Essays in Microeconomics

Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften der Universität Mannheim

Christoph Wolf

Frühjahrssemester 2017

Abteilungssprecher: Prof. Dr. Jochen Streb
Referent: Prof. Volker Nocke, Ph.D.
Korreferent: Prof. Konrad Stahl, Ph.D.

Tag der mündlichen Prüfung: 21. Juni 2017

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbständig angefertig habe und die benutzten Hilfsmittel vollständig und deutlich angegeben habe.

Mannheim, 02. Mai 2017 Christoph Wolf

Acknowledgements

First of all, I would like to thank my supervisor Volker Nocke for his continuous support during my doctoral studies in Mannheim. I am very grateful for the productive discussions about my research and the highly valuable advise on the projects in this thesis. The interaction with him has shaped the way I am thinking about economic problems to a large extent.

I am indebted to Konrad Stahl for his support over the past years. Without all his time, effort, advise and the many discussions about research, I would not be at the stage at which I am now. I developed the interest in economic research during my undergraduate time in Mannheim being a research assistant for him. I am very grateful that he convinced me to stay in Mannheim.

I am also very thankful that I had the chance to present my work to the Mannheim Micro Group multiple times and to discuss it with them in detail. I benefited a lot from the constructive comments and suggestions. A special thanks goes to Malin Arve, Takakazu Honryo, Raphael Levy, Andras Niedermayer, Martin Peitz, Nicolas Schutz, Michelle Sovinsky, Hidenori Takahashi, Emanuele Tarantino and Thomas Tröger. I especially thank Renato Gomes for his very detailed and extremely helpful comments. Moreover, I would like to thank my coauthors Stefan Weiergräber, Sinem Hidir and André Stenzel for their effort in the projects and their patience with me.

I would like to thank many of the doctoral students in Mannheim. In particular, I would like to mention Florian Exler for making the last year more enjoyable (#wwdf1jm), Niccolo Lomys, Johannes Schneider and André Stenzel for being great office mates to discuss not only research with and the Theory Early Lunch Group.

Moreover, I thank the whole University of Mannheim Handball team. I have found many friends in this group over the past 9 years and enjoyed the distraction in stressful periods. I want to (and have to) mention Matthias Carius. Thanks for being a good friend, gym buddy, and handball coach: Mens sana in corpore sano.

One special thank you goes to David Lindequist, a close friend and coauthor. Thank you for bringing FIFA to my life and for giving very thoughtful advise while losing against me. I always enjoy talking to you about lending standards, macroeconomics, Bayes' rule and life in general.

Finally, I wish to thank my parents, Marion and Jürgen, and my sister, Tanja, for everything. It goes without saying that without you I would not be where I am now. Thank you.

Contents

1 General Introduction

A central question in many economic problems is how information about the environment affects the interaction of decision makers. Typically, information is not symmetrically distributed across economic agents for several reasons. For example, one agent may be the owner of a product, while the other is willing to buy the product. Naturally, the owner has access to more information about the product than the potential buyer. It is well-known that asymmetric information can have drastic consequences such as leading to the breakdown of markets as famously discussed in Akerlof (1970). Another reason for asymmetric information is learning. In dynamic environments, players can acquire additional information on projects or objects that is valuable to them. However, learning is not necessarily symmetric because the economic agents may have different tasks or roles in the interaction and may not observe the same information.

This thesis contains three self-contained articles, each studying the role of information on the economic interaction in a specific environment. Chapter 3 is joint work with Stefan Weiergräber, Chapter 4 is joint work with Sinem Hidir. Chapters 2 and 4 are theoretical papers that study dynamic environments in which players can learn about the fundamental variables that determine the value of the interaction. Chapter 2 analyzes the effect of learning about a project's profitability in a principal-agent relationship with an application to the financing of projects with uncertain quality, e.g. venture capital finance. Chapter 4 considers a dynamic team work problem in which the ability of a team member is his private information but learned over time. Chapter 3 is empirical work that addresses asymmetries in the bidders' information about the future profitability of tracks in German short haul railway passenger service procurement auctions.

Chapter 2. In this chapter, the underlying question is how learning about a project's feasibility affects the interaction of a principal, who funds the project but cannot directly observe the agent's actions and therefore what the agent learns, with an agent, who works on the project. Towards this, I study a continuous-time moral hazard problem with private learning about a project of unknown quality. There is ex ante symmetric information and full commitment for the principal. The project generates a profit if two consecutive stages are completed. The amount of experimentation required to complete the first stage (milestone) is informative but not conclusive about the quality of the project. The informativeness of the milestone yields an incentive to privately shirk in the first stage. This increases the principal's pessimism in the second stage and thereby induces more favorable second-stage contract terms for the agent.

In the optimal contract, the reward for a first-stage success is decreasing in its arrival time to prevent effort delays. The reward's composition changes with the success time: early successes are rewarded with long second-stage deadlines and no bonus payments in the first stage, while later successes are rewarded with first-stage bonus payments and less continuation value from second-stage experimentation. Allowing for agent replacement between stages, I show that the principal wants to replace the agent in the second stage if the success arrives late.

Chapter 3. This chapter, which is joint work with Stefan Weiergräber, is, in contrast to chapters 2 and 4, an empirical project, in which we take a theoretical model to a data set on German short haul railway passenger service auctions to estimate the effects of asymmetric access to information of an incumbent and its competitors. Many procurement auctions involve both private value and common value elements. Bidding firms are often asymmetric in both dimensions. First, former state monopolists or incumbents may be better informed about the common value, for example the revenue component of a procurement contract. Second, incumbents and entrants may have very different cost distributions (a typical private value component). Understanding the bidding behavior in a setting with private and common value components and asymmetries among bidders in both dimensions is essential to evaluate auction outcomes. We develop and estimate a structural auction model using a detailed contract-level data set of the market for short-haul railway passenger services in Germany. This allows us to disentangle the effects of asymmetries in the cost distribution between the incumbent and the entrants from the effects of asymmetric information about the revenue component of a SRPS contract. Data on *gross auctions*, in which firms do not face revenue risk, allow us to back out the cost distribution for each firm. Data on *net auctions*, in which firms bear the revenue risk, enable us to quantify the effect of revenue uncertainty on bidding strategies. Our results indicate that (1) bidding behavior is indeed systematically different in gross and net auctions, (2) the incumbent is only slightly more cost-efficient on most lines, (3) the incumbent has substantially more information about future ticket revenues than an entrant. We use our parameter estimates to run a series of counterfactuals. If net auctions were procured as gross auctions, we find that (1) entrants would have bid much more aggressively than in the status quo, (2) on average, the probability of selecting the efficient firm would increase from 64% to 75%.

Chapter 4. Chapter 4 is joint work with Sinem Hidir and studies the role of learning about a team member's productivity and dynamic freeriding incentives. We analyze a model of dynamic collaboration in the presence of asymmetric information about a player's ability. There is a team of two working to achieve a one time success on a project, and only the ability of one player is common knowledge (senior) while the ability of the other player (junior) is private information. This leads to gradual pessimism of the senior about the junior being of high ability as time passes without a success. The senior increases his effort over time in order to compensate for the junior's (in expectation) lower ability. This is anticipated by the junior and therefore induces him to reduce his effort early on. We show numerically that overall effort can increase when instead of adding a productive junior to the team with certainty, the junior's ability is random and he is unproductive with positive probability. This uncertainty reduces the freeriding incentive of the senior.

2 Informative Milestones in Experimentation

2.1 Introduction

Intermediate milestones are frequently observed in principal-agent relationships where the feasibility of the project is unknown to both parties. A fundamental reason for this is that the performance in early stages conveys valuable information about subsequent stages to the principal. Given this informational spillover and the agent's ability to affect the observable information, several questions arise. How are the agent's incentives to exert effort affected by the informativeness of the milestone? How does the optimal contract adapt to these incentives? How does the role of bonus payments, deadlines and continuation contracts change with an informative milestone?

I show that the informational spillover across the stages introduces an endogenous ratchet effect: By privately shirking in the first stage the agent increases the principal's pessimism in the second stage which yields higher rents for the agent through a more favorable second-stage contract. To prevent the delay in effort, the optimal contract rewards early successes with higher rents. These rents are increasing in the continuation value from second-stage experimentation because a higher continuation value amplifies the ratchet effect. In contrast to a setting with independent stages, rewarding the agent with a continuation contract is therefore costly. Thus, the optimal contract will also use bonus payments for first-stage successes to reduce the information rents. Moreover, if the principal has access to a new agent for the second stage, the informative milestone gives rise to replacement of the agent if he succeeds too late in the first stage.

As an application of the setup, consider the venture capital industry. Innovative projects involve uncertainty about their quality. They also require the effort and knowledge of experts as well as substantial amounts of capital. As entrepreneurs rarely have the necessary funds themselves, they contract with financial investors. One of the main funding sources for high-risk startups is venture capital.[1] The high degree of uncertainty about the project's future profitability together with the substantial size of the investment may make investors reluctant to invest. To overcome this problem, intermediate stages, so-called *milestones*, are introduced to gather information about the project's quality at a reduced cost. A typical example for such a milestone is the development of a prototype. One important feature of prototypes is that they are a tool to learn about the prospects of the project.

[1] Gornall and Strebulaev (2015) show that 42% of independent U.S. public firms founded after 1974 are venture-capital backed. Notably, 85% of the R&D expenditures of these firms stem from venture-capital backed firms.

Staging of venture capital contracts is a well-documented feature: Kaplan and Stroemberg (2004) show that 72.8% of contracts in their sample involve staging. It has been argued that staging helps in mitigating agency costs (see for example, Gompers (1995), Neher (1999) or Cumming (2012)). However, the literature is surprisingly silent about learning and the informational value of milestones. I show that introducing informational content of early stages has important consequences for the incentives of the startup as well as the design of the optimal contract.

I develop and study a continuous-time principal-agent bandit experimentation model with an informative milestone. A project of uncertain quality has to complete each of two sequential stages to realize its benefits. Any success is immediately and publicly observed. The intensity rate of obtaining a breakthrough for a given level of effort in the first stage stage is higher for a good than for a bad project. Hence, the total effort required until the first success is informative but not conclusive about the project's quality.[2] As effort is costly and unobserved, this is a dynamic moral hazard problem with private learning. I solve for the full-commitment profit-maximizing contract that conditions on the publicly observable success times. I allow for arbitrary payment rules subject to limited liability.

The optimal contract will feature a deadline for each of the stages because if the principal becomes too pessimistic she will terminate the project. Moreover, it is without loss of generality to focus on bonus contracts that have payments to the agent only at success times. Hence, there are at most two payments to the agent. First-stage bonus payments are short-term incentives that do not condition on the long-run success of the project and their value is independent of the current belief. However, the second-stage bonus payment and deadline together induce an expected value for the agent that depends on the belief about the project's quality at the beginning of the second stage which is determined by the first-stage performance. This continuation value can be interpreted as the value of equity given to the agent after the first success.

The informational spillover across the stages implies that effort choices in the first stage do not only affect the belief about the project in the current stage but also the initial belief of the second stage. If more effort is exerted until the first success is obtained, players are more pessimistic in the second stage. Deviations from the expected effort path persistently divert the agent's private belief from the principal's belief. Therefore, he holds a different belief *after* the first success than the principal. This is the key novelty of this paper and the underlying reason for the main results. The effect of a deviation from the expected effort path on the belief in the current stage gives rise to *procrastination rents*, while the effect of a deviation on the belief in the following stage gives rise to a novel rent that I call *informativeness rents*.

First, consider the interaction of moral hazard and private learning in the current stage. The agent's incentives are driven by his private belief about the success probability: the reward has to be chosen such that the agent is at least compensated with an *expected* utility that outweighs the cost of effort. However, the agent has the ability to privately shirk and divert the beliefs

[2] Hence, the stochastic process of the breakthrough in the first stage is the same as, for example, in Keller and Rady (2010), while the stochastic process for a breakthrough in the second stage is the same as, for example, in Keller et al. (2005).

because the principal cannot distinguish whether the absence of a success was due to bad luck or due to a deviation. Therefore, the principal becomes overly pessimistic and the reward is misspecified: if the principal is overly pessimistic she believes that she has to pay a higher reward to the agent. Hence, the agent has an incentive to delay effort if the contract does not account for private learning about the current stage. This effect is present in both stages and the agent has to be granted *procrastination rents* to prevent belief manipulation about the current stage.

Second, consider the interaction of moral hazard and private learning in the first stage about the second stage. By privately shirking in the first stage the agent induces the principal to be overly pessimistic in the second stage. A low second-stage belief of the principal implies that the continuation contract has to promise the agent a high bonus if he succeeds in the second stage: the less likely she thinks it is to obtain a success, the higher this bonus payment must be upon obtaining the second success to make the agent willing to exert effort. Hence, conditional on reaching the second stage, the agent wants the principal to be pessimistic in the second stage to enjoy higher payments upon second-stage success.

The intuition for the agent's incentive to manipulate the principal's belief can be related to the *ratchet effect*:[3] the agent wants the principal to be sufficiently optimistic to continue the project; however, conditional on continuation he wants the principal to be pessimistic to be granted a high bonus after the second success.

This effect is neither present with independent stages nor in a one-stage setting. With independent stages, the belief at the beginning of the second stage is exogenously given and the agent cannot affect this by off-path effort choices in the first stage. In a one-stage setting, the interaction ends after the first success. To prevent the delay in effort, the principal has to reward early successes with higher rents than later successes. I call these rents *informativeness rents* as they only arise due to the informativeness of the milestone. The rate at which the total reward for the first success decreases is exactly the gain in value of the private information at the beginning of the second stage; that is, the value of holding a marginally more optimistic belief than the principal in the second stage. While the procrastination rents prevent deviations that directly affect the level of the reward for a success in the current stage, the informativeness rents prevent deviations that alter the assessment of the second-stage contract through the persistence of the agent's private information into the second stage.

Procrastination rents are unaffected by the way in which the reward is delivered to the agent. However, the informativeness rent has to be provided because the agent can gain from the persistence in his private information in the continuation contract. I show that the informativeness rent is increasing in the on-path value of the continuation contract. The higher is the value to be delivered in the second stage, the higher is the value of being more optimistic in the second stage. Therefore, the principal faces a tradeoff when choosing how to deliver the first-stage reward. The principal would like to incentivize the agent to work until an extended deadline in the second stage: due to the procrastination rents, experimentation stops inefficiently early in the second stage. By extending the deadline, additional overall surplus is generated which makes continuation contracts an attractive reward mechanism.

[3] See, for example, Laffont and Tirole (1988) or, more recently, Bhaskar (2014).

The downside of delivering utility through more valuable continuation contracts is that these increase the informativeness rents. In particular, at all times prior to a particular success time, the agent's incentive to delay effort increases if additional utility is delivered through a continuation contract at that success time because it increases the value of the private information. Hence, the cost of using the long-term reward continuation contract is increasing in the success time.

In the optimal contract, early successes are rewarded with continuation contracts only, implemented through long second-stage deadlines, because the gain of extending the deadline is large while the cost is very low. As time elapses without a breakthrough, the composition changes such that the reward consists of an increasing share of bonus payments and less of a continuation contract to reduce the informativeness rents for all earlier successes.

It is worthwile to note that I can use backward induction in this setting with full commitment. The reason is that reducing deviation incentives in the first stage through the choice of and maximizing profits with a second-stage contract are aligned. Delivering utility with extended deadlines in the second stage reduces the deviation incentives in the first stage compared to the alternative of using higher second-stage bonus payments. Higher bonus payments in the second stage are indeed the reason why the agent is willing to delay effort in the first stage. Moreover, extending deadlines increases the probability of succeeding in the second stage and therefore maximizes the principal's profits subject to the promised utility of the agent.

In an extension, I consider the possibility of replacing the agent after the first stage. In contrast to other models of experimentation or staged financing, my model can rationalize managerial turnover in young startup firms.[4] Hannan et al. (1996) show that 40% of CEOs are replaced within the first 40 months of a startup. I show that the presence of the informativeness rent gives rise to turnover: The principal always wants to introduce two deadlines for the first stage in the present setting: (i) if the agent succeeds before the first deadline, he is rewarded with a continuation contract, (ii) if the agent succeeds after the first and before the second deadline, he receives a payment and is replaced by a new agent in the second stage, (iii) if the agent has not obtained a success before the second deadline, the project is terminated. To see why replacement is optimal, recall that the agent receives the informativeness rent only if he is working on the second stage. Hence, if the agent is replaced when he succeeds after a certain deadline he can be incentivized at a lower cost in the first stage. This also implies that the agent receives lower rents in the continuation region because delaying effort becomes less attractive as the replacement deadline approaches. Thus, the principal faces a tradeoff between the cost of more expensively rewarding agents in the replacement region with a bonus payment instead of a continuation contract and the benefit of reduced informativeness rents. I show that the principal always prefers to have both, a replacement and a continuation region, in the optimal contract if there is an agency conflict in the first stage.

