive the optimal incentive contract for both agents. Even though the agents cannot observe each other's project outcome and wages, they know that their wages differ in certain states of the world because an incentive contract must condition wages on project realizations. In order not to transfer information about the other agent's project outcome, each agent's wage can only depend on his own output realization. Thus, there are two wage levels only. The principal therefore maximizes

$$P_2^s = 2x_l + 2\pi^2[\Delta x - h(u_h)] + 2\pi(1-\pi)[\Delta x - h(u_h) - h(u_l)] - 2(1-\pi)^2 h(u_l) \quad (4.26)$$

with respect to u_h, u_l, and under the incentive and participation constraint

$$\begin{array}{lrl} \text{(IC")} & \pi(1+\alpha\pi)(u_h - u_l) - \psi & \geq 0 \\ \text{(PC")} & \pi u_h + (1-\pi)(u_l - \pi\alpha(u_h - u_l)) - \psi & \geq 0 \end{array}$$

Superscript s stands for 'secrecy contract'. Solving the resulting first-order conditions yields

$$u_h^* = \frac{\psi}{\pi} \qquad \text{and} \qquad u_l^* = \frac{\alpha\psi}{1+\alpha\pi}. \quad (4.27)$$

At $\alpha = 0$ wages and profit equal (twice) the single agent solution. With α increasing the low wage increases in order to reduce inequity, and the principal's expected profit falls. Differentiating P_2^s with respect to α yields

$$\frac{\partial P_2^s}{\partial \alpha} = -\frac{2(1-\pi)\psi(1+\alpha(\pi + r\psi))}{(1+\alpha\pi)^3} \quad (4.28)$$

which is unambiguously negative. We can establish the following proposition.

Proposition 4.9 (Secrecy of Salaries)
Separating the agents such that project outcomes and wages are unobservable amplifies the negative effect of inequity aversion if the agents reference group remains the respective other agent and contracts are common knowledge.

Proof: Comparing (4.28) to (4.20) it can be seen that the 'secrecy profit' falls faster in α than the profit in the two agents case. Subtracting (4.28) from (4.20) yields $(2(1-\pi)\alpha\pi(1+\alpha(\pi+r\psi))(1+\alpha(\pi(\pi(3+\alpha\pi(3+\alpha\pi)))+(2+\alpha\pi(4+\alpha(1+2\pi)))r\psi))/(rk^2(1+\alpha\pi)^3)$ which is always positive. q.e.d.

The 'secrecy contract' is therefore never optimal. Covering up the respective other agent's output realization and wage payments with intent to avoid social comparisons does not mitigate but amplifies the principal's problem. In the 'secrecy contract' wage payments cannot depend on both agent's output realizations since this would reveal the respective other agent's outcome realization and thus wage payment. In states with diverging output realizations wages can therefore not be compressed as it was found optimal in the previous sections. This restriction on the contract design renders the 'secrecy contract' too costly.

4.7 Discussion

4.7.1 Rent Comparison

Suppose now that agents compare rents, that is they explicitly account for effort cost in their inequity term. Notice first that in equilibrium both agents exert effort such that effort terms cancel out in the PC. We must, however, reconsider the IC because an agent now has to account for the difference in effort costs in the inequity term when considering to shirk. The IC can now be written as

$$(\mathrm{IC}_\psi) \qquad \begin{aligned} &\pi^2 u_{hh} + \pi(1-\pi)u_{hl} - \pi^2 u_{lh} - \pi(1-\pi)\alpha(u_{hl} - u_{lh}) \\ &\qquad + \pi\alpha \max\left[u_{hl} - \psi - u_{lh}, 0\right] - \pi(1-\pi)u_{ll} - \psi \quad \geq 0. \end{aligned}$$

Subtracting the l.h.s. of (IC_ψ) from the l.h.s. of (IC), the IC if effort costs are not considered in the inequity term, yields

$$\pi\alpha((2-\pi)(u_{hl} - u_{lh}) - \max\left[u_{hl} - \psi - u_{lh}, 0\right]) \tag{4.29}$$

which is always positive. Hence, considering effort costs in the inequity term can only increase agency costs thereby reinforcing our results. The intuition is straightforward. The expected utility when shirking increases because suffering from being behind is now lower. The difference in utility from wages is reduced by the amount of effort costs, if not cancelled. The incentive to exert effort is thus reduced.

4.7.2 Disutility from Being Better Off

In their original formulation of inequity aversion Fehr and Schmidt (1999) assume that inequity averse individuals dislike both unfavorable and favorable inequity. In this section we discuss the implications if suffering from being better off is incorporated in our model. It now makes a crucial difference whether effort costs enter the comparison or not.

Consider first the case in which agents compare utility from wages only. Instead of the simplified version of inequity aversion assumed in the previous sections, agents' utility function is now given by

$$v_i(w_i, w_j) = u(w_i) - \psi - \alpha \cdot \max[u(w_j) - u(w_i), 0] - \beta \cdot \max[u(w_i) - u(w_j), 0]. \quad (4.30)$$

If $\beta > 0$, agents suffer from receiving a higher wage than the respective other agent. Suppose the principal offers incentive contracts to both agents. Incorporating disutility from being better off into our model has two effects. First, the agents' PCs are tightened. If agents are paid different wages in case their project outcomes differ, an agent now also suffers from inequity whenever he is fortunate whereas the other agent is not. As this happens with positive probability agents have to be compensated. Second, incentive provision is impaired because suffering from being better off clearly reduces the incentive to exert effort. Recall that the results in our model are driven by the observation that the overall impact of inequity aversion on the principal's profit is negative – even when neglecting the utility loss from being better off. Incorporating this disutility adds an unambiguously negative effect and would thus only reinforce our results.

Consider now the case in which effort costs enter the inequity term. Again, the PC is tightened if $\beta > 0$. In equilibrium both agents exert effort and effort costs thus cancel in the inequity term. However, effort cost enter the IC and suffering from being better off may now *facilitate* incentive provision. To see this, assume the most extreme case, which is $\psi > u_{hl} - u_{lh}$. A shirking agent that saves on effort costs is then always better off than the other agent (who works) even if the other agent receives the higher wage in case of diverging output realizations. The IC can then be written as

$$(\text{IC}_\beta) \qquad \begin{aligned} &\pi^2 u_{hh} + \pi(1-\pi)u_{hl} - \pi^2 u_{lh} - \pi(1-\pi)u_{ll} - \psi \\ &\qquad -\pi(1-\pi)\alpha(u_{hl} - u_{lh}) + \beta\left[\psi - \pi(2-\pi)(u_{hl} - u_{lh})\right] \;\geq 0. \end{aligned}$$

The positive effect of β on the IC_β may be very strong. As long as Δx is sufficiently large to ensure $B \geq 0$, ψ can become very large without violating our assumption that incentive contracts are optimal without inequity aversion. Intuitively, if an agent shirks he saves on effort costs and may thus be better off than the other agent who exerts effort. If agents suffer from being better off incentives to exert effort are increased. This effect could, in principle, be so strong that agency costs are *lowered* in comparison to the case without inequity aversion.[50]

4.7.3 Status Seeking

In the previous section we have discussed the possibility that agents suffer from being better off than others. In contrast, suppose now that agents are status seekers, that is they receive *additional* utility from being better off than others. In the context, of this model this translates into $\beta < 0$. Incorporating status seeking into our model has two effects. First, the agents' participation constraints are relaxed. Whenever diverging project outcomes realize the successful agent receives additional utility from being better off than the unsuccessful agent. Second, there is an positive effect on incentives because on top of a high wage an agent receives 'status utility' whenever he is successful whereas the other agent is not. In summary, the unambiguously positive

[50] This effect is analyzed in Bartling and von Siemens (2004).

effect of status seeking on the principal's profit opposes the negative effect of inequity aversion that we have identified in this paper. Since there is no natural lower bound on β agency costs could, in principle, be reduced without bounds. Carrying this effect to the extremes, status seeking would eventually result in contracts in which agents actually pay the principal in order to be employed and sometimes receive 'status utility'. This is only reinforced if effort costs are considered in the inequity terms. However, in this paper we focus on the more natural and more interesting case in which other-regarding preferences provoke a trade-off between the positive effect on the incentive and the negative effect on the participation constraint.