I extend the analysis to allow for the principal's choice of informativeness and endogenous staging. Assuming that the principal can choose the intensity rate of the first stage and that

[4] Garrett and Pavan (2012) provide a dynamic model of managerial turnover. However, they assume ex ante asymmetric information and that the productivity of the agent is changing.

the second stage vanishes as the first stage becomes fully informative, I show that the optimal two-stage contract converges to the optimal one-stage contract. The principal faces a non-trivial tradeoff in the staging decision: when introducing an informative milestone, the principal gains from the additional information provided in the first stage and can condition the second-stage funding and contract on the first stage outcome. However, introducing an informative milestone generates the informativeness rents. Numerically, I show that the staging decision depends on parameter values and it cannot be argued that one mode dominates the other generically. However, I find that staging takes place more likely if the initial success probability is low; that is, when the value of additional information is relatively high. In particular, I find that introducing an informative milestone can facilitate funding for projects with low initial success probabilities that would not receive funding as a one-stage project.

The insights I derive can be applied in several other contexts. For example, it could describe the interaction of a CEO with the leader of the research department to work on a risky and expensive project. Alternatively, there could be uncertainty about the worker's type instead of the project's quality. In this case, an employer could offer a contract with a probationary period in which a first signal about the employee's competence can be obtained. The result on the composition of an agent's reward conditional on the performance may also give another perspective on regulating CEO compensation.

My analysis generates several empirical predictions that can be of interest in different applications: (i) The composition of the agent's compensation changes with performance. In particular, if performance gets worse, the total reward is lower and consists of relatively more short-term than long-term rewards. For example, a well-performing CEO is rewarded with stock options that are tied to future performance. A CEO that performs worse is rewarded with bonus payments and less with stock options. The total worth of the reward is higher for the well-performing CEO. (ii) Deadlines are relatively more responsive to early performance while final-stage bonus payments are less responsive to early performance compared to a setting without informativeness rents. (iii) Early-stage deadlines are relatively short if there is a learning spillover to future stages. (iv) Even successful agents may be replaced if they do not perform sufficiently well although the project is continued and the agent known to be able to complete future tasks. (v) Staging occurs more frequently if the initial risk is high.

Short early-stage deadlines may explain the observation of high failure rates of startups. Shikhar Gosh is cited in The Wall Street Journal that 35% of startups survive to the age of 10 years. High failure rates are not necessarily due to high risk only. If deadlines are used by the investors and agency conflicts induce deadlines to be inefficiently short, too many startups fail and too few innovations are obtained in a society. Concerning the replacement of successful agents, Noam Wasserman notes in a Harvard Business Review article that: *"[o]thers invest in a start-up only when they're confident the founder has the skills to lead it in the long term. Even these firms, though, have to replace as many as a quarter of the founder-CEOs in the companies they fund."* Hence, replacement also occurs even though there is no doubt about the agent's qualification to succeed with the project. Although empirical implication (v) about staging and initial risk is only a numerical outcome, it is fairly intuitive and empirical evidence has been found in Bienz and Hirsch (2011).

Related Literature. My paper contributes to the growing literature on principal-agent models with ex ante symmetric uncertainty about a project's feasibility. Most of the early work focuses on the case where one success suffices to complete the project and in the absence of a success players become pessimistic about the project's quality; see for example, Bergemann and Hege (1998; 2006), Halac et al. (2016), Hörner and Samuelson (2013). These models apply the exponential bandit model of Keller et al. (2005) in which one success is fully informative about the project's quality. By contrast, I assume that there may be a first stage that is informative but not conclusive about the quality of the project and a second breakthrough is required to complete the project. First, it allows to assess the impact of staging that is widely used in contracting relationships when there is uncertainty about the project's value. I show that introducing an additional and informative stage can facilitate funding of projects that would not be undertaken if they were forced to involve only a single stage. Seond, it allows for more flexibility in the learning process compared to the one-stage experimentation literature; during the course of the project it may be that players become more optimistic instead of increasingly pessimistic over time.

To the best of my knowledge, there are three papers that consider staged projects which are closely related. First, Moroni (2016) considers a principal contracting with several agents on a project that requires multiple breakthroughs to yield the final payoff. She shows that agents have an additional free-riding incentive because another agent may start a subsequent stage. Because she assumes that early stages carry no information about later stages, staging has no informational value and therefore, there is no informativeness rent present. If there is only one agent, her analysis may serve as a benchmark to the present paper without learning across stages and with a fixed second-stage belief. In that case the agent would be incentivized with a constant continuation value in the first and a constant bonus payment in the second stage. Hence, the continuation contract was independent of the first-stage performance and there would be no replacement of the agent.

Second, Green and Taylor (2016) and Hu (2014) study dynamic moral hazard problems in which the agent also has to obtain two success. However, the quality of the project is known to be good but the agent has the ability to divert the flow funding of the principal. In their case deadlines arise to prevent the agent from diverting cash. Early successes also have to be rewarded with higher continuation values. However, the reason is fundamentally different: In both, Green and Taylor (2016), Hu (2014), the agent has a direct benefit from delaying effort which is the flow benefit from diverting the cash. In my paper, delaying effort creates an informational advantage because it persistently drives a wedge between the principal's and the agent's belief about the project's value. If the project was known to be good, the principal could achieve the first best in my model. While Green and Taylor (2016) focus on the role of communication and private observability of progress, they do not consider general payment schemes and restrict attention to contracts that only pay the agent after a final success.

This paper also relates to the ongoing discussion on staged contracts. Examples of this literature are, among others, Bolton and Scharfstein (1990), Neher (1999), Cuny and Talmor (2005) and Booth et al. (2004). These papers discuss the value of staging contracts to mitigate agency conflicts through the threat of termination and to reduce the hold-up problem. However, these papers do not consider the possibly uncertain feasibility of the project about which the agent can privately learn by exercising effort. While this may be realistic in some cases, learning plays an important role in the financing of innovation. Therefore, I take a different perspective and study the informational value of staging if the project's feasibility is unknown. Pindyck (1993) also discusses the informational value of early investments that can reduce uncertainty over later costs. I show that if learning is private, then there is a tradeoff of introducing informative milestones: On the one hand, an informative milestone can be beneficial because a signal can be generated at a lower cost. On the other hand, private learning of the agent gives rise to an additional agency rent due to the possibility to manipulate the principal's belief. To my knowledge, this is the first paper to address this potential drawback of informative milestones.

Bhaskar (2014) considers a related two-period model with learning about a project's difficulty but without commitment. In the first period, a signal is generated that depends on the agent's effort and the project's type. Similar to the present paper, he shows that the agent has an incentive to manipulate the principal's belief such that he obtains higher payoffs in the second period. Different to his paper, I study the interplay between learning and dynamic moral hazard. Thereby, I can shed light on the use of different reward instruments, deadlines and bonus payments, to incentivize the agent.

On a more abstract level, this paper is related to DeMarzo and Sannikov (2016) who also study a model where agent's private deviations affect the assessment of the promised continuation value. In their model, the principal also has to pay an additional information rent to prevent the agent's deviation to get an informational advantage over the principal. The model differs in the underlying learning process: DeMarzo and Sannikov (2016) study a Brownian model in which informative outputs are produced continuously, while I assume that news arrive at exponentially distributed times. While their model is suitable to analyze an unknown profitability of a continuously producing firm in which news arrive continuously, my model highlights the aspects of an innovative project in which drastic news arrive at random times. This focus allows me to study staged contracts and the role of deadlines and bonus payments. Prat and Jovanovic (2014) consider a model similar to DeMarzo and Sannikov (2016) with a risk averse agent and a constant quality. As information arrives continuously in their model, early deviations have a stronger impact on the belief diversion than later deviations.

2.2 Model

There is an agent (entrepreneur, he) with access to a project of unknown quality that has to complete two sequential stages, $i \in \{1, 2\}$. The agent has no wealth and contracts with a principal (e.g. a venture capitalist, she) to receive the necessary funds, f_i, that are required to work on stage i. After the completion of the final stage the project immediately generates a value π to the principal. The project can be either good or bad, $\omega \in \{g, b\}$. Only a good project can complete both stages; however, a bad project may complete the first stage. The agent has to undergo experimentation to learn about the quality of the project and to advance it towards completion. Experimentation is modeled as a two-armed bandit in continuous time, $t \in [0, \infty)$. The agent chooses in time interval $[t, t + dt)$ how much effort to exert, i.e., chooses $a_{i,t} \in [0, 1]$, which comes at cost $a_{i,t}c$.

A project of quality ω generates a success in stage i with probability $\lambda_i^\omega a_{i,t}dt$ if effort $a_{i,t}$ has been exerted in time interval $[t, t + dt)$. I assume that the intensity rate of a good project is higher than the intensity of a bad project, $\lambda_i^g > \lambda_i^b$. Moreover, only a good project can succeed in the second stage, i.e., $\lambda_2^g > \lambda_2^b = 0$.[5] The principal and the agent hold a common initial belief $p_0 \in (0, 1)$ that the project is of good quality. Let $p_{i,t}(\{a_{i,s}\}_{0 \leq s < t})$ denote the belief that the project is of good quality at time t in stage i given effort path $\{a_{i,s}\}_{0 \leq s < t}$.

Breakthroughs are immediately publicly observed. The public history at time t, $h^t \in \mathcal{H}^t$, consists of the success times, i.e., $\{\tau_i\}_{i \in \{1,2\}}$ with $\tau_i \in \{\emptyset \cup \mathbb{R}_+\}$ where the emptyset refers to the case that no success in stage i has been obtained yet. Note that if I did assume that the success is verifiable but not publicly observable, the most profitable deviation of an agent could be to exert effort, but hide a potential success. However, I show in an extension that under the optimal contract with public observability of successes, the agent would have no incentive to hide a success even if he could do so. Denote by $h_\alpha^t \in \mathcal{H}_\alpha^t$ the private history of the agent at time t that consists of the public history as well as the agent's effort choices in each of the stages, $\{a_{i,t}\}_{0 \leq s < t}$. The history of past effort choices matters only through its aggregation in each stage $A_{i,t} = \int_0^t a_{i,t}dt$ as this determines the agent's belief. Hence, I can restrict attention to private histories of the form $\mathcal{H}_\alpha^t \in \{\mathcal{H}^t \times \{\emptyset \cup [0, t]\} \times \{\emptyset \cup [0, t]\}\}$. The agent's strategy is therefore a measurable map from calendar time and his private history into the unit interval, $a_{i,t}(h_\alpha^t) : \mathcal{R}_+ \times \mathcal{H}_\alpha^t \to [0, 1]$. To simplify notation, I drop the explicit dependence on the history and keep only the time index t.

The principal offers the agent a profit-maximizing payment process conditioning on the public history to which she is fully committed from time zero on. I restrict attention to deterministic contracts. A payment process consists of a flow payment, $w_f(h^t)$, and a lump-sum payment, $w_l(h^t)$ at every history $h^t \in \mathcal{H}^t$. In the Appendix, I show that it is without loss of generality to restrict attention to bonus contracts; that is, to contracts that have payments only at time zero, and the success times. Denote by $\tilde{\mathcal{H}}^t$ the subset of public histories at time t with a breakthrough at time t. Hence, a bonus contract maps for every history $h^t \in \tilde{\mathcal{H}}^t$ a bonus payment $b(h^t) \in \mathbb{R}$ to the agent and chooses a payment b_0 at time zero. To simplify notation, I drop the dependence on the history of the bonus payment and denote a bonus payment in stage i given the history h^t as $b_{i,t}$. No payments take place at histories $h^t \notin \{\tilde{\mathcal{H}}^t \cup \mathcal{H}^0\}$. I assume that the agent is subject to limited liability. Hence, at every history h^t the agent's bonus payment is nonnegative. Note

[5]The results would not change qualitatively if a bad project could also succeed in the second stage when a bad project's success probability in the second stage is sufficiently low that it would not be funded if it was known to be of bad quality.

that without limited liability the principal could obtain the first-best by having the agent make a payment equal to the expected value of the project at time zero.

For expositional purposes and to help building intuition, I consider in the main text that discounting is in the limit $r = 0$. All results remain qualitatively unchanged with a common and positive but sufficiently small discount rate $r > 0$. All proofs are carried out with $r > 0$ in the Appendix.[6]

Given a terminal history h with success times $\{\tau_i\}_{i\in\{1,2\}}$ and bonus payments $b_{i,t}$, the principal's payoff is given by

$$e^{-r\tau_2}\left(\pi - b_{2,\tau_2}\right) - e^{-r\tau_1} b_{1,\tau_1}.$$

Similarly for the agent

$$e^{-r\tau_1} b_{1,\tau_1} - \int_0^{\tau_1} e^{-rt} c a_{1,t} dt + e^{-r\tau_2} b_{2,\tau_2} - e^{-r\tau_1} \int_0^{\tau_2} e^{-rt} c a_{2,t} dt.$$

The agent's outside option is normalized to zero.

2.2.1 Learning

A breakthrough is informative but not conclusive. After a success at time τ in stage 1 the belief jumps according to Bayes' rule to[7]

$$p_{2,0}(A_{1,\tau}) = \frac{\lambda_i^g p_{1,\tau}(A_{1,\tau})}{\lambda_i^g p_{1,\tau}(A_{1,\tau}) + \lambda_i^b (1 - p_{1,\tau}(A_{1,\tau}))}.$$

Hence, the less effort has been exerted until a success is achieved, the higher is the upward jump of the belief after a success. When effort is exerted but no breakthrough is observed the agent becomes more pessimistic about project quality as $\lambda_i^g > \lambda_i^b$. The belief follows the differential equation[8]

$$dp_{i,t} = -p_{i,t}(1 - p_{i,t})\Delta\lambda_i a_t$$

where $\Delta\lambda_i \equiv \lambda_i^g - \lambda_i^b$ and initial condition $p_{1,0} = p_0$ and $p_{2,0} = p_{2,0}(A_{1,\tau})$ as defined above. Hence, the belief drifts downwards if the agent exerts effort. To simplify notation, denote $\lambda_1^g = \lambda^g$, $\lambda_1^b = \lambda^b$ and $\lambda_2^g = \lambda$ while $\lambda_2^b = 0$ by assumption. $\lambda^g > \lambda^b$ implies that the absence of a

[6] To make the analysis of the informative milestone interesting, I assume that discounting is sufficiently small. Strategic incentives in the present setting are driven by the possibility to delay effort. However, if the agent discounts the future more (if r becomes large), the agent becomes less strategic. In the limit case of a myopic agent, the efficient outcome is obtained. Assuming $r = 0$ throughout introduces a technical difficulty in the proof that full effort will be implemente by the principal. This can be circumvented by assuming a strictly positive discount rate.

[7] Note that to be precise, $p_{1,\tau}$ in this equation is $p_{1,\tau-}$, i.e., the left-limit of the belief held at τ. For almost all t, that is whenever no success occurs, $p_{1,t-} = p_{1,t}$. This is to say that the action at t cannot condition on the arrival of a success at t.

[8] This follows from calculating the belief at $t + dt$ via Bayes' rule and taking the limit $dt \to 0$.

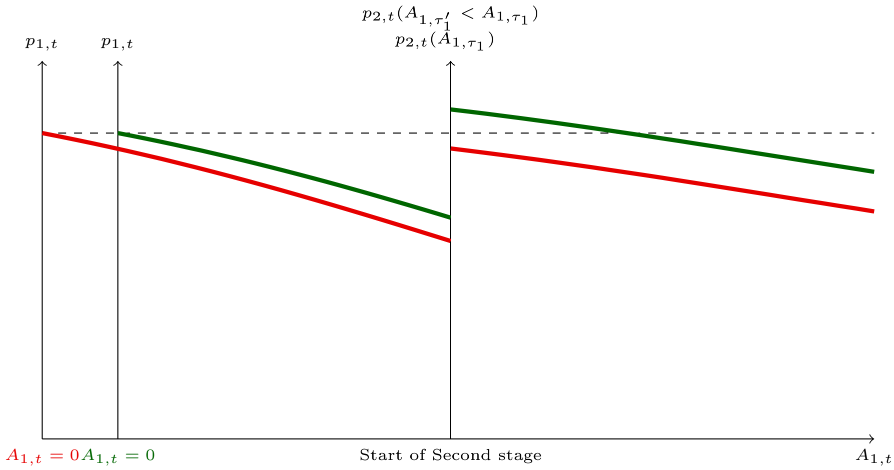


Figure 2.1: Belief path depending on total amount of effort up to success.
The red line plots the belief path if more effort is required to obtain the first breakthrough, while the green line plots belief path obtains the first breakthrough if less effort is required.

breakthrough makes players more pessimistic about the state of the project.

Note that beliefs do depend on the total effort that has been exerted in each of the stages, but not on how it was distributed over time. Hence, the higher is total effort until t, the lower is the belief about the project quality at t. This is illustrated in Figure 2.1. Learning is private because the agent's effort choices are unobserved by the principal. However, the principal holds a belief about the agent's effort. If the agent's choices coincide with the principal's belief about these, their beliefs about the project's quality coincide. Otherwise, if the agent has exerted less (more) effort than expected by the principal, the principal is more pessimistic (optimistic) than the agent.

2.3 First-Best Benchmark

As a benchmark consider a social planner that maximizes the sum of payoffs. This optimization is solved by backward induction through the stages. Hence, consider the second stage and assume that the first stage was completed at τ_1. The initial belief at the beginning of the second stage is therefore $p_{2,0}(A_{1,\tau_1})$ with $A_{1,\tau_1} = \int_0^{\tau_1} a_{1,t}dt$. Note first that the probability of reaching time t in the second stage is given by[9]

$$e^{-\int_0^t p_{2,s}(A_{1,\tau_1})\lambda a_{2,s}ds}$$

[9]To ease notation, assume that the clock is restarted when the second stage is reached.

and the instantaneous success probability by

$$p_{2,t}(A_{1,\tau_1})\lambda a_{2,t}dt$$

which implies that the probability of a success in $[t, t+dt)$ is given by

$$e^{-\int_0^t p_{2,s}(A_{1,\tau_1})\lambda a_{2,s}ds} p_{2,t}(A_{1,\tau_1})\lambda a_{2,t}dt.$$

Therefore, in the second stage the social planner chooses $\{a_t\}_{t\geq 0}$ to maximize

$$\int_0^\infty e^{-\int_0^t p_{2,s}(A_{1,\tau_1})\lambda a_{2,s}ds} a_{2,t}\left(p_{2,t}(A_{1,\tau_1})\lambda\pi - c\right)dt.$$

This gives as optimal choice $a_{2,t} = 1$ for all t such that $p_{2,t}(A_{1,\tau_1})\lambda a_{2,t}\pi \geq c$ and $a_{2,t} = 0$ otherwise. The optimal experimentation duration is given by

$$\begin{aligned} p_{2,T_2^{FB}(A_{1,\tau_1})}(A_{1,\tau_1})\lambda a_{2,T_2^{FB}}\pi &= c \\ T_2^{FB}(A_{1,\tau_1}) &= \frac{1}{\lambda}\ln\left(\frac{p_{2,0}(A_{1,\tau_1})}{1-p_{2,0}(A_{1,\tau_1})}\frac{\pi\lambda - c}{c}\right). \end{aligned}$$

This optimal deadline generates value

$$\Pi_2(A_{1,\tau_1}) = \int_{\tau_1}^{\tau_1 + T_2^{FB}(A_{1,\tau_1})} e^{-\int_{\tau_1}^t p_{2,s}(A_{1,\tau_1})\lambda ds}\left(p_{2,t}(A_{1,\tau_1})\lambda\pi - c\right)dt - f_2$$

and therefore, in the first stage the principal solves

$$\Pi = \max_{a_{1,t}} \int_0^\infty e^{-(p_{1,s}\lambda^g + (1-p_{1,s})\lambda_b)a_{1,s}ds} a_{1,t}\left((p_{1,t}\lambda^g + (1-p_{1,t})\lambda^b)\Pi_2(A_{1,t}) - c\right)dt - f_1.$$

The optimal effort policy has $a_{1,t} = 1$ if $(p_{1,t}\lambda^g + (1-p_{1,t})\lambda^b)\Pi_2(A_{1,t}) \geq c_2$ and $a_{1,t} = 0$ otherwise.

Note that not only does the probability of succeeding in the first stage decrease in the absence of a success but also the continuation value of reaching the next stage. Because increasing pessimism in the first stage induces also higher pessimism in the second stage, the initial belief of the second stage is lower if more effort was required to reach that stage. This follows from the logic that bad quality projects require more effort to successfully complete the first stage.

As a consequence, the better the project performed in the first stage, the higher is the optimal amount of total experimentation in the second stage. Earlier successes are better news about the project and therefore generate more optimisim about its quality. The first-best experimentation policy therefore uses the first-stage performance to adjust the optimal amount of experimentation in the second stage.

2.4 Derivation of the Optimal Contract

In this section, I derive the optimal contract via backward induction. That is, I first derive the continuation contract for the second stage. Taking this continuation contract as given, I move to the first stage and study the agent's incentives and solve for the optimal contract. To apply backward induction, I need to ensure that the principal cannot improve upon the optimal continuation contract by committing to a suboptimal continuation contract that reduces deviation incentives in the first stage. I show that the optimal continuation contract subject to promise keeping is indeed the contract that gives the least incentives to deviate which allows me to use backward induction.

2.4.1 Second-Stage Continuation Contract

I first study the optimal continuation contract after a first-stage success. I proceed in several steps. First, I define the principal's optimization problem. Second, I derive the incentive-compatible bonus payment process that implements any desired effort path. Third, I derive the optimal continuation contract. Finally, I consider the agent's value after a deviation in the first stage.

The principal enters this stage with a belief $p_{2,0}(\hat{A}_{1,\tau_1})$ that depends on her belief about the total effort that has been exerted up to the success time τ_1 in the first stage, where $\hat{A}_{1,\tau_1} \equiv \int_0^{\tau_1} \hat{a}_{1,t} dt$. The principal can only condition on this belief as she can condition on the public history which consists of the success times only. In this subsection, I first assume that the agent has not deviated in the first stage implying that the belief held by principal and agent at the beginning of the stage coincide. To induce the desired effort, the contract has to satisfy the agent's incentive-compatibility condition. Hence, the principal solves the following optimization problem[10]

$$(OBJ_2) \qquad \Pi(\hat{A}_{\tau_1}, v(\tau_1)) = \max_{a_{2,t}, b_{2,t}(\tau_1)} \int_0^\infty e^{-\int_0^t p_{2,s}(\hat{A}_{1,\tau_1})\lambda a_{2,s} ds} a_{2,t} p_{2,t}(\hat{A}_{1,\tau_1})\lambda(\pi - b_{2,t}(\tau_1))dt$$

$$(IC_2) \qquad s.t.\ a_{2,t} \in \arg\max_{\tilde{a}} \int_0^\infty e^{-\int_0^t p_{2,s}(\hat{A}_{1,\tau_1})\lambda \tilde{a}_{2,s} ds} \tilde{a}_{2,t}(p_{2,t}(\hat{A}_{1,\tau_1})\lambda b_{2,t}(\tau_1) - c)dt$$

$$(PK) \qquad v(\hat{A}_{1,\tau_1}) \geq (=) \int_0^\infty e^{-\int_0^t p_{2,s}(\hat{A}_{1,\tau_1})\lambda \tilde{a}_{2,s} ds} \tilde{a}_{2,t}(p_{2,t}(\hat{A}_{1,\tau_1})\lambda b_{2,t}(\tau_1) - c)dt$$

where $v(\hat{A}_{1,\tau_1})$ is the utility the agent is promised from the first stage. Hence, condition PK means that the agent's utility from the second-stage contract has to equal to $v(\hat{A}_{1,\tau_1})$. Note that it depends on the first stage whether the promise-keeping constraint has to hold with equality or inequality. It may be optimal to commit to a value less than the desired level if this reduces first-stage information rents. This will be discussed in the analysis of the first stage. The principal maximizes her payoff by choosing an effort path $\{a_{2,t}\}_{t\geq 0}$ she wants to induce.

[10] Note that deadlines are always implemented by the bonus dropping to zero at the desired point in time. Also, note that it is without loss to restart the time variable at the beginning of the second stage. Hence, I use for the second stage the interval $[0, T_2]$ instead of $[\tau_1, \tau_1 + T_2]$.

For the agent to follow that recommendation the bonus payment has to be chosen such that the agent finds it indeed optimal to choose that effort path. That is, $b_{2,t}(\hat{A}_{1,\tau_1})$ has to satisfy IC_2 as well.

Incentive Compatibility. I first study the agent's effort choice and derive the incentive-compatible second-stage bonus payment that induces effort of the agent up to a deadline. The agent's effort choices in the second stage have two effects. First, effort is required to obtain a success at the current instant. Second, effort determines the learning; if more effort has been exerted without a success, the more pessimistic is the agent. Whenever the agent has followed the principal's effort recommendations, their beliefs coincide. However, by deviating from the recommended effort path, the agent can divert his private belief from the principal's belief. In this case, the bonus payment is tailored to the belief the principal holds. This induces a dynamic agency rent. To build intuition, consider a dynamic programming heuristic similar to Bonatti and Hörner (2011). Recall that only a good project can succeed and conditional on a success the value of the project, π realizes and the agent receives bonus $b_{2,t}$.

$$V_t = (1 - e^{-a_{2,t}p_{2,t}\lambda dt})b_{2,t} - ca_{2,t}dt + e^{-a_{2,t}p_{2,t}\lambda dt}V_{t+dt}$$

Using the analogous approximation for V_{t+dt}, approximating the exponentials with a second-order Taylor expansion, dividing by dt^2 and taking the limit as $dt \to 0$ yields for the effect of delaying effort

$$\left(-\frac{\partial V_t}{\partial a_{2,t}} + \frac{\partial V_t}{\partial a_{2,t+dt}}\right)/dt^2 = \dot{b}_{2,t}p_{2,t}\lambda.$$

By shifting effort from today to tomorrow, the agent loses the marginal payoff from effort today, $p_{2,t}\lambda b_{2,t}$, but gains in return the marginal benefit of effort, $p_{2,t}\lambda b_{2,t+dt}$. If the principal were to use an increasing bonus process, the agent had an incentive to delay effort; this induces a *procrastination rent*. If the principal wants to make the agent indifferent between all effort levels, incentive compatibility implies $\dot{b}_{2,t} = 0$. If the bonus process were decreasing, the agent preferred to frontload effort.

Lemma 2.1. *The time-independent profit-sharing rule $b_{2,t}(A_{1,\tau_1}) = b_2(A_{1,\tau_1})$ that makes the agent indifferent between all effort levels at the second-stage deadline $T_2(A_{1,\tau_1})$, i.e.*

$$b_2(A_{1,\tau_1}) = \frac{c}{\lambda}\frac{1}{p_{2,T_2(A_{1,\tau_1})}(A_{2,T_2})}$$

induces the agent to exert full effort for all $t \in [\tau_1, \tau_1 + T_2(A_{\tau_1})]$. For $t > \tau_1 + T_2(A_{\tau_1})$, $b_2(A_{1,\tau_1}) = 0$.

Note that this bonus payment depends on the total effort that has been required in the first stage. This, as a consequence of the informativeness of the first stage, affects the continuation contract because it determines the belief about the project quality in the second. The bonus chosen by the principal depends on her belief about the effort choices of the agent, i.e., on $\hat{A}_{1,\tau_1}$. However, for now, I assume that the agent has not deviated in the first stage and therefore $\hat{A}_{1,\tau_1} = A_{1,\tau_1}$. With positive discounting, a delay in effort were less attractive and the principal could save on some procrastination rents. The bonus payment was slightly increasing; however, the intuition for the incentives to delay effort were unaltered.

Principal's Optimization. Next, I study the principal's preferred contract subject to the incentive-compatibility and promise-keeping constraints. Because the absence of a success is bad news and the principal as well as the agent become increasingly pessimistic, she will terminate experimentation in finite time. It will turn out that the principal frontloads effort in the second stage. That is, she wants to induce $a_{2,t} = 1$ for all times up to a deadline. Hence, the problem boils down to determining a maximum level of total effort that she wants to induce in the second stage. Because she wants to frontload experimentation and the effort level is at its maximum, total experimentation on path coincides with calendar time, $A_{2,t} = t$.

Recall that incentive compatibility induces a weakly decreasing bonus process. Moreover, promise-keeping requires that the agent's expected utility in the second stage is at least as high as the promise from the first stage, $v(A_{1,\tau_1})$. If the promise-keeping constraint is binding, the principal has to choose how to deliver additional utility to the agent. She can either pay higher bonuses for a success or she can extend the deadline and thereby increase the probability of obtaining the bonus. It is optimal for the principal to deliver additional utility by incentivizing agents to work until extended deadlines. To see why this is optimal, note that the experimentation deadline in the second stage will be distorted downwards from the efficient level derived in Section 2.3 due to the procrastination rents. Having the agent exert more effort before terminating experimentation increases the total surplus as well as the agent's expected utility generated in the second stage. Therefore, the principal chooses the contract such that the agent receives the promised utility at the highest total surplus. The level of the bonuses is pinned down by the agent's incentive compatibility condition at the deadline.

$$b_{2,T_2}(A_{1,\tau_1}) = \frac{c}{\lambda} \frac{1}{p_{2,T_2}(A_{1,\tau_1})}. \tag{2.1}$$

If a longer deadline is chosen, the agent is more pessimistic at the deadline and therefore a higher bonus is required to incentivize him to exert effort. The agent's value of a contract with constant bonus process and deadline T_2 is given by

$$v(T_2; A_{1,\tau_1}) = c\,(1 - p_{2,0}(A_{1,\tau_1})) \left(\frac{e^{\lambda T_2} - 1}{\lambda} - T_2 \right). \tag{2.2}$$

I denote the outcome of a maximization of the principal without promise-keeping constraint by *second-best* contract. The corresponding deadline is denoted by $T_2^{SB}(A_{1,\tau_1})$ and the corresponding utility by $v(T_2^{SB}(A_{1,\tau_1}))$. Note that it may be optimal to commit to a continuation utility that is lower than the second-best utility in the first stage to reduce deviation incentives. When this will occur, will be discussed in the following subsection.

Proposition 2.1. *The principal-optimal second-stage contract given first-stage success time τ_1, corresponding total effort in the first stage A_{1,τ_1} and agent's promised utility $v(A_{1,\tau_1})$ is given by*

$$T_2(A_{1,\tau_1}) = \begin{cases} T_2^{SB}(A_{1,\tau_1}) & , \text{ if } v(T_2^{SB}(A_{1,\tau_1})) \geq v(A_{1,\tau_1}) \\ T_2(A_{1,\tau_1}, v(A_{1,\tau_1})) & , \text{ if } v(T_2^{SB}(A_{1,\tau_1})) < v(A_{1,\tau_1}) \end{cases}$$

or if the principal commits to providing value less than the second-best by

$$T_2(A_{1,\tau_1}) = T_2(A_{1,\tau_1}, v(A_{1,\tau_1})), \quad \text{for all } v(A_{1,\tau_1})$$

where $T_2(A_{1,\tau_1}, v(A_{1,\tau_1}))$ is defined as the solution, T, to

$$v(A_{1,\tau_1}) = c(1 - p_{2,0}(A_{1,\tau_1})) \left(\frac{e^{\lambda T} - 1}{\lambda} - T \right)$$

which is given by[11]

$$T_2(A_{1,\tau_1}, v(A_{1,\tau_1})) = -\frac{v(A_{1,\tau_1})}{c(1 - p_{2,0}(A_{1,\tau_1}))} - \frac{1}{\lambda} \left(1 + W_{-1} \left(-e^{-1-\lambda \frac{v(A_{1,\tau_1})}{c(1-p_{2,0}(A_{1,\tau_1}))}} \right) \right).$$

$T_2^{SB}(A_{1,\tau_1})$ is given by

$$T_2^{SB}(A_{1,\tau_1}) = \frac{1}{2\lambda} \ln \left(\frac{p_{2,0}(A_{1,\tau_1})}{1 - p_{2,0}(A_{1,\tau_1})} \frac{\pi\lambda - c}{c} \right).$$

The corresponding bonus payment is given by

$$b_{2,t}(A_{1,\tau_1}) = \frac{c}{\pi\lambda} \frac{1}{p_{T_2(A_{1,\tau_1})}}.$$

One important feature of the optimal second-stage continuation contract is that it uses deadlines as the main instrument to deliver utility to the agent: given a deadline, the principal always uses the lowest possible bonus payment that incentivizes the agent to exert effort until that deadline. The underlying reason is that extending deadlines reduces inefficiencies in the total amount of experimentation in the second stage and therefore increases the overall surplus. In addition, by rewarding with extended deadlines, the bonus payment is as low as possible given the promise-keeping condition. Hence, generating the maximum surplus and keeping the bonus payment as low as possible while keeping the promise from the first stage are both obtained through extended deadlines. This observation will lateron lead to the conclusion that this contract is not only profit-maximizing in the second stage but also the contract that yields the lowest incentives to deviate in the first stage.