4.8 Conclusion

Recent insights from experimental economics have shown that many people are not fully selfish but have some kind of social preferences. This, in turn, raises the question of how other-regarding behavior interacts with incentive provision. In a moral hazard model with risk averse agents we have shown that inequity aversion among agents unambiguously *increases* agency costs unless agents compare rents *and* suffer from being better off. As a result, optimal contracts for inequity averse agents may be 'low powered', equitable flat wage contracts even when 'high powered' incentive contracts are optimal with selfish agents. Accounting for inequity aversion may thus offer an explanation for the scarcity of incentive contracts many real world situation – in which verifiable performance measures would be available but are not contracted upon.

More specifically, assuming that social comparison are pronounced within firms but less so in the market, we have argued that inequity aversion helps to understand Williamson's (1985) observation that incentives offered to employees within firms are generally low powered as compared to 'high powered' incentive in markets.

Furthermore, we have argued that inequity aversion among agents and the resulting increased agency costs contribute to the old question of the boundary of the firm. In an enriched setting of the basic model, the principal could set up a second firm to separate the agents with intent to avoid social comparisons. If this involves costs, the principal faces the trade-off to either bear increased agency costs or the cost of operating the second firm. The solution to this trade-off defines an optimal size of the firm.

Incorporating our findings into richer models with, for example, heterogeneous agents with respect to the degree of inequity aversion or productivity, or allowing for multi-tasking promises to yield further insights into the determinants of real world wage contracts, the optimal design of institutions, and the boundary of the firm.

4.9 Appendix

Throughout the paper we have assumed an explicit utility function in order to obtain simple closed form solutions. As in Fehr and Schmidt (1999) we also assumed a linear inequity term. In this appendix we show that our results hold true for any concave utility function and irrespective of the functional form the inequity term. To show and illustrate the basic reasoning we, firstly, maintain the assumption that there are only two possible output realizations. Later we will drop this restriction and allow for arbitrary numbers of possible output realizations.

As benchmark, consider the single agent case. With only two possible outcome realizations, wage levels are well-defined by the incentive and participation constraints. Recall that the utility level arising from the wage payment in case of a high output realization is given by u_h, in analogy we defined u_l. From

$$\begin{array}{lrcl}(\tilde{IC}) & \pi u_h + (1-\pi)u_l - \psi & \geq & \pi' u_h + (1-\pi')u_l \\ (\tilde{PC}) & \pi u_h + (1-\pi)u_l - \psi & \geq & 0\end{array}$$

we thus get

$$u_h^* = \frac{(1-\pi')\psi}{\pi - \pi'} \qquad \text{and} \qquad u_l^* = -\frac{\pi'\psi}{\pi - \pi'}. \tag{4.31}$$

If now a second agent is introduced and wages are contingent on the respective other agent's output realization, each agent faces an additional lottery. Suppose an agent's outcome realization is high. If the other agent works, he will also receive a high output realization with probability π, and a low output realization with probability $1-\pi$.

Recall that u_{ij} was defined as an agent's utility from wage w_{ij}, if the agent's output is i and the other agent's output is j. Absent inequity aversion we must have

$$\pi u_{hh} + (1-\pi)u_{hl} = u_h^* \qquad \text{and} \qquad \pi u_{lh} + (1-\pi)u_{ll} = u_l^* \tag{4.32}$$

The inverse function $h = u^{-1}$ specifies the wage payment that is necessary to generate a certain utility level. The principal minimizes wage payments $h(u_{hh})$ and $h(u_{hl})$, and $h(u_{lh})$ and $h(u_{ll})$ such that (4.32) holds. From the first-order condition

$$\pi h'(u_{hh}) + (1-\pi)h'(u_{hl})(-\pi)/(1-\pi) = 0 \tag{4.33}$$

and convexity of $h(\cdot)$ it follows that $u_{hh}^* = u_{hl}^* = u_h^*$ and, equivalently, $u_{lh}^* = u_{ll}^* = u_l^*$. The intuition is straightforward. The second-best utility levels that induce the agent to exert effort are given buy u_h^* and u_l^*. If an agent's wages depend on the other agent's output realization, in expectation he should nevertheless receive u_h^* and u_l^*. Incentives are thus not affected but contingent wages introduce an additional lottery, and agents must be compensated for the associated risk. Absent inequity aversion, wages will thus be independent of the other agent's output realization.

Consider now inequity averse agents. The second-best optimal utility levels in case of high and low output realizations are still given by u_h^* and u_l^*, respectively. However, in case of diverging output realizations there is now a utility loss arising from the inequity. In analogy to (4.32) wage levels must now be such that

$$\pi u_{hh} + (1-\pi)(u_{hl} - \alpha \max[u_{lh} - u_{hl}, 0]) = u_h^*, \qquad \text{and} \tag{4.34}$$

$$\pi(u_{lh} - \alpha \max[u_{hl} - u_{lh}, 0]) + (1-\pi)u_{ll} = u_l^*. \tag{4.35}$$

It can be seen that the cost of providing the second-best optimal utility level are weakly increasing in the level of inequity aversion α. Consider the following reasoning.

1. Fix u_{hl} at some level.

2. Consider the set of (u_{lh}, u_{ll}) such that an agent with a low output realization

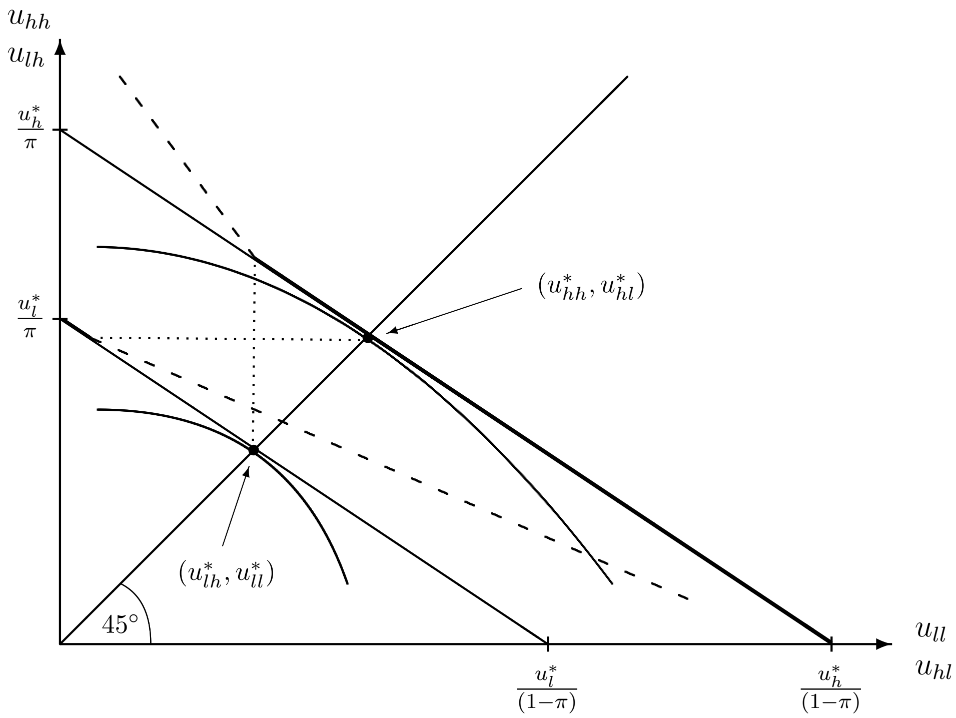

Figure 4.3: **Inequity aversion increases agency costs.**
The negatively sloped, parallel lines depict the constraints subject to which the principal minimizes wages. Without inequity aversion, the lowest iso-cost curves that satisfy the restrictions are tangent where $u_{hh} = u_{hl}$ and $u_{lh} = u_{ll}$. With inequity aversion, the constraints become weakly more restrictive, depicted by the dashed lines. Utility combinations that satisfy the constraints cannot lie on lower iso-cost curves.

receives an expected utility level of u_l^*.

3. Given any u_{lh}, the level of u_{ll} to yield u_l^* is given by

$$u_{ll} = \frac{u_l^* - \pi u_{lh} + \pi\alpha \max[u_{hl} - u_{lh}, 0]}{1 - \pi} \tag{4.36}$$

4. Hence, the cost to implement u_l^* weakly increases in α.

The reasoning for u_h^* is analogous.