[11] $W_{-1}(x)$ denotes the negative branch of the Lambert-W-function.

The main comparative statics that are relevant for the analysis of the first stage are summarized in the following corollary.

Corollary 2.1 (Comparative Statics of the Continuation Contract.).
The bonus and the deadline in the second stage are (weakly) increasing in the promised utility for a given initial second-stage belief.
The bonus in the second stage is increasing and the deadline decreasing in the initial second-stage belief for a given level of promised utility.

That bonus and deadline are weakly increasing in the promised utility follows from the way the principal provides the agent with additional utility: she extends the deadline and to incentivize the agent to exert effort until the new deadline she has to promise a higher bonus payment conditional on success. The deadline is decreasing in the initial belief because for every deadline, the principal has to provide the agent with higher bonuses to incentivize him. However, she does not want to reduce the deadline too much as this also reduces the probability of obtaining the final breakthrough.

Agent's Continuation Value after a First-Stage Deviation. To study the agent's incentives in the first stage I need to evaluate his continuation payoff after a deviation in the first stage. An agent could deviate by making effort choices that are different from the principal's recommendation. Off-path effort choices have no direct benefit but divert the agent's from the principal's belief. Due to the informativeness of the milestone and the resulting persistence of the private information the deviation has two consequences. First, it affects the agent's belief in the first stage. Second, it affects the initial belief of the second stage because the required effort in the first stage is informative about project quality. If the agent has exerted less effort in the first stage than the principal believes he has, he is more optimistic about the project's quality than the principal. Therefore, the agent will value the continuation contract differently than the principal believes he does. The value of a continuation contract given that the principal beliefs the exerted effort is $\hat{A}_{1,\tau_1}$ while the true exerted effort is A_{1,τ_1} is given by

(2.3)
$$v(t, A_{1,\tau_1}, \hat{A}_{\tau_1}) = c\left(p_{2,0}(A_{1,\tau_1})\frac{1 - p_{2,0}(\hat{A}_{1,\tau_1})}{p_{2,0}(\hat{A}_{1,\tau_1})}\frac{e^{\lambda T_2(A_{1,\tau_1})} - 1}{\lambda} - (1 - p_{2,0}(\hat{A}_{1,\tau_1}))T_2(A_{1,\tau_1})\right).$$

It is straightforward to show that the agent's value is decreasing in $A_{1,t}$; that is, for every continuation contract he prefers to hold a higher belief than the principal. This already foreshadows that the agent has an incentive in the first stage to shirk in order to become more optimistic than the principal and thereby increase his continuation payoff. I will lateron show that the contract derived in this section is also the contract that yields the smallest incentive to deviate to the agent given the promised utility from the first stage.

2.4.2 First-Stage Analysis

Given the analysis of the second stage, I now move to the first stage. I analyze the agent's incentives to exert effort first and then study the principal's optimal contract. To incentivize the agent to work in the first stage the principal has to promise a reward in case of a success. As the first stage is followed by the second stage, the principal can use the experimentation assignment in the second stage as a reward instrument. However, the principal can also use a bonus payment to reward the agent for a first-stage success that is independent of future performance. The total reward of the agent consists of both, the bonus payment and the value of the continuation contract

$$w(\tau_1) = b_{1,\tau_1} + v(\tau_1).$$

To understand the agent's incentives in the first stage it is important to note that, in contrast to settings with independent stages, an off-path effort choice has two consequences. First, it diverts the agent's from the principal's belief in the first stage as it is the case in the second stage. Second, it also diverts the initial belief at the beginning of the second stage because the effort required to complete the first stage is informative about the project quality.

To see the impact of the latter on the agent's incentives, consider the implementation of the second-stage contract: the principal chooses the continuation contract such that it delivers in expectation the promised utility from the first stage to the agent. This expectation is calculated based on the principal's belief about the project quality. By first-stage deviations the agent can divert his private belief from the principal's in the second stage. This will affect the agent's true expected value of the continuation contract: recall the off-path value of the agent from equation (2.3). The gain of holding a marginally more optimistic belief is

$$\frac{\partial v(t, \hat{A}_{1,\tau_1}, A_{1,\tau_1})}{\partial p_{2,0}(A_{1,\tau_1})} = \frac{c}{\lambda}\left(\frac{1-p_{2,0}(A_{1,\tau_1})}{p_{2,0}(A_{1,\tau_1})}\left(e^{\lambda T_2(A_{1,\tau_1})} - 1\right) + \lambda T_2(A_{1,\tau_1})\right) > 0. \tag{2.4}$$

How this affects the agent's incentives can again be seen in a dynamic programming heuristic:

$$V_t = (1 - e^{-a_{1,t}(p_{1,t}\lambda^g + (1-p_{1,t})\lambda^b)dt})w_t - ca_{1,t}dt + e^{-a_{1,t}(p_{1,t}\lambda^g+(1-p_{1,t})\lambda^b)dt}V_{t+dt}$$

Using the analogous approximation for V_{t+dt}, approximating the exponentials with a second-order Taylor expansion, dividing by dt^2 and taking the limit as $dt \to 0$ yields for the effect of delaying effort

$$\left(-\frac{\partial V_t}{\partial a_{1,t}} + \frac{\partial V_t}{\partial a_{1,t+dt}}\right)/dt^2 = \left(\dot{w}_t - \frac{\partial v(t, \hat{A}_{1,\tau_1}, A_{1,\tau_1})}{\partial p_{2,0}(A_{1,\tau_1})}\frac{\partial p_{2,0}(A_{1,\tau_1})}{\partial A_{1,\tau_1}}a_{1,t}\right)\left(p_{1,t}\lambda^g + (1-p_{1,t})\lambda^b\right).$$

This heuristic mirrors the two effects of a deviation: first, a delay in effort affects the level of the total reward, w_t, because the agent can divert the belief in the current stage. Second, the delay in effort also affects the agent's belief in the second stage and thereby the assessment of the continuation contract. Becoming more optimistic than the principal increases the value of a continuation contract that is tailored to a more pessimistic agent. Therefore, the persistence in the private information across stages creates an endogenously arising *ratchet effect*: the agent wants the principal to think that the success probability in the second stage is low because in that case the principal believes that she has to promise high payments conditional on second-stage success to incentivize the agent.

So far, I have assumed that the second-stage contract is implemented as derived in the previous section. By full commitment this is not necessarily the case because it could be better for the principal to commit to a suboptimal second-stage contract that reduces the deviation incentives in the first stage. The following lemma shows that the optimal second-stage contract is the implementation of the promised utility from the first stage that induces the lowest incentive to deviate allowing me to use backward induction.

Lemma 2.2. *The continuation contract derived in Proposition 2.1 induces the lowest incentives to deviate in the first stage while satisfying the promise-keeping condition.*

Intuitively, this result holds because the contract in Proposition 2.1 is the implementation with the lowest bonus payment after a second-stage success. The ratchet effect in the first stage arises because by making the principal more pessimistic the agent is promised a higher bonus conditional on success in the second stage. This effect is increasing in the second-stage bonus and therefore the incentive to deviate is increasing in the bonus payment. Hence, the principal wants to implement the continuation contract such that the bonus payment is as low as possible. As a consequence, she rather extends the deadline further and thereby increases the success probability rather than increasing only the bonus payment keeping the deadline at the second-best level.

Taking this continuation contract as given, I next characterize the minimal total reward process that induces incentive compatibility of an effort path $\{a_{1,t}\}_{t\geq 0}$ in the following proposition.

Proposition 2.2. *The minimal required continuaton utility to induce effort $\{a_{1,t}\}_{t\geq 0}$ in the first stage solves the following differential equation*

$$\dot{w}(t) = \frac{\partial v(t, \hat{A}_{1,t}, A_{1,t})}{\partial A_{1,t}} a_{1,t}$$

with boundary condition

$$w(T_1) = \frac{c}{p_{1,T_1}\lambda^g + (1 - p_{1,T_1})\lambda^b}$$

and $w_t = 0$ for all $t > T_1$, if there is a T_1 such that $a_{1,t} = 0$ for all $t > T_1$.

This proposition shows that due to the informativeness of the first stage the agent has to receive an additional rent to exert effort if he is assigned experimentation in the second stage. Note that in the absence of the informativeness of the milestone this rent would not be required as then $\frac{\partial v(t,\hat{A}_{1,t},A_{1,t})}{\partial A_{1,t}} = 0$ which is the case, for example, in Moroni (2016). As $\frac{\partial v(t,\hat{A}_{1,t},A_{1,t})}{\partial A_{1,t}} < 0$, the agent's continuation value is decreasing over time and can be disentangled into three components

$$\dot{w}_t = \underbrace{-\frac{c}{\dot{p}_{1,t}}}_{\text{static MH}} + \underbrace{\frac{c}{\dot{p}_{1,t}}}_{\substack{\text{procrastination}\\ \text{rent}}} + \underbrace{\frac{\partial p_{2,0}(A_{1,t})}{\partial A_{1,t}} a_{1,t}}_{\substack{\text{effect of effort}\\ \text{on belief}}} \underbrace{\frac{\partial v(t, \hat{A}_{1,t}, A_{1,t})}{\partial p_{2,0}(A_{1,t})}}_{\substack{\text{effect of belief}\\ \text{on value}}} \tag{2.5}$$

The first term corresponds to the agent's instantaneous cost of effort that he has to be compensated for. This changes over time as the agent becomes more pessimistic when exerting effort. However, this induces the procrastination incentive to obtain a higher reward and the agent has to be granted a procrastination rent. Moreover, due to the persistence of the private information across the stages, the agent has to obtain the informativeness rent. To incentivize effort the total reward on path has to decrease sufficiently steeply over time. The rate at which it decreases is such that the gain from delaying effort and thereby becoming more optimistic than the principal in the second stage is at most as large as the value the agent loses from not succeeding today. Importantly, it has to decrease more steeply if the value of the continuation contract is higher because it implies that a higher bonus payment is required in the second stage. Hence, the more utility the agent receives through a continuation contract, the higher is the incentive to divert the beliefs. This induces first-stage information rents to increase in the value of the continuation contract at a given success time for all earlier success times. This creates a downside of using continuation contracts and therefore long-term incentives because they induce informativeness rents in the first stage. However, using continuation contracts also has an advantage over bonus payments: by using continuation contracts the principal generates a continuation value to herself as she only receives the benefits of the project if second-stage experimentation is successful. Moreover, recall that second-stage experimentation is inefficiently short due to procrastination rents. Suppose that the principal has to provide the agent with an additional unit of total reward after a first-stage success. Then, she has to choose whether to deliver this through a first-stage bonus payments, which implies lower information rents for earlier successes in the first stage, or through a more valuable continuation contract, which generates additional surplus in the second stage through more efficient second-stage experimentation. This tradeoff is illustrated in Figure 2.2. Introducing a bonus payment after late first-stage successes reduces the information rents at all previous success times but costs profits at the respective success time.

Therefore, the informativeness of the first stage induces a tradeoff between short-term incentives that only condition on current performance, i.e., bonus payments after the first success, and long-term incentives that condition on future performance as well. This tradeoff does not arise in a setting without persistent information across stages because the agent does not have the ability to divert the beliefs that the second-stage contract terms condition on. Therefore, in a setting with independent stages the principal would never use bonus payments that only condition on first-stage success. In the following I am restricting attention to the case of *costly incentives* defined below.

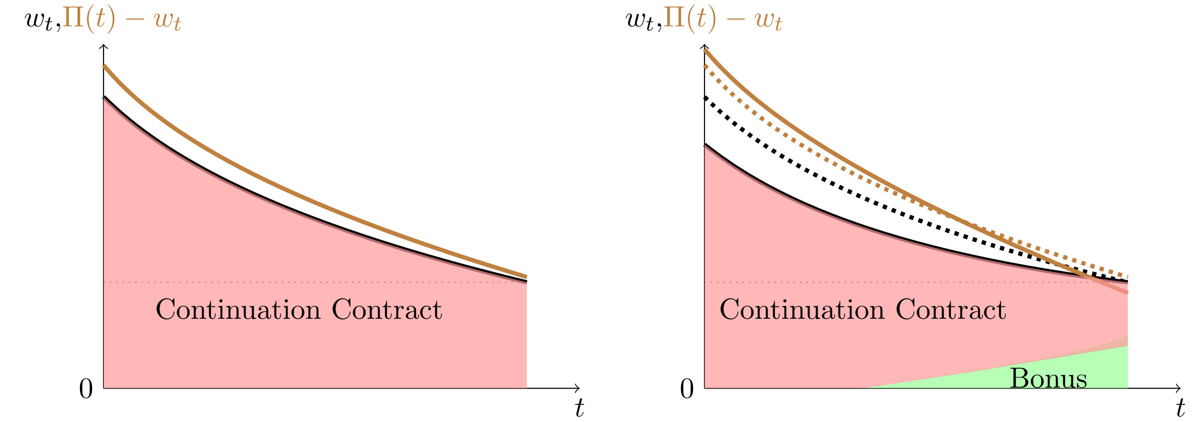


Figure 2.2: Tradeoff between Continuation Contract and Bonus Payment.
The left panel shows a hypothetical total reward of the agent as a function of the success time if the agent were rewarded with a continuation contract only. The total reward therefore decreases relatively steeply. The brown line illustrates the corresponding profits of the agent. The right panel shows how the optimal contract improves on the hypothetical contract illustrated in the left panel: by introducing bonus payments at the end, it reduces the informativeness rents for all earlier success. The dotted lines are the total reward and profits from the left panel as benchmark.

Definition 1 (Costly Incentives)**.** *First stage incentives are costly if the agent's first-stage incentive constraint is binding for all $t \in [0, T_1]$. That is*

$$\gamma_t \equiv \dot{w}_t - \frac{\partial v(t, \hat{A}_{1,t}, A_{1,t})}{\partial A_{1,t}} a_{1,t} \tag{2.6}$$

is such that $\gamma_t = 0$ for all $t \in [0, T_1]$. A sufficient condition for costly incentives to occur is that $\dot{w}_t \leq \dot{v}_t^{SB}$ and $w_{T_1} \geq v_{T_1}^{SB}$.

This implies that the principal does not provide the agent with more utility than necessary to ensure incentive compatibility in the first stage. That is, the first-stage incentive constraint is binding for all $t \in [0, T_1]$. It may be the case that under some parameter values, the principal is willing to give more utility to the agent than necessary. This can occur if the incentives in the first stage are relatively cheap such that a continuation value less than the value of the second stage alone would incentivize the agent to work in the first stage. If incentives are relatively cheap, the incentive constraint still requires $\gamma_t \geq 0$ as in Proposition 2.2. As the main contribution of my paper lies in the analysis of the first-stage incentives and the corresponding optimal contract, I am assuming that first-stage incentives are costly.

The following theorem shows how the optimal contract solves this tradeoff in the costly incentives case.

Theorem 2.1. *Suppose that first-stage incentives are costly. The total reward w_t the agent receives conditional on completing the first stage at time t induces full effort, is strictly decreasing* *for all $t \in [0, T_1]$ and solves*

$$\dot{w}_t = \frac{\partial p_{2,0}(A_t)}{\partial A_t} \frac{\partial v(t, A_t)}{\partial p_{2,0}(A_t)} a_t \quad s.t.\ w(T_1) = \frac{c}{p_{T_1}\lambda^g + (1 - p_{T_1})\lambda^b}$$

and has $w_t = 0$ for all $t > T_1$. There is a $\hat{t} \in (0, T_1)$ such that $w_t = v_t$ and $b_{1,t} = 0$ for all $t \in [0, \hat{t}]$; i.e., early successes are rewarded with continuation contracts only.
For all $t \in (\hat{t}, T_1]$, $w_t > v(t)$ and $b_{1,t} > 0$ with $\frac{b_{1,t}}{w_t}$ increasing in t; i.e., if the success is obtained after $\hat{t}$, the total reward consists of a continuation contract and a bonus payment with the share of the bonus payment in the total reward increasing in the success time.
The continuation contract, $v(t)$, is implemented according to the optimal second-stage contract as in Proposition 2.1.

Theorem 2.1 implies that the composition of the reward changes over time. Early successes are rewarded with continuation contracts only and the second-stage contract has deadlines close to the first best. If the success arrives late, the reward consists of bonus payments as well as less valuable experimentation assignment in the second stage. The part of the reward that is provided to the agent with a bonus payment is increasing, the later the breakthrough is obtained.

The reason that a lower share of the reward is provided with continuation contracts over time is that the additional gain in overall surplus from extended deadlines is decreasing in the belief about the project's quality. This belief is decreasing in the first-stage success time. Moreover, the arising ratchet effect implies that higher information rents have to be paid for *all earlier success times* if more utility is delivered through a continuation contract. By choosing the share of the total reward that is delivered through a continuation contract, the principal can control the information rents for all earlier successes. At earlier success times, the agent has to be granted a sufficiently higher reward to prevent him from delaying effort. If a success at a later time is rewarded with a relatively high share of utility through a continuation contract, the gain from holding a more optimistic belief then is high. Therefore, a higher reward for earlier successes is needed to prevent a deviation. Hence, bonus payments become more favorable for later successes for two reasons: the effect on information rents for earlier successes increases and the gain of extending deadlines decreases in the belief.

It is interesting to note that the optimal contract provides a decreasing amount of value through a continuation contract and therefore induces the continuation contract instruments to vary with success times. The second-stage deadline is strongly dependent on the first-stage success time while the second-stage bonus is not as dependent on performance as in the second-best contract. The reason is, as discussed above, that the incentive cost is lower if the agent is rewarded with extended deadlines rather than with higher bonus payments. However, having the bonus payment become less dependent on performance reduces the ratchet effect.

Corollary 2.2. *Second-stage deadlines are decreasing in the first-stage success time and more responsive to it than in the second-best contract.*
Bonus payments are increasing in the first-stage success time and less responsive to it than in

the second-best contract.

How the second-stage contract terms depend on the first stage outcome is illustrated in Figure 2.3.

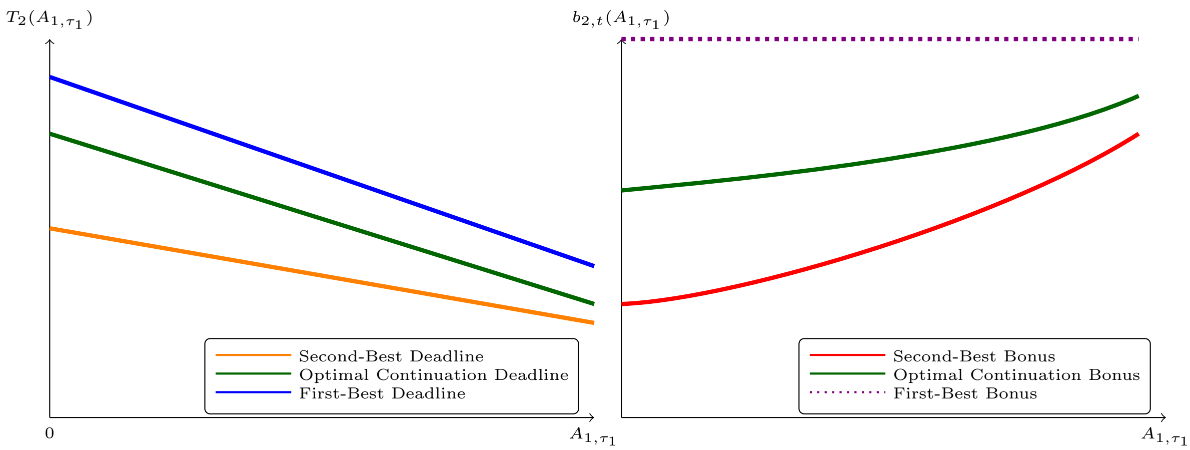


Figure 2.3: Deadline and Bonus in Optimal Contract.
The left panel shows how the second-stage deadline in the optimal contract varies with the first-stage success time compared to the first-best and second-best deadline.
The right panel show how the second-stage bonus in the optimal contract varies with the first-stage success time compared to the first-best and second-best bonus.

2.5 Endogenous Agent Replacement

In this section, I consider the case in which the principal has access to another agent for the second stage. This gives her the additional choice whether to keep the agent from the first stage to work in the second stage as well or whether she rather has a new agent and get the second-best value in the second stage. If the agent is replaced, he is rewarded with a bonus payment only and no continuation value. As shown in the previous section, this implies that at those success times at which the agent is replaced, he does not need to receive the informativeness rent. Therefore, the principal saves on rents for earlier success times when replacing the agent. However, she foregoes the possibility to gain from extended deadlines in the second stage. Replacement is the most extreme bonus payment and saves the most informativeness rents. Without a new agent this corresponds to terminating the project as no continuation value is granted to the agent and no second-stage experimentation takes place. Because the principal can obtain the second-best value after replacement if she has access to a new agent, I show that she will *always* make use of this possibility for success times close to the deadline. The optimal contract with replacement is illustrated in Figure 2.4. I assume that parameters are such that we are in the costly incentives case. If the agent has not to be granted rents that exceed the second-best value it is straightforward that replacement may not be desirable from the principal's point of view.

Theorem 2.2. *Suppose that first-stage incentives are costly. If the principal has access to another agent in the second stage, she will choose two deadlines,* $\hat{T}$ *and* T_1 *in the first stage. For all* $t \in [0, \hat{T}]$*, the agent receives* w_t *upon a breakthrough as in the optimal contract with boundary condition given by* $w_{\hat{T}} = \frac{c}{p_{T_1}\lambda^g + (1-p_{T_1})\lambda^b}$ *and works in the second stage. For all* $t \in (\hat{T}, T_1]$*, the agent receives* $w_t = b_t = w_{T_1}$ *upon a breakthrough and a new agent works on the second-stage contract according to the second-best value and the belief given by* $p_{2,0}(A_{1,\tau_1})$*. If no success has been obtained by* T_1*, the project is terminated.*

In light of the optimal contract without replacement this result seems intuitive. However, it

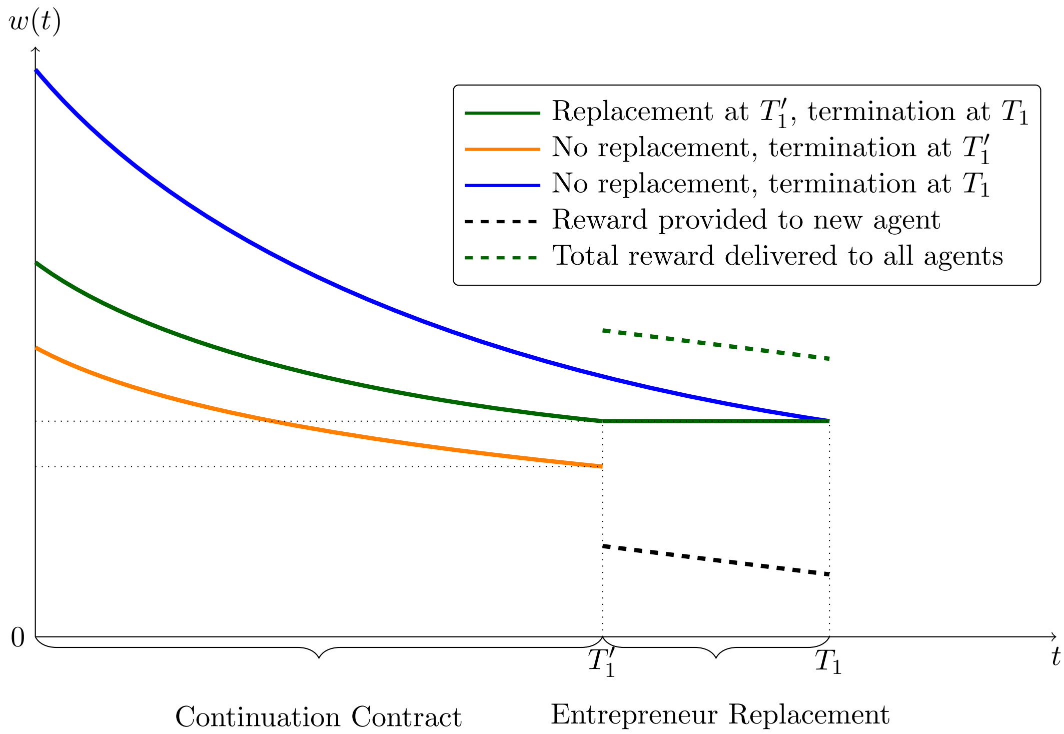


Figure 2.4: Agent Replacement.
The orange line shows a hypothetical reward process with a short deadline T_1'*. The blue line show a hypothetical reward process with longer deadline* T_1'*. The green line shows the evolution of a contract that rewards the agent at least partially with a continuation contract for a success before* T_1' *and with a bonus payment only for a success before the deadline* T_1*. The dashed black line corresponds to the new agent's value in the second stage if the first-stage agent is replaced. The dashed green line is the total reward that both agents receive in the replacement region.*

is not obvious: there is no learning about the agent's type but still the successful agent gets replaced, although continuation contracts are "cheaper" to provide the required utility than bonus payments. In particular, when the first milestone is not informative about the second stage, replacement between stages never occurs because there are no information rents to be saved by replacement. Moreover, if agents could be continuously replaced even within a stage at no cost, the principal would replace the agent continuously and induce first-best experimentation because no dynamic agency rents have to be paid at all. Theorem 2.2 shows that informative milestones give rise to replacement of agents that do succeed in their assigned task but took relatively long to do so. The underlying reason is that continuation contracts give rise to the informativeness rents caused by the persistence of learning. Hence, replacement occurs to reduce information rents in the first stage. This may be one explanation for high managerial turnover rates in the innovative industries found by Hannan et al. (1996), for example.

2.6 Extensions

2.6.1 Project Design

In this section, I study the choice of the informativeness of the first stage and endogenous staging, i.e., the decision of choosing one or two stages for the project. In many instances, a milestone may not be necessary to implement the full project but still the principal requires it: when a prototype is required, it may serve as an informative signal about the project. Typically, the principal is able to decide on the informativeness of the prototype; that is, how many details of the final product should be incorporated. The tradeoff of the principal is that on the one hand she wants the signal to be as informative as possible to prevent funding a bad-quality project in the second stage. On the other hand, if the initial informativeness is low, then if the signal becomes marginally more informative, the agent's ability to divert the beliefs increases and therefore the informativeness rent in the first stage does as well. However, the capital required for the first stage is increasing in the informativeness because the more informative is the first tak, the closer it is to the final product. Hence, the optimal level of informativeness is not obvious and may not be extreme.

The ratio $\frac{\lambda^g}{\lambda^b}$ can be interpreted as the informativeness of the first stage. The higher is the ratio the higher is the upward jump in the belief after a success. If $\frac{\lambda^g}{\lambda^b} \to \infty$, there is certainty after a first-stage success that the project is of good quality. If $\frac{\lambda^g}{\lambda^b} \to 1$, there is no learning at all. I assume that the more has been learned in the first stage, the faster is a success obtained in the first stage. Towards this, I assume that the good project's intensity rate in the second stage is given by $\lambda = \frac{\lambda^g}{\lambda^b}$. This implies that when the first stage is perfectly informative, there is no second stage because $\lambda = \infty$. I can show that the two-stage contract converges to the second-best one-stage contract as $\lambda^b \to 0$. Also, I change λ^g with λ^b such that the expected duration of project completion conditional on the quality being good remains constant, which yields $\lambda^g = 1 + \lambda^b$ when normalizing $\lambda^g(\lambda^b = 0) = 1$. An additional advantage of this formulation is that for all combinations of λ^g and λ^b under this restriction the belief evolution in the first stage is identical because $\lambda^g - \lambda^b = 1$. However, the upwards jump after a success depends on the ratio $\frac{\lambda^g}{\lambda^b}$ as the posterior is given by

$$p_{2,0} = \frac{1}{\frac{1-p_0}{p_0}\frac{\lambda^b}{1+\lambda^b}e^t + 1}$$

which is decreasing in λ^b and goes to one as λ^b goes to zero. To capture the feature of capital infusions that are contingent on milestones, I let the fixed cost per stage, f_i depend on the informativeness of the first stage. The more informative the first stage, the higher is the cost for this stage, $f_i(\frac{\lambda^g}{\lambda^b})$. I assume that if $\frac{\lambda^g}{\lambda^b} \to \infty$ the cost of the second stage converges to zero and the first stage cost converges to the one-stage case. Moreover, I continue to assume that the parameters are such that we are in the costly incentives case. Otherwise, the principal would get the informative first stage signal without having to deliver any additional rents and the staging decision became trivial.

As a first result I show that the optimal two-stage contract converges to the optimal one-stage contract as the first stage becomes perfectly informative.

Lemma 2.3. *With $\lambda = \frac{\lambda^g}{\lambda^b}$ and $\lambda^b \to 0$, the two-stage optimal contract from the previous sections, converges to the optimal one-stage contract with $\lambda^{onestage} = \lambda^g(\lambda^b = 0)$.*

This result allows me to study endogenous staging as a choice of λ^b numerically quite straightforwardly. When the principal chooses $\lambda^b = 0$ the problem collapses to a one-stage problem. Note that in the present setting, the principal cannot choose an entirely uninformative first stage except for the limit case $\lambda^b \to \infty$ because $\lambda^g = 1 + \lambda^b > \lambda^b$. In this case, again, the problem would collapse to a one-stage problem because the first stage is immediately completed and there is no way to divert beliefs for the agent. It is immediate that the principal would never choose a perfectly uninformative first stage $\lambda^b = \lambda^g$. This would make the first stage a pure moral hazard stage without any signal. Still, the agent has to be incentivized to exert effort. That is, the principal would have to deliver additional rents to the agent without gaining from the first stage.

It follows from the comparative statics of the optimal contract in the informativeness that the informativeness rent is inversely u-shaped in the informativeness. If the first stage is entirely uninformative, then the informativeness rent is zero. If the first stage is fully informative, it is zero as well. In between, it is strictly positive. Hence, moving from a one-stage project to a two-stage project with a somewhat informative first stage has the following effects: (i) a positive informativeness rent has to be delivered to the agent in expectation (ii) the principal can condition the second capital infusion on the first-stage outcome. These two effects work against each other and it depends on the parameters which one dominates. Note that if the initial belief is sufficiently high, investing all capital at once and avoiding the informativeness rent is more attractive. If the initial belief is lower, investing all capital at once is less attractive because it is lost with a high probability. However, the principal may then introduce a first stage at a cost that is lower than investing into the full project immediately to generate an informative signal and condition the second infusion on this signal.

Numerical Results. In a numerical analysis in Mathematica, I study the endogenous choice of staging and the optimal degree of informativeness of the first stage. This reveals that the previously discussed tradeoff between terminating bad projects and agent's information rents is relevant when designing a project. In several specifications, the optimal degree of informativeness is interior and hence the choice of two stages dominates a one-stage project. However, not requiring a milestone may also be optimal under other parameter values. The analysis reveals intuitive comparative statics, as the agency conflict increases, the optimal informativeness decreases. This follows because the cost of the informativeness is increasing in the agency conflict and hence, the optimal informativeness is reduced. Also, the principal chooses an inefficiently low level of informativeness.

The conjecture that projects with lower initial beliefs; i.e., more risky projects are more likely to be staged investments seems to be true in numerical examples. This is in line with the findings in Bienz and Hirsch (2011).

2.6.2 Privately Observable Successes

One important feature of the optimal contract derived is that even if successes were not publicly but only privately observable, the agent would not make use of the possibility to strategically hide a success. In principle, a profitable deviation of the agent could be to hide a success if it is obtained instead of shirking to divert the beliefs. However, the gain from hiding a success is the same as the gain from shirking for an instant: it alters the principal's belief in the following stage and therefore increases the value of the continuation contract from the agent's perspective. The optimal contract precludes this behavior by rewarding earlier successes with higher rents. Hence, the agent would immediately reveal a private breakthrough.