Figure 4.3 illustrates the above reasoning and shows how inequity aversion tightens the constraints subject to which the principal minimizes costs. The decreasing, parallel lines depict combinations of u_{hh} and u_{hl}, and u_{lh} and u_{ll} that lead to expected utility levels of u_h^* and u_l^*, respectively. The total differential of (4.32) at constant utility levels yields their slope with $-(1 - \pi)/\pi$. The iso-cost curves in the case without inequity aversion are tangent at $u_{hh}^* = u_{hl}^* = u_h^*$, and $u_{lh}^* = u_{ll}^* = u_h^*$, as argued above. Consider now the case with inequity aversion. Algebraically, the combinations of u_{hh} and u_{hl}, and u_{lh} and u_{ll} that lead to expected utility levels of u_h^* and u_l^* are now given by (4.34) and (4.35), respectively. The total differential of (4.34) while setting $du_h = 0$ yields

$$\frac{du_{hh}}{du_{hl}} = -\frac{(1-\pi)(1+\alpha)}{\pi}, \tag{4.37}$$

if we have $u_{lh} > u_{hl}$. For $u_{lh} \leq u_{hl}$ we get $-(1-\pi)/\pi$. Graphically, with inequity aversion, the dashed line depicting the combinations of u_{hh} and u_{hl} such that the agents receives an expected utility level of u_h^* is *steeper* for $u_{lh} > u_{hl}$ and has the same slope otherwise. Equivalently, the total differential of (4.35) while setting $du_l = 0$ yields

$$\frac{du_{lh}}{du_{ll}} = -\frac{(1-\pi)}{\pi(1+\alpha)}, \tag{4.38}$$

if we have $u_{lh} < u_{hl}$, and $-(1-\pi)/\pi$ otherwise. The dashed line depicting the combinations of u_{lh} and u_{ll} such that the agents receives an expected utility level of u_l^* is *flatter* for $u_{lh} < u_{hl}$ and has the same slope otherwise. Hence, the constraints subject to which the principals minimizes wage payments thus become (weakly) more restrictive.

The above reasoning generalizes straightforwardly to the case where one agent has N and the other agent has M possible output realizations. Now, the solution to the principal's profit maximization problem is not determined by IC and PC alone any longer. The principal first derives the contract that implements each action at the least cost. She then implements the action that maximizes her profit. Incentive and participation constraint for agent i in case the principal wants to implement a_h can now be written as

$$(\hat{IC}) \qquad \sum_{n=1}^{N}\sum_{m=1}^{M} U_i(x_n, x_m) f(x_n, x_m \mid a_h, a_h) - \psi \;\geq\; 0$$

$$(\hat{PC}) \qquad \sum_{n=1}^{N}\sum_{m=1}^{M} U_i(x_n, x_m)[f(x_n, x_m \mid a_h, a_h) - f(x_n, x_m \mid a_l, a_h)] - \psi \;\geq\; 0$$

where $f(\cdot)$ denotes the conditional joint density function over output realizations, and

$U_i(x_n, x_m)$ is given by

$$U_i(x_n, x_m) = u_i(x_n, x_m) - \alpha \max[u_i(x_n, x_m) - u_i(x_n, x_m), 0]. \tag{4.39}$$

$u_i(x_n, x_m)$ denotes the utility that arises from the wage payment in case the own output realization is x_n and the other agent's output realization is x_m. $U_i(x_n, x_m)$ denotes the utility level in this case net of a possible utility loss due to suffering from inequity aversion. Equivalently for agent j. Denote by $\mathcal{U}$ the set of all $U_i(x_n, x_m)$ and $U_j(x_n, x_m)$ such that $(\hat{IC})$ and $(\hat{PC})$ are binding,

$$\mathcal{U} := \{U_i(\cdot), U_i(\cdot) \mid (\hat{IC}) \text{ and } (\hat{PC}) \text{ binding}\}. \tag{4.40}$$