Corollary 2.3. *If successes are privately observed by the agent, the agent immediately reveals a success.*

It follows from this corollary that the assumption of publicly observable breakthroughs is without loss of generality. In principle, with private observability the agent has an incentive to strategically delay the arrival of a success due to the arising ratchet effect. The principal needs to impose an additional truthtelling/revelation constraint on her optimization problem. However, by inspecting the incentive to hide a success, it becomes apparent that this condition coincides with the decision to exert effort. Hence, the optimal contract satisfies the revelation constraint as well and the agent immediately reveals a success.

2.6.3 Learning About Agent's Type

If the learning is about the agent's type instead of the project and the project's quality is known, the optimal contract without replacement is as in Theorem 2.1. However, replacement is different in this case because a new agent is hired, that is at the beginning of the second stage the initial belief is back at the initial prior p_0. This changes the optimal replacement deadline, but not the incentive to introduce a replacement region. It may even be more attractive to introduce replacement because it allows to increase the belief at the beginning of the second stage if the principal became too pessimistic in the first stage. The continuation value for the principal upon replacement would therefore be unaffected by first stage outcomes and she can always guarantee herself at least this continuation payoff after a first-stage success.

Corollary 2.4. *If the agent's type is unknown, the optimal contract without replacement is as in Theorem 2.1. With replacement and independent agents, the continuation value after replacing the agent is independent of the timing of the first breakthrough and given by the second-best value under the initial prior belief* $v^{SB}(p_{2,0} = p_0)$.

This shows that replacement may also occur due to information rent reasons if learning is about the agent's type. That is, replacement does not only occur because the agent is too likely to be of low quality but to reduce informativeness rents that he would have to be provided if he would receive a continuation contract. However, the possibility to obtain a new agent whose quality is drawn according to the initial prior makes replacement attractive as well. It increases the continuation value for the principal if the belief about the current agent's quality is low.

2.7 Conclusion

In this paper, I study the optimal contract for a two-stage project under full commitment in a dynamic moral hazard setting with ex ante symmetric information and learning within and across stages.

I show that the informativeness of the first stage gives rise to an endogenously arising ratchet effect. As a consequence, the optimal contract has to provide the agent with additional rents for good performance. Moreover, using long-term rewards with continuation contracts that condition on future performance is costly in that they amplify the ratchet effect and therefore increase the agent's information rents. This induces the composition of the total reward to change with performance: bad performance cannot be identified as either bad luck or simple shirking. Thus, the agent receives a higher share of the total reward as a bonus payment rather than a continuation contract if a success is obtained later. Good performance is rewarded more with continuation contracts because these reduce the inefficiencies in the second stage caused by procrastination rents. If the principal has the ability to replace agents after stages, she will make use of this possibility for the latest success times that still induce continuation. By replacing the agent, the principal eliminates the incentive to manipulate the performance within the replacement region and therefore reduces informativeness rents for all success times.

My analysis has several empirical implications: (i) The composition of the agent's compensation changes with performance. In particular, if performance gets worse, the total reward is lower and consists of relatively more short-term than long-term rewards. For example, a well-performing CEO is rewarded with stock options that are tied to future performance. A CEO that performs worse is rewarded with bonus payments and less with stock options. The total worth of the reward is higher for the well-performing CEO. (ii) Deadlines are relatively more responsive to early performance while final-stage bonus payments are less responsive to early performance compared to a setting without informativeness rents. (iii) Early-stage deadlines are relatively short if there is a learning spillover to future stages. (iv) Even successful agents may be replaced if they do not perform sufficiently well although the project is continued and the agent known to be able to complete future tasks. (v) Staging occurs more frequently if the initial risk is high.

Besides studying these empirical implications, there are still open avenues for future research: First, it would be interesting to consider the case of no or only partial commitment of the principal. Second, in a setting with ex ante private information of the agent, one may wonder whether a menu of differently staged contracts can elicit the agent's superior information about the project. Third, analyzing how competition between agents and free-riding interact in the presence of the arising ratchet effect is another possible extension.

2.8 Appendix

2.8.1 Preliminaries

The following results will be used frequently throughout the analysis.

Probability that no success has occurred until t. The Poisson distribution implies that no success occurs in an interval $[0,t]$ with probability

$$e^{-\int_0^t (p_{i,s}\lambda^g + (1-p_{i,s})\lambda^b) a_{1,s} ds}.$$

Using the definition of the posterior and its law of motion, $dp_t = -p_t(1-p_t)\Delta\lambda a_{1,t}$, we can rewrite this probability. First, note that the law of the posterior can be written as

$$-p_t \lambda^g a_{1,t} = \frac{dp_{i,t}}{1-p_{i,t}} - p_{i,t}\lambda^b a_{1,t}.$$

Second, I apply this on the probability of no success:

$$\begin{aligned}
&= e^{\int_0^t \frac{dp_{i,s}}{1-p_{i,s}} - p_{i,s}\lambda^b a_{1,s} ds} e^{-\int_0^t [(1-p_{i,s})\lambda^b] a_{1,s} ds} \\
&= e^{\int_0^t \frac{dp_{i,t}}{1-p_{i,t}}} e^{-\int_0^t [(1-p_{i,s})\lambda^b + p_{i,s}\lambda^b] a_{1,s} ds} \\
&= e^{\int_0^t \frac{dp_{i,t}}{1-p_{i,t}}} e^{-\int_0^t \lambda^b a_{1,s} ds} \\
&= e^{-\ln(1-p_{i,t})} e^{-\lambda^b \int_0^t a_{1,s} ds} \\
&= \frac{1-p_{i,0}}{1-p_{i,t}} e^{-\lambda^b \int_0^t a_{1,s} ds}.
\end{aligned}$$

Using the posterior at time t, $1-p_{i,t} = \frac{e^{-\lambda^b \int_0^t a_{1,s} ds}(1-p_{i,0})}{e^{-\lambda^g \int_0^t a_{1,s} ds} p_{i,0} + e^{-\lambda^b \int_0^t a_{1,s} ds}(1-p_{i,0})}$, we get

$$= e^{-\lambda^g \int_0^t a_{1,s} ds} p_{i,0} + e^{-\lambda^b \int_0^t a_{1,s} ds}(1-p_{i,0})$$

This allows me to rewrite the agent's objective functions as follows

$$\begin{aligned}
&\int_0^\infty e^{-rt} e^{-\int_0^t (p_{i,s}\lambda^g + (1-p_{i,s})\lambda^b) a_{1,s} ds} \cdot a_{1,t} \left((p_{i,t}\lambda^g + (1-p_{i,t})\lambda^b) v_i(t) - c \right) dt \\
&= \int_0^\infty e^{-rt} a_{1,t} \left(e^{-\lambda^g \int_0^t a_{1,s} ds} p_{i,0} + e^{-\lambda^b \int_0^t a_{1,s} ds}(1-p_{i,0}) \cdot ((p_t\lambda^g + (1-p_t)\lambda^b) v_i(t) - c) \right) dt \\
&= \int_0^\infty e^{-rt} a_{1,t} \left(\left(p_{i,0}\lambda^g e^{-\int_0^t a_{1,s}\lambda^g ds} + (1-p_{i,0})\lambda^b e^{-\int_0^t \lambda^b a_{1,s} ds} \right) v_i(t) \right. \\
&\left. - \left(e^{-\lambda^g \int_0^t a_{1,s} ds} p_{i,0} + e^{-\lambda^b \int_0^t a_{1,s} ds}(1-p_{i,0}) \right) c \right) dt
\end{aligned}$$

where $v_i(t)$ is the value of succeeding in stage i at time t. For the principal, the value is given by

$$\int_0^\infty e^{-rt}\left(p_{i,0}\lambda^g e^{-\int_0^t a_{1,s}\lambda^g ds} + (1-p_{i,0})\lambda^b e^{-\int_0^t \lambda^b a_{1,s} ds}\right)\pi_i(t)dt$$

where $\pi_i(t)$ is the principal's value of a success in stage i at time t.

Posteriors and Odds Ratios. The belief in the second stage depends on the success time in the first stage and parameters of the model. The initial belief in terms of primitives and success time is given by

$$p_{2,0}(\tau_1) = \frac{\lambda^g p_0 e^{-\int_0^{\tau_1}\lambda^g a_{1,s}ds}}{\lambda^g p_0 e^{-\int_0^{\tau_1}\lambda^g a_{1,s}ds} + \lambda^b(1-p_0)e^{-\int_0^{\tau_1}\lambda^b a_{1,s}ds}}$$

and the posterior after an experimentation duration of t in the second stage

$$p_{2,0}(\tau_1,t) = \frac{\lambda^g p_0 e^{-\int_0^{\tau_1}\lambda^g a_{1,s}ds-\int_0^t \lambda a_{1,s}ds}}{\lambda^g p_0 e^{-\int_0^{\tau_1}\lambda^g a_{1,s}ds-\int_0^t \lambda a_{1,s}ds} + \lambda^b(1-p_0)e^{-\int_0^{\tau_1}\lambda^b a_{1,s}ds}}.$$

The odds ratio is then given by

$$\frac{p_{2,0}(\tau_1,t)}{1-p_{2,0}(\tau_1,t)} = \frac{p_0}{1-p_0}\frac{\lambda^g}{\lambda^b}e^{-\int_{\tau_1}^{\tau_1+t}\lambda a_{1,s}ds-(\lambda^g-\lambda^b)\int_0^{\tau_1}a_{1,s}ds}.$$

Bonus contracts are without loss of generality. The same argument as in Moroni (2016) yields the result. Denote the general payment process $\{w_f dt + w_l\}_{t\geq 0}$ by w. w maps histories into payments, $w:\mathcal{H}^t\to\mathbb{R}$. Consider a bonus contract b that only has payments at time zero ($\tau_0 = 0$) and breakthrough times τ_1 and τ_2. Define $w_i(\emptyset, h^{\tau_{i-1}})$ as discounted payoff that payment process w delivers to the agent given the history if the game ended without a breakthrough at h^{τ_i}. Then, let $b_0 = w_1(\emptyset, h^0)$ and $b_{\tau_1}(h^{\tau_1}) = e^{r\tau_1}\left(w_{i+1}(\emptyset, h^{\tau_i}) - w_i(\emptyset, h^{\tau_{i-1}})\right)$. This is a bonus contracts giving the same expected payoff after every history to the agent as the initial contract w. Limited liability is satisfied in the bonus contract as well if no positive payments are made if no breakthrough is obtained in a stage. Such a payment rule is clearly suboptimal for the principal.

2.8.2 Proofs

If it does not cause confusion, I drop stage indices in the proofs to simplify notation.

Proof of Lemma 2.1. [12] The proof relies on Pontryagin's maximum principle. The second-stage analysis in my model resembles the one of Moroni (2016). The agent's problem is to choose an effort path $\{a_t\}$ given a contract $b_{2,t}$ to maximize expected payoffs, that is

$$\max_{\{a_t\}} \int_0^\infty e^{-rt} e^{-\int_0^t p_s \lambda a_s ds} a_t \left(p_t \lambda b_{2,t} - c\right) dt.$$

Using the definition of the posterior as well as the differential equation determining its law, this can be rewritten as

$$\max_{\{a_t\}} \int_0^\infty e^{-rt} a_t \left(p_0 e^{-\lambda \int_0^t a_s ds} \lambda b_{2,t} - (p_0 e^{-\lambda \int_0^t a_s ds} + 1 - p_0) c\right) dt.$$

Defining $A_t = \int_0^t a_s ds$, we can rewrite the maximization as optimal control problem

$$\max_{\{a_t\}} \int_0^T e^{-rt} a_t \left(p_0 e^{-\lambda A_t} \lambda b_{2,t} - c p_0 e^{-\lambda A_t} - c(1-p_0)\right) dt$$
$$\text{s. t. } \dot{A}_t = a_t.$$

The Hamiltonian and the costate law are given by

$$\mathcal{H} = e^{-rt} \left(p_0 e^{-\lambda A_t} \lambda b_{2,t} - c p_0 e^{-\lambda A_t} - c(1-p_0)\right) a_t + \eta_t a_t$$
$$\dot{\eta}_t = e^{-rt} p_0 (b_{2,t} \lambda - c) e^{-\lambda A_t} \lambda a_t.$$

Note that the objective is linear in a_t is binary. Let

$$\gamma_t \equiv e^{-rt} \left(p_0 e^{-\lambda A_t} \lambda b_{2,t} - c p_0 e^{-\lambda A_t} - c(1-p_0)\right) + \eta_t,$$

then if $\gamma_t > 0$, the agent will exert effort, $a_t = 1$, and if $\gamma_t < 0$, he will exert no effort $a_t = 0$. If the principal wants to induce effort, she will choose $\gamma_t = 0$ at which the agent is indifferent between working and shirking. This is optimal, because whenever $\gamma_t > 0$, the principal can increase her payoff by slightly reducing $b_{2,t}$ without altering the agent's incentives.
The standard boundary condition gives $\eta_T = 0$ implying $\gamma_T = e^{-rT} \left(p_0 e^{-\lambda A_t} \lambda b_{2,T} - c p_0 e^{-\lambda A_t} - c(1-p_0)\right)$. Choosing $\gamma_t = 0$ implies

$$\eta_t = -e^{-rt} \left(p_0 e^{-\lambda A_t} (\lambda b_{2,t} - c) - c(1-p_0)\right).$$

Differentiating this with respect to time and equating it with (OBJ) delivers

$$e^{-rt} p_0 (b_{2,t} \lambda - c) e^{-\lambda A_t} \lambda a_t = e^{-rt} p_0 (b_{2,t} \lambda - c) e^{-\lambda A_t} \lambda a_t - e^{-rt} p_0 e^{-\lambda A_t} \dot{b}_{2,t} \lambda$$
$$+ r e^{-rt} \left(p_0 e^{-\lambda A_t} (\lambda b_{2,t} - c) - c(1-p_0)\right)$$

[12] The existence and sufficiency results to the optimal control problem analyzed for the second stage in this paper follow directly from Moroni (2016) and the references therein.

and hence, we may conclude that

$$\dot{b}_{2,t} = r\left(b_{2,t} - c\left(1 + \frac{1-p_0}{p_0}\frac{e^{\lambda A_t}}{\lambda}\right)\right)$$

together with the boundary condition

$$b_{2,T} = c\left(1 + \frac{1-p_0}{p_0}\frac{e^{\lambda A_T}}{\lambda}\right)$$

induces effort path $\{a_t\}$ up to time T.
In the limit $r \to 0$, the bonus payment is constant over time $\dot{b}_{2,t} = 0$ and pinned down by the static moral hazard constraint at the deadline.

Proof of Proposition 2.1. First, consider the second-best second-stage contract without a promise-keeping constraint. Recall that incentive compatibility requires $\dot{b}_{2,t}$ follows from Lemma 2.1. The principal wants to induce full effort which follows from Moroni (2016) in the second stage. The bonus payment can be integrated to

$$b_{2,t} = \frac{c}{p_{2,0}\lambda(\lambda - r)}\left(\lambda(p_{2,0} + e^{-r(T-t)+\lambda T}(1-p_{2,0})) - r(p_{2,0} + (1-p_{2,0})e^{-r(T-t)+\lambda T})\right).$$

Hence, the principal chooses a total effort A_T equal to calendar time T. The objective of the principal is

$$\begin{aligned}&\max_T \int_0^T e^{-rt}e^{-\lambda A_t}p_0\lambda(\pi - b_{2,t})dt\\ &\text{s. t. } b_{2,t} = \frac{c}{p_{2,0}\lambda(\lambda - r)}\left(\lambda(p_{2,0} + e^{-r(T-t)+\lambda T}(1-p_{2,0})) - r(p_{2,0} + (1-p_{2,0})e^{-r(T-t)+\lambda T})\right).\end{aligned}$$

Applying the constraint and full effort in the objective delivers

$$\begin{aligned}\max_T \int_0^T e^{-rt}e^{-\lambda t}p_{2,0}\lambda\Bigg(\pi - \frac{c}{p_{2,0}\lambda(\lambda - r)}\Big(&\lambda(p_{2,0} + e^{-r(T-t)+\lambda T}(1-p_{2,0}))\\ &-r(p_{2,0} + (1-p_{2,0})e^{-r(T-t)+\lambda T})\Big)\Bigg)dt\end{aligned}$$

This simplifies to

$$\max_T \frac{1}{r-\lambda}\left(-c(1-p_{2,0})((1+e^{(\lambda-r)T})) + \frac{(\pi\lambda - c)p_{2,0}(r-\lambda)(1-e^{-(r+\lambda)T})}{r+\lambda}\right)$$

and taking the first-order condition delivers

$$T = \frac{1}{2\lambda}\ln\left(\frac{p_{2,0}}{1-p_{2,0}}\frac{\pi - c}{c}\right).$$

Consider the case of the promise-keeping constraint. The promise-keeping constraint is given by

$$v(\tau_1, T_2(\tau_1)) \geq v(\tau_1)$$

where $v(\tau_1)$ is the value promised to the agent. Note that the agent's second-stage value can be written as

$$\frac{ce^{-rT}(1-p_{2,0})}{r(r-\lambda)}\Big(r(e^{\lambda T}-1)+\lambda(1-e^{rT})\Big)$$

Again Lemma 2.1 pins down the incentive compatibility condition. With promise-keeping constraint the principal maximizes

$$\text{(OBJ)} \qquad \max_{T_2, b_{2,t}} \int_0^{T_2} e^{-rt} e^{-\int_0^t p_{2,s}(A_{1,\tau_1})\lambda a_s ds} p_{2,t}(A_{1,\tau_1})\lambda a_t(\pi - b_{2,t})dt.$$

$$\text{(IC)} \qquad s.t. \quad a_t \in \arg\max_{a_t \in \{0,1\}} \int_0^{T} e^{-rt} e^{-\int_0^t p_{2,s}(A_{1,\tau_1})\lambda a_s ds} a_t(p_{2,t}(A_{1,\tau_1})\lambda b_{2,t} - c)dt$$

$$\text{(PK)} \qquad \max_{a_t \in [0,1]} \int_0^{T_2} e^{-rt} e^{-\int_0^t p_{2,s}(A_{1,\tau_1})\lambda a_s ds} a_t(p_{2,t}(A_{1,\tau_1})\lambda b_{2,t} - c)dt \geq v(\tau_1).$$

Applying Lemma 2.1 and assuming frontloading of effort, this reduces to

$$\max_{T_2(\tau_1)} \frac{1}{r-\lambda}\left(-c(1-p_{2,0})((1+e^{(\lambda-r)T_2(\tau_1)})) + \frac{(\pi\lambda - c)p_{2,0}(r-\lambda)(1-e^{-(r+\lambda)T_2(\tau_1)})}{r+\lambda}\right)$$

$$s.t. \quad v(\tau_1, T_2(\tau_1)) = \frac{ce^{-rT_2(\tau_1)}(1-p_{2,0})}{r(r-\lambda)}\Big(r(e^{\lambda T_2(\tau_1)}-1)+\lambda(1-e^{rT_2(\tau_1)})\Big) \geq v(\tau_1).$$

Which yields as Lagrangean

$$\begin{aligned}\mathcal{L} = &\frac{1}{r-\lambda}\left(-c(1-p_{2,0})((1+e^{(\lambda-r)T_2(\tau_1)})) + \frac{(\pi\lambda - c)p_{2,0}(r-\lambda)(1-e^{-(r+\lambda)T_2(\tau_1)})}{r+\lambda}\right)\\ &- \mu\left(\frac{ce^{-rT_2(\tau_1)}(1-p_{2,0})}{r(r-\lambda)}\Big(r(e^{\lambda T_2(\tau_1)}-1)+\lambda(1-e^{rT_2(\tau_1)})\Big) - v(\tau_1)\right)\end{aligned}$$

that can be solved for $T_2(A_{1,\tau_1})$ by the Kuhn-Tucker Theorem. This is solved by the second-best bonus, whenever $v(\tau_1, T_2(\tau_1)) > v(\tau_1)$ as then $\mu = 0$ and we are in the case of the second best. If the constraint is binding, we require $T_2(\tau_1)$ to be chosen such that

$$\frac{ce^{-rT_2(\tau_1)}(1-p_{2,0})}{r(r-\lambda)}\Big(r(e^{\lambda T_2(\tau_1)}-1)+\lambda(1-e^{rT_2(\tau_1)})\Big) = v(\tau_1).$$

If $r \to 0$, this is solved by

$$T_2(\tau_1, v(\tau_1)) = -\frac{v(\tau_1)}{c(1-p_{2,0}(\tau_1))} - \frac{1}{\lambda}\left(1 + W_{-1}\left(-e^{-1-\lambda\frac{v(\tau_1)}{c(1-p_{2,0}(\tau_1))}}\right)\right)$$

where W_{-1} denotes the negative branch of the Lambert-W-function.

Proof of Proposition 2.2. The agent's value, $w_t \equiv b_{1,t} + v_t$, consists of a promised utility from the second stage, v_t, and a bonus payment after the first success, $b_{1,t}$. Note that the continuation value of the agent depends on the agent's private information. The principal promises the continuation value conditional on the expected exerted effort, $\hat{A}_{1,t}$. Conditional on this, she implements a bonus and a deadline in the second stage. However, the true total effort that the agent has exerted is private information, $A_{1,t} = \int_0^t a_s ds$. The promised utility is given by $v(t)$ and implemented through the continuation contract in Proposition 2.1. The value from the agent's view is given by $v(t, A_{1,t})$ and depends on the true effort because she may hold a different belief than the principal in the second stage. The Hamiltonian of the agent is given by

(2.7)
$$\mathcal{H} = a_t e^{-rt}\left((p_0\lambda^g e^{-A_t\lambda^g} + (1-p_0)\lambda^b e^{-\lambda^b A_t})(b_{1,t} + v(t, A_t)) - (p_0 e^{-\lambda^g A_t} + (1-p_0)e^{-\lambda^b A_t})c\right) + \eta_t a_t.$$

He will exert effort if

(2.8)
$$\gamma_t \equiv e^{-rt}\left((p_0\lambda^g e^{-A_t\lambda^g} + (1-p_0)\lambda^b e^{-\lambda^b A_t})(b_{1,t} + v(t, A_t)) - (p_0 e^{-\lambda^g A_t} + (1-p_0)e^{-\lambda^b A_t})c\right) + \eta_t \geq 0$$

and the cheapest way to do so is $\gamma_t = 0$. Hence, if $\gamma_t = 0$

(2.9)
$$\eta_t = e^{-rt}\left((p_0\lambda^g e^{-A_t\lambda^g} + (1-p_0)\lambda^b e^{-\lambda^b})(b_{1,t} + v(t, A_t)) - (p_0 e^{-\lambda^g A_t} + (1-p_0)e^{-\lambda^b A_t})c\right).$$

The boundary condition is given by $\eta_T = 0$ which yields

$$(b_T + v(T, A_T)) = c\frac{p_0 e^{-A_T\lambda^g} + (1-p_0)e^{-A_T\lambda^b}}{p_0\lambda^g e^{-A_T\lambda^g} + (1-p_0)\lambda^b e^{-A_T\lambda^b}}. \tag{2.10}$$

The costate evolution is given by $\dot{\eta}_t = -\partial_{A_t}\mathcal{H}$:

$$\begin{aligned}\dot{\eta}_t = e^{-rt}\Big(&a_t(p_0{\lambda^g}^2 e^{-A_t\lambda^g} + (1-p_0){\lambda^b}^2 e^{-A_t\lambda^b})(b_{1,t} + v(t, A_t)) \\ &-(p_0\lambda^g e^{-\lambda^g A_t} + (1-p_0)\lambda^b e^{-\lambda^b A_t})\frac{\partial v(t, A_t)}{\partial A_t}\Big) \\ &- e^{-rt}a_t(p_0\lambda^g e^{-\lambda^g A_t} + (1-p_0)\lambda^b e^{-\lambda^b A_t})c.\end{aligned} \tag{2.11}$$

Differentiating (2.9) with respect to time delivers

$$\begin{aligned}\dot{\eta}_t &= re^{-rt}\Big((p_0\lambda^g e^{-A_t\lambda^g} + (1-p_0)\lambda^b e^{-\lambda^b})v(t,A_t) - (p_0 e^{-\lambda^g A_t} + (1-p_0)e^{-\lambda^b A_t})c\Big)\\ &+ a_t((p_0\lambda^{g^2} e^{-A_t\lambda^g} + (1-p_0)\lambda^{b^2} e^{-A_t\lambda^b})v(t,A_t) - (p_0\lambda^g e^{-\lambda^g A_t} + (1-p_0)\lambda^b e^{-\lambda^b A_t})c\\ &- (p_0\lambda^g e^{-\lambda^g A_t} + (1-p_0)\lambda^b e^{-\lambda^b A_t})\left(\dot{b}_t + \dot{v}(t,A_t)\right))\end{aligned} \tag{2.12}$$

or with γ_t not fixed

(2.13)

$$\begin{aligned}\dot{\eta}_t &= \dot{\gamma}_t + re^{-rt}\Big((p_0\lambda^g e^{-A_t\lambda^g} + (1-p_0)\lambda^b e^{-\lambda^b})v(t,A_t) - (p_0 e^{-\lambda^g A_t} + (1-p_0)e^{-\lambda^b A_t})c\Big)\\ &+ a_t((p_0\lambda^{g^2} e^{-A_t\lambda^g} + (1-p_0)\lambda^{b2} e^{-A_t\lambda^b})v(t,A_t) - (p_0\lambda^g e^{-\lambda^g A_t} + (1-p_0)\lambda^b e^{-\lambda^b A_t})c\\ &- (p_0\lambda^g e^{-\lambda^g A_t} + (1-p_0)\lambda^b e^{-\lambda^b A_t})\left(\dot{b}_t + \dot{v}(t,A_t)\right))\end{aligned}$$

or (2.8) Equating this with (2.11) yields

$$\begin{aligned}\dot{b}_t + \dot{v}(t,A_t) &= r\left(b_{1,t} + v(t,A_t) - \frac{p_0 e^{-\lambda^g A_t} + (1-p_0)e^{-\lambda^b A_t}}{p_0\lambda^g e^{-\lambda^g A_t} + (1-p_0)\lambda^b e^{-\lambda^b A_t}}c\right) + \frac{\partial v(t,A_t)}{\partial A_t}a_t\\ &+ \frac{\dot{\gamma}_t}{p_0\lambda^g e^{-\lambda^g A_t} + (1-p_0)\lambda^b e^{-\lambda^b A_t}}.\end{aligned} \tag{2.14}$$

Hence, if the promised utility follows (2.42) together with the boundary equation (2.10), the principal can induce the effort in the interval $[0,T]$.
If $r \to 0$, this reduces to

$$\dot{b}_t + \dot{v}(t,A_t) = \frac{\partial v(t,A_t)}{\partial A_t}a_t + \frac{\dot{\gamma}_t}{p_0\lambda^g e^{-\lambda^g A_t} + (1-p_0)\lambda^b e^{-\lambda^b A_t}}. \tag{2.15}$$

Proof of Lemma 2.2. Alternatively, the principal could provide the agent with utility by varying the boundary condition, B, for the bonus payment and differing the deadline. The payment rule, however, still has to satisfy incentive-compatibility. The bonus payment for each t under an alternative deadline is given by

$$b(t,B) = Be^{-r(T-t)} - c\left(1 - e^{-r(T-t)} + \frac{r}{\lambda - r}\frac{1-p_{2,0}}{p_{2,0}}\left(e^{\lambda t} - e^{-r(T-t)+\lambda T}\right)\right)$$

This determines the agent's value if, $p_{2,0}(\hat{A}_t)$ is the principal's and $p_{2,0}(A_t)$ the agent's belief, to be given by

$$
\begin{aligned}
&(2.16)\\
&v(T,B,p_{2,0}(A_t),p_{2,0}(\hat{A}_t))\\
&=\frac{1}{p_{2,0}(\hat{A}_t)r(r-\lambda)(r+\lambda)}\\
&\cdot\Big((\lambda+1)(\lambda-r)rp_{2,0}(\hat{A}_t)p_{2,0}(A_t)+e^{-rT}p_{2,0}(A_t)r(ce^{\lambda T}(p_{2,0}(\hat{A}_t)-1)r+(B+c)p_{2,0}(\hat{A}_t)(r-\lambda))(\lambda+r)\\
&+(\lambda+r)(c(p_{2,0}(A_t)-1)p_{2,0}(\hat{A}_t)r)+c(p_{2,0}(A_t)r-p_{2,0}(\hat{A}_t)(-1+p_{2,0}(A_t)+rp_{2,0}(A_t)))\lambda\\
&-(r-\lambda)e^{-(r+\lambda)T}(ce^{\lambda T}((p_{2,0}(A_t)-1)p_{2,0}(\hat{A}_t)+p_{2,0}(A_t)(p_{2,0}(\hat{A}_t)-1)r)(r+\lambda))\\
&-(r-\lambda)e^{-(r+\lambda)T}(p_{2,0}(A_t)p_{2,0}(\hat{A}_t)r(c(r-1)+B(\lambda+r)))\Big).
\end{aligned}
$$

This delivers as deviation incentive in the first stage

$$
\begin{aligned}
&(2.17)\\
&\frac{\partial v(T,B,p_{2,0}(A_t),p_{2,0}(\hat{A}_t))}{\partial A_t}\\
&=\frac{\partial p_{2,0}(A_t)}{\partial A_t}\frac{1}{p_{2,0}(\hat{A}_t)r(r-\lambda)(r+\lambda)}\\
&\Big((1+\lambda)cp_{2,0}(\hat{A}_t)r(\lambda-r)+re^{-rT}(ce^{\lambda T}(p_{2,0}(\hat{A}_t)-1)r+(B+c)p_{2,0}(\hat{A}_t)(r-\lambda))(r+\lambda)\\
&(r+\lambda)(cp_{2,0}(\hat{A}_t)r+c(r-p_{2,0}(\hat{A}_t)(1+r))\lambda)\\
&(r-\lambda)e^{-(r+\lambda)T}(r-\lambda)(cp_{2,0}(\hat{A}_t)(r-1)r+Bp_{2,0}(\hat{A}_t)r(r+\lambda)+ce^{\lambda T}(p_{2,0}(\hat{A}_t)+(p_{2,0}(\hat{A}_t)-1)r)(r+\lambda))\Big).
\end{aligned}
$$

To see how this varies with the composition of the continuation contract note that it follows from the implicit function theorem applied on the promise-keeping condition that the deadline and the bonus payment vary according to

$$
\frac{dT}{dB}=\frac{(1-e^{\lambda T})p_{2,0}(\hat{A}_t)}{(B-c)p_{2,0}(\hat{A}_t)(r+\lambda)+e^{\lambda T}(-c(1-p_{2,0}(\hat{A}_t))(\lambda+re^{\lambda T})+r(p_{2,0}(\hat{A}_t)B-c))}<0. \tag{2.18}
$$

That is, as intuitive, if the boundary condition increases, the deadline decreases. Considering now the effect of the deadline on the deviation incentive I find that this is affected as follows

$$
\frac{\partial p_{2,0}(A_t)}{\partial A_t}\frac{e^{-(r+\lambda)T}}{p_{2,0}(\hat{A}_t)}\Big(re^{2\lambda T}(1-p_{2,0}(\hat{A}_t))c+e^{\lambda T}(cp_{2,0}(\hat{A}_t)-(c+Bp_{2,0}(\hat{A}_t))r) \tag{2.19}
$$

$$
+p_{2,0}(\hat{A}_t)(c(r-1)+B(\lambda+r)))\Big) \tag{2.20}
$$

which is positive. Hence, by extending the deadline, the principal reduces the incentive to deviate in the first stage.