The principal chooses those $U_i(x_n, x_m)$ and $U_j(x_n, x_m)$ from $\mathcal{U}$ that minimize her cost. The wage cost $w(\cdot)$ of providing the respective utility levels is given by

$$w(U_i(x_n, x_m)) = h(u_i(x_n, x_m)) = w_{nm}. \tag{4.41}$$

Recall that $h(\cdot) = u^{-1}$. As can be seen from equation (4.39), for any strictly positive level of α, the utility from wage to attain any fixed level of 'net utility' U must be higher whenever $u_j(x_n, x_m) > u_i(x_n, x_m)$. Since $h'(\cdot) \geq 0$, the wage payment w_{nm} must be higher. Hence, if the principal wants to implement the high effort choice a_h her costs are weakly increased by inequity aversion. If the principal want to implement a_l, she will pay a fixed wage and inequity aversion is thus irrelevant. With additional expenses on notation, this reasoning generalizes to the cases with any finite number of possible effort levels and more than two agents.

Bibliography

AGGARWAL, R. (2000): "Stabilization Activities by Underwriters after Initial Public Offerings," *Journal of Finance*, 55, 1075–1103.

AKERLOF, G. (1970): "The Market for Lemmons: Quality Uncertaintiy and the Market Mechanism," *Quarterly Journal of Economics*, 89, 488–500.

ALLEN, F., AND G. FAULHABER (1989): "Signalling by Underpricing in the IPO Market," *Journal of Financial Economics*, 23, 303–323.

BARTLING, B., AND F. VON SIEMENS (2004): "Efficiency in Team Production with Inequity Averse Agents," Discussion Paper, University of Munich.

BENVENISTE, L., W. BUSABA, AND W. WILHELM JR. (1996): "Price Stabilization as a Bonding Mechanism in Equity Issues," *Journal of Financial Economics*, 42, 223–255.

BENVENISTE, L., AND P. A. SPINDT (1989): "How Investment Bankers Determine the Offer Price and Allocation of Initial Public Offerings," *Journal of Financial Economics*, 24, 343–362.

BENZION, U., A. RAPOPORT, AND J. YAGIL (1989): "Discount Rates Inferred from Decisions: An Experimental Study," *Management Science*, 35, 270–84.

BEWLEY, T. F. (1999): *Why Wages Don't Fall During a Recession.* Harvard University Press, Cambridge, MA.

BOEHMER, E., AND R. FISHE (2001): "Who ends up short from Underwriter Short Covering – A detailed Analysis of IPO Price Stabilization," working paper, NYSE.

Bolton, G. E., and A. Ockenfels (2000): "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, 90, 166–193.

Booth, J. R., and R. L. Smith (1986): "Capital Raising, Underwriting and the Certification Hypothesis," *Journal of Financial Economics*, 15, 261–281.

Busaba, W. Y., L. M. Benveniste, and R.-J. Guo (2001): "The Option to Withdraw IPOs during the Premarket: Empirical Analysis," *Journal of Financial Economics*, 60, 73–102.

Camerer, C. F. (2003): *Behavioral Game Theory*. Princeton University Press, Princeton and Oxford.

Camerer, C. F., G. Loewenstein, and M. Rabin (eds.) (2003): *Advances in Behavioral Economics*. Princeton University Press, Princeton and Oxford.

CESR (2002): "Stabilisation and Allotment, a European Supervisory Approach," Consultative paper, cesr 02/020b, The Committee of European Securities Regulators.

Chen, H.-C., and J. Ritter (2000): "The seven percent solution," *Journal of Finance*, 55, 1105–1131.

Cho, I.-K., and D. Kreps (1987): "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102, 179–222.

Chowdhry, B., and V. Nanda (1996): "Stabilization, Syndication, and Pricing of IPOs," *Journal of Financial and Quantitative Analysis*, 31, 25–42.

Coase, R. (1937): "The Nature of the Firm," *Economica*, 4, 386–405.

DellaVigna, S., and U. Malmendier (2003): "Overestimating Self-Control: Evidence from the Health Club Industry," mimeo, Stanford University.

——— (2004): "Contract Design and Self-Control: Theory and Evidence," *Quarterly Journal of Economics*, forthcoming.

DEMOUGIN, D., AND C. FLUET (2003): “Group vs. Individual Performance Pay When Workers Are Envious,” CIRPEÉ Working Paper 03-18.

DUNBAR, C. G. (2000): “Factors Affecting Investment Bank Initial Public Offering Market Share,” *Journal of Financial Economics*, 55, 3–41.

DUR, R., AND A. GLASER (2004): “Optimal Incentive Contracts When Workers Envy Their Boss,” Tinbergen Institute Discussion Paper 2004-046/1.

ELLIS, K., R. MICHAELY, AND M. O’HARA (2000): “When the Underwriter Is the Market Maker: An Examination of Trading in the IPO Aftermarket,” *Journal of Finance*, 55, 1039–1074.

ENGLMAIER, F., AND A. WAMBACH (2003): “Contracts and Inequity Aversion,” CESifo Working Paper No. 809.

FEHR, E., AND S. GÄCHTER (2002): “Do Incentive Contracts Crowd Out Voluntary Cooperation?,” Institute for Empirical Research in Economics Working Paper No. 34, University of Zurich.

FEHR, E., AND K. SCHMIDT (1999): “A Theory of Fairness, Competition, and Cooperation,” *Quarterly Journal of Economics*, 114, 817–68.

——— (2003): “Theories of Fairness and Reciprocity - Evidence and Economic Applications,” in *Advances in Economics and Econometrics - 8th World Congress*, ed. by M. Dewatripont, L. Hansen, and S. Turnovsky, pp. 208–57. Cambridge University Press.

FRANCIS, B. B., AND I. HASAN (2001): “The Underpricing of Venture and Nonventure Capital IPOs: An Empirical Investigation,” *Journal of Financial Services Research*, 19, 99–113.

FREDERICK, S., G. LOEWENSTEIN, AND T. O’DONOGHUE (2002): “Time Discounting and Time Preference: A Critical Review,” *Journal of Economic Literature*, 40, 351–401.

FREY, B. S., AND R. JEGEN (2001): "Motivation Crowding Theory," *Journal of Economic Surveys*, 15, 589–611.

FUDENBERG, D., AND J. TIROLE (1991): *Game Theory*. MIT-Press, Cambridge, Massachusetts.

GNEEZY, U., AND A. RUSTICHINI (2000a): "A Fine is a Price," *Journal of Legal Studies*, 29, 1–17.

——— (2000b): "Pay Enough or Don't Pay at All," *Quarterly Journal of Economics*, 115, 791–810.

GRINBLATT, M., AND C. Y. HWANG (1989): "Signalling and the Pricing of New Issues," *Journal of Finance*, 44, 393–420.

GROSSMAN, S., AND O. HART (1983): "An Analysis of the Principal-Agent Problem," *Econometrica*, 51, 7–45.

——— (1986): "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94, 691–719.

GUL, F., AND W. PESENDORFER (2001): "Temptation and Self-Control," *Econometrica*, 69, 1403–35.

HANSEN, R. S. (2001): "Do Investment Banks compete in IPOs?: The Advent of the 7% Plus Contract," *Journal of Financial Economics*, 59, 313–346.

HART, O., AND J. MOORE (1990): "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 98, 1119–1158.

HELWEGE, J., AND N. LIANG (1996): "Is There a Pecking Order? Evidence from a Panel of IPO Firms," *Journal of Financial Economics*, 40, 429–458.

HOLMSTRÖM, B. (1979): "Moral Hazard and Observability," *Bell Journal of Economics*, 10, 74–91.

——— (1982): "Moral Hazard in Teams," *Bell Journal of Economics*, 13, 324–40.

Holmström, B., and P. Milgrom (1987): "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55, 303–28.

——— (1991): "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization*, 7, 24–52.

Itoh, H. (2003): "Moral Hazard and Other-Regarding Preferences," Discussion Paper, Hitotsubashi University.

Jenkinson, T., and A. Ljungqvist (2001): *Going Public: The Theory and Evidence on How Companies Raise Equity Finance*. Oxford University Press, (2nd edition).

Kirby, K. N. (1997): "Bidding on the Future: Evidence Against Normative Discounting of Delayed Rewards," *Journal of Experimental Psychology*, 126, 54–70.

Kirby, K. N., and R. J. Herrnstein (1995): "Preference Reversals Due to Myopic Discounting of Delayed Reward," *Psychological Science*, 6, 83–89.

Laibson, D. (1997): "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112, 443–77.

Madrian, B. C., and D. F. Shea (2001): "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior," *Quarterly Journal of Economics*, 116, 1149 – 1187.

Megginson, W. L., and K. A. Weiss (1991): "Venture Capitalist Certification in Initial Public Offerings," *Journal of Finance*, 46, 879–903.

Mookherjee, D. (1984): "Optimal Incentive Schemes with Many Agents," *Review of Economic Studies*, 51, 433–446.

Nanda, V., and Y. Yun (1997): "Reputation and Financial Intermediation: An Empirical Investigation of the Impact of IPO Mispricing on Underwriter Market Value," *Journal of Financial Intermediation*, 6, 39–63.

Neilson, W. S., and J. Stowe (2003): "Incentive Pay for Other-Regarding Workers," Discussion Paper, Texas A&M University.

O'Donoghue, T., and M. Rabin (1999a): "Doing It Now or Later," *American Economic Review*, 89, 103–124.

——— (1999b): "Incentives for Procrastinators," *Quarterly Journal of Economics*, 114, 769–816.

——— (1999c): "Procrastinaton in Preparing for Retirement," in *Behavioral Dimensions of Retirement Economics*, ed. by H. Aaron, pp. 125–156. Brookings Institution Press and Russell Sage Foundation.

——— (2001): "Choice and Procrastination," *Quarterly Journal of Economics*, 116, 121–160.

Phelps, E. S., and R. A. Pollack (1968): "On Second-Best National Saving and Game-Equilibrium Growth," *Review of Economic Studies*, 35, 185–99.

Pollack, R. A. (1968): "Consistent Planning," *Review of Economic Studies*, 35, 201–08.

Rabin, M. (2002): "A Perspective on Psychology and Economics," *European Economic Review*, 46, 657–685.

Rey Biel, P. (2003): "Inequity Aversion and Team Incentives," Discussion Paper, University College London.

Ritter, J., and I. Welch (2002): "A Review of IPO Activity, Pricing, and Allocations," *Journal of Finance*, 57, 1795–1828.

Rock, K. (1986): "Why New Issues Are Underpriced," *Journal of Financial Economics*, 15, 187–212.

SEC (1997): "Regulation M," Release no. 34-38067, Securities and Exchange Commission (SEC).

SPENCE, M. (1973): "Job Market Signaling," *Quarterly Journal of Economics*, 87, 355–74.

STROTZ, R. H. (1956): "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies*, 23, 165–80.

TITMUSS, R. M. (1970): *The Gift Relationship*. Allen and Unwin, London.

WELCH, I. (1989): "Seasoned Offerings, Imitation Costs, and the Underpricing of Initial Public Offerings," *Journal of Finance*, 44, 421–449.

WILLIAMSON, O. (1975): *The Economic Institutions of Capitalism: Firms, Markets, and Relational Contracting*. Free Press, New York.

——— (1985): *The Economic Institutions of Capitalism*. Free Press, New York.

Eidesstattliche Versicherung

Ich versichere hiermit eidesstattlich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie mir gegebene Anregungen sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

München, den 5. März 2004

Lebenslauf

22. Mai 1974	Geboren in Herford
1993	Abitur am Friedrichs-Gymnasium Herford
1993/94	Zivildienst in Krankenhäusern auf Norderney und in München
1994-1997	Studium der Volkswirtschaftslehre an der Ludwig-Maximilians-Universität München
1997/98	Master's Course in Economics an der London School of Economics and Political Science
1998-2000	Wissenschaftlicher Mitarbeiter am Seminar für Wirtschaftstheorie der LMU München
2000/01	Marie Curie Fellow an der Faculty of Economics and Politics der Cambridge University
2001-2004	Wissenschaftlicher Mitarbeiter am Seminar für Wirtschaftstheorie der LMU München
21. Juli 2004	Promotion zum Dr. oec. publ.
2004/05	Post-Doctoral Fellow am Department of Economics der Harvard University