Evolution of Agent's Continuation Utility. The agent's utility from the second stage may evolve different than the $v(t)$ because it depends on his private information about the true effort. The promised utility to the agent is denoted by $v(t)$ which coincides with the agent's continuation utility if he is on path, i.e., if $A_{1,t} = t$. However, if the agent has deviated, $v(t) \neq v(t, A_{1,t})$ where the latter denotes the agent's continuation utility given his private information $A_{1,t}$. We know that the on-path utility evolves according to

$$\dot{w}(t, A_{1,t}) = r\left(b_{1,t} + v(t, A_{1,t}) - \frac{p_0 e^{-\lambda^g A_{1,t}} + (1-p_0)e^{-\lambda^b A_{1,t}}}{p_0\lambda^g e^{-\lambda^g A_{1,t}} + (1-p_0)\lambda^b e^{-\lambda^b A_{1,t}}}c\right) + \frac{\partial v(t, A_{1,t})}{\partial A_{1,t}}a_t. \tag{2.21}$$

For an agent that has exerted effort $\hat{A}_{1,t}$, the value of succeeding the first stage at t is given by (net the bonus payment, $b_{1,t}$)

$$\begin{aligned} v(\hat{A}_{1,t}, A_{1,t}) = &\frac{ce^{-rT}}{r(r-\lambda)p_{2,0}(A_{1,t})} \\ &\cdot \Big(p_{2,0}(\hat{A}_{1,t})(re^{rT} - \lambda p_{2,0}(A_{1,t})e^{rT} - r(1-p_{2,0}(A_{1,t}))e^{\lambda T}) \\ &+(1-p_{2,0}(\hat{A}_{1,t}))p_{2,0}(A_{1,t})(r-\lambda) + p_{2,0}(A_{1,t})e^{rT}(\lambda - r)\Big) \end{aligned}$$

where $p_{2,0}(A_{1,t})$ is the principal's belief at the beginning of the second stage and $p_{2,0}(\hat{A}_{1,t})$ is the agent's belief. This can be simplified substantially to

$$v(\hat{A}_{1,t}, A_{1,t}) = v(t)\frac{p_{2,0}(\hat{A}_{1,t})}{p_{2,0}(A_{1,t})} - cr(1 - e^{rT_2(A_{1,t})})\frac{p_{2,0}(\hat{A}_{1,t}) - p_{2,0}(A_{1,t})}{p_{2,0}(A_{1,t})}.$$

The total value is evolving according to

$$\dot{w}(\hat{A}_{1,t}, A_{1,t}) = \dot{b}_t + \dot{v}(\hat{A}_{1,t}, A_{1,t}). \tag{2.22}$$

Hence, if $r \to 0$

$$\dot{v}(\hat{A}_{1,t}, A_{1,t}) = \dot{v}(t)\frac{p_{2,0}(\hat{A}_{1,t})}{p_{2,0}(A_{1,t})} + v(t)\left(\frac{\dot{p}_{2,0}(\hat{A}_{1,t})}{p_{2,0}(A_{1,t})} - \frac{p_{2,0}(\hat{A}_{1,t})}{p_{2,0}(A_{1,t})}\frac{\dot{p}_{2,0}(A_{1,t})}{p_{2,0}(A_{1,t})}\right) \tag{2.23}$$

Agent's Problem. Given this law of the agent's continuation value, the agent solves the following problem

$$\max_a \int_0^\infty e^{-\int_0^t e^{-rt}(p_s\lambda_1^g + (1-p_s)\lambda_1^b)a_s ds} a_t \left((p_t\lambda_1^g + (1-p_t)\lambda_1^b)v(t, A_{1,t}) - c\right) dt \tag{2.24}$$

$$s.t. \quad \dot{w}(t, A_{1,t}) = \dot{b}_t + \dot{v}(t, A_{1,t}) \tag{2.25}$$

$$\dot{w}(t) = r\left(b_{1,t} + v(t, A_{1,t}) - \frac{p_0 e^{-\lambda^g A_{1,t}} + (1-p_0)e^{-\lambda^b A_{1,t}}}{p_0\lambda^g e^{-\lambda^g A_{1,t}} + (1-p_0)\lambda^b e^{-\lambda^b A_{1,t}}}c\right) + \frac{\partial v(t, A_{1,t})}{\partial A_{1,t}}a_t \tag{2.26}$$

$$\dot{A}_{1,t} = a_t. \tag{2.27}$$

Existence of Solution to Agent's Problem Existence follows from Clarke (2013), Theorem 23.11. The theorem applies as:

- the laws of motion of the state variables, A_t, $w(t, A_t)$ and $w(t)$ are measurable in t and continuous in A_t
- the control set $a_t \in [0, 1]$ is closed and convex
- the running cost is
  - Lesbesgue measurable in t and (A, a)
  - lower semicontinuous in (A, a)
  - convex in a for any (t, A)
- the effort path $a_t = 0$ for all t and $A_t = 0$ for all t is admissible and delivers a finite value.

Principal's Problem. The Hamiltonian of the principal's problem is given by

(2.28)

$$\mathcal{H} = e^{-rt}\left(p_0\lambda^g e^{-\lambda^g A_t} + (1-p_0)\lambda^b e^{-\lambda^b A_t}\right)(\Pi(T_2(t), t) - c(T_2, t) - w_t) + \gamma_t a_t$$

(2.29)

$$+\eta_t\left(r\left(b_t + v(t, A_t) - \frac{p_0 e^{-\lambda^g A_t} + (1-p_0)e^{-\lambda^b A_t}}{p_0\lambda^g e^{-\lambda^g A_t} + (1-p_0)\lambda^b e^{-\lambda^b A_t}}c\right) + \frac{\partial v(t, A_t)}{\partial A_t}a_t\right) - \zeta_t(w_t - v(T_2(t), t))$$

(2.30)

$$\dot{A}_t = a_t$$

(2.31)

$$\dot{w}_t = r\left(b_t + v(t, A_t) - \frac{p_0 e^{-\lambda^g A_t} + (1-p_0)e^{-\lambda^b A_t}}{p_0\lambda^g e^{-\lambda^g A_t} + (1-p_0)\lambda^b e^{-\lambda^b A_t}}c\right) + \frac{\partial v(t, A_t)}{\partial A_t}a_t$$

(2.32)

$$\dot{\gamma}_t = e^{-rt}\left(p_0\lambda^{g^2} e^{-\lambda^g A_t} + (1-p_0)\lambda^{b^2} e^{-\lambda^b A_t}\right)(\Pi(T_2(t), t) - c(T_2, t) - w_t) - \eta_t\frac{\partial \dot{w}_t}{\partial A_t}$$

(2.33)

$$\dot{\eta}_t = \left(p_0\lambda^g e^{-\lambda^g A_t} + (1-p_0)\lambda^b e^{-\lambda^b A_t}\right) - \eta_t r + \zeta_t$$

with constraint from agent's problem $w_{T_1} = c\frac{p_0\lambda^g e^{-\lambda^g A_{T_1}} + (1-p_0)\lambda^b e^{-\lambda^b A_{T_1}}}{p_0 e^{-\lambda^g A_{T_1}} + (1-p_0)e^{-\lambda^b A_{T_1}}}$ with associated multiplier μ and moreover boundary conditions $\gamma_T = 0, \eta_0 = 0, \eta_T = \mu, A_0 = 0$. Note that $\Pi(T_2(t), t)$ is the total expected profit from choosing deadline $T_2(t)$ after a success at t, i.e.,

$\int_0^{T_2(t)} p_{2,0}(A_t)e^{-\lambda A_s}\pi ds$ and $c(T_2(t),t)$ is the total expected experimentation cost from choosing deadline $T_2(t)$ after a success at t, i.e., $\int_0^{T_2(t)} p_{2,0}(A_t)e^{-\lambda A_s}cds$.

Maximization with respect to a_t. To see that the principal wants to implement full effort consider a dynamic programming heuristic

$$\begin{aligned}
&\left(-\frac{\partial\Pi_t}{\partial a_t}+\frac{\partial\Pi_t}{\partial a_{t+dt}}\right)\\
&= dt\left((p_t\lambda^g+(1-p_t)\lambda^b)\left(-\Pi(A_t)-a_t dt\Pi'(A_t)+\Pi(A_{t+dt})\right)\right)\\
&+dt^2\left(\Pi_{t+2dt}\left(\underbrace{(a_t-a_{t+dt})}_{=0,\text{ if } a_t \text{ continuous}}\;((1-p_t)^2\lambda^{b^2}+2(1-p_t)p_t\lambda^b\lambda^g+p_t\lambda^{g^2})+(1-p_t)\lambda^{b^2}+p_t\lambda^{g^2}\right)\right)\\
&-dt^2\left(c\underbrace{\left(((1-p_t)\lambda^b+p_t\lambda^g)-((1-p_t)\lambda^b+p_t\lambda^g)\right)}_{=0}\right)\\
&+dt^2\left(\underbrace{\frac{\partial\Pi(A_{t+dt})}{\partial a_t}-\frac{\partial\Pi(A_{t+dt})}{\partial a_{t+dt}}}_{=0}\right)a_{t+dt}\left(a_t((1-p_t)\lambda^b+p_t\lambda^g)+\frac{1}{2}(1-p_t)^2\lambda^{b^2}p_t(1-p_t)\lambda^g\lambda^b+\frac{1}{2}p_t^2\lambda^{g^2}\right)\\
&-dt^2\;\underbrace{(a_t-a_{t+dt})}_{=0 \text{ if } a_t \text{ continuous}}\;\left((1-p_t)\lambda^{b^2}+p_t\lambda^{g^2}\right)\Pi(A_{t+dt})\\
&+dt^2\left((1-p_t)^2\lambda^{b^2}+2(1-p_t)p_t\lambda^b\lambda^g+p_t\lambda^{g^2})+(1-p_t)\lambda^{b^2}+p_t\lambda^{g^2}\right)a_t\\
&\left(\Pi(A_t)+\frac{1}{2}a_t dt\Pi'(a_t)-\underbrace{\frac{a_{t+dt}}{a_t}}_{=1 \text{ if } a_t \text{ continuous}}\Pi(A_{t+dt})\right)\\
&-dt^2\left(r\left((p_t\lambda^g+(1-p_t)\lambda^b)\Pi(A_{t+dt})-c+a_{t+dt}(p_t\lambda^g+(1-p_t)\lambda^b)\underbrace{\left(\frac{\partial\Pi(A_{t+dt})}{\partial a_{t+dt}}-\frac{\partial\Pi(A_{t+dt})}{\partial a_{t+}}\right)}_{=0}\right)\right)\\
&+o(dt^3)
\end{aligned}$$

Assuming a_t to be continuous this simplifies after division by dt^2 to

$$\begin{aligned}
&\left(-\frac{\partial \Pi_t}{\partial a_t}+\frac{\partial \Pi_t}{\partial a_{t+dt}}\right)/dt^2\\
&=\left((p_t\lambda^g+(1-p_t)\lambda^b)\left(\frac{\Pi(A_{t+dt})-\Pi(A_t)}{dt}-a_t\Pi'(A_t)\right)\right)\\
&-r\left((p_t\lambda^g+(1-p_t)\lambda^b)\Pi(A_{t+dt})-c\right)\\
&+\left((1-p_t)^2\lambda^{b^2}+2(1-p_t)p_t\lambda^b\lambda^g+p_t\lambda^{g^2})+(1-p_t)\lambda^{b^2}+p_t\lambda^{g^2}\right)a_t\\
&\left(\Pi(A_t)+\frac{1}{2}a_t dt\Pi'(a_t)-\Pi(A_{t+dt})\right).
\end{aligned}$$

Take the limit as $dt\to 0$ and get

$$\begin{aligned}
&\left(-\frac{\partial \Pi_t}{\partial a_t}+\frac{\partial \Pi_t}{\partial a_{t+dt}}\right)/dt^2\\
&=\left((p_t\lambda^g+(1-p_t)\lambda^b)\left(\dot{\Pi}(A_t)-a_t\Pi'(A_t)\right)\right)\\
&-r\left((p_t\lambda^g+(1-p_t)\lambda^b)\Pi(A_{t+dt})-c\right)
\end{aligned}$$

Note that $\dot{\Pi}(A_t)-a_t\Pi'(A_t)$ is zero up to the second order and therefore the expression is negative as long as t is less than the first-best. Thus, welfare increases if effort is frontloaded. Moreover, note that the expected payment to the agent is decreasing if effort is frontloaded. If discounting dominates learning in a way that the expected payment to the agent is increasing for some measure of time an argument similar to the one in Moroni (2016) delivers optimality of frontloading. Hence, the principal prefers frontloading of effort, that ist $a_t=1$ for $t\in[0,T_1]$.

Maximization with respect to $T_2(t)$. Note that whenever $w_t-v(T_2(t),t)>0$ a bonus is paid to the agent and $\zeta_t=0$. ζ_t follows from the first-order condition with respect to time as

$$\zeta_t=-\frac{\partial\mathcal{H}}{\partial T_2(t)}/\frac{\partial v(T_2(t),t)}{\partial T_2(t)} \tag{2.34}$$

$$=-\frac{e^{-rt}\left(p_0\lambda^g e^{-\lambda^g A_t}+(1-p_0)\lambda^b e^{-\lambda^b A_t}\right)\left(\frac{\partial\Pi(T_2(t),t)}{\partial T_2(t)}-\frac{\partial c(T_2(t),t)}{\partial T_2(t)}\right)+\eta_t\frac{\partial \dot{w}_t(A_t)}{\partial T_2(t)}}{\frac{\partial v_t(A_t)}{\partial T_2(t)}}. \tag{2.35}$$

Using this in (2.33) delivers

(2.36)

$$\dot{\eta}_t=e^{-rt}\left(p_0\lambda^g e^{-\lambda^g A_t}+(1-p_0)\lambda^b e^{-\lambda^b A_t}\right)\left(1-\frac{\frac{\partial\Pi(T_2(t),t)}{\partial T_2(t)}-\frac{\partial c(T_2(t),t)}{\partial T_2(t)}}{\frac{\partial v_t(A_t)}{\partial T_2(t)}}\right)+\eta_t\left(\frac{\frac{\partial \dot{w}_t(A_t)}{\partial T_2(t)}}{\frac{\partial v_t(A_t)}{\partial T_2(t)}}-r\right)$$

which is a differential equation of η_t that can be integrated to

$$\eta_t = -e^{-\int_0^t r + \frac{\partial \dot{w}_s}{\partial T_2(s)} ds} \int_0^t e^{\int_0^s \frac{\partial \dot{w}_\tau}{\partial T_2(\tau)} d\tau} \left(p_0 \lambda^g e^{-\lambda^g A_s} + (1-p_0)\lambda^b e^{-\lambda^b A_s}\right) \tag{2.37}$$

$$\frac{\frac{\partial \Pi(T_2(t),t)}{\partial T_2(t)} - \frac{\partial c(T_2(t),t)}{\partial T_2(t)} - \frac{\partial v_t(A_t)}{\partial T_2(t)}}{\frac{\partial v_t(A_t)}{\partial T_2(t)}} ds \tag{2.38}$$

using the boundary condition $\eta_0 = 0$. Note that η_t is increasing over time as $T_2(t) > T_2^{SB}(t)$ by optimality and costly incentives. Continuity of η_t implies that there is a $\hat{t}$ such that $\zeta_t > 0$ for all $t \in [0, \hat{t}]$ because if $\zeta_t = 0$, we get from (2.34)

$$e^{-rt}\left(p_0 \lambda^g e^{-\lambda^g A_t} + (1-p_0)\lambda^b e^{-\lambda^b A_t}\right)\left(\frac{\partial \Pi(T_2(t),t)}{\partial T_2(t)} - \frac{\partial c(T_2(t),t)}{\partial T_2(t)}\right) = \eta_t \frac{\partial \dot{w}_t(A_t)}{\partial T_2(t)} \tag{2.39}$$

which cannot be satisfied in a neighborhood of $t = 0$ if the principal is optimizing. To see why, note that the left-hand side is the first-order condition for the social planner. To induce effort in that solution, the bonus has to be equal to 1 and the principal makes zero profits. It is easy to construct a contract that induces positive profit for the principal. Thus, we have a contradiction and $\zeta_t > 0$ for $t \in [0, \hat{t}]$. Moreover, the left-hand side is decreasing

Moreover, it can be seen that as soon as a positive bonus payment is used, that is, when $\zeta_t = 0$, the bonus of the reward that is given to the agent with a bonus payment is increasing over time because η_t is increasing over time.

$$\eta_t = \frac{e^{-rt}\left(p_0 \lambda^g e^{-\lambda^g A_t} + (1-p_0)\lambda^b e^{-\lambda^b A_t}\right)\left(\frac{\partial \Pi(T_2(t),t)}{\partial T_2(t)} - \frac{\partial c(T_2(t),t)}{\partial T_2(t)}\right)}{\frac{\partial \dot{w}_t(A_t)}{\partial T_2(t)}}. \tag{2.40}$$

Because η_t measures the marginal cost of changing the state v_t (note that only changing v_t affects the evolution of w_t) it follows that this is increasing over time and hence, for any level w_t of current promised value of succeeding, the principal provides more of this utility through a bonus payment $w_t - v_t(T_2)$.

To see that $\hat{t} < T_1$, equate the condition for $\zeta_{\hat{t}} = 0$ with the equation for η_t which yields

$$\begin{aligned} &- e^{-\int_0^t r + \frac{\partial \dot{w}_s}{\partial T_2(s)} ds} \int_0^t e^{\int_0^s \frac{\partial \dot{w}_\tau}{\partial T_2(\tau)} d\tau} \left(p_0 \lambda^g e^{-\lambda^g A_s} + (1-p_0)\lambda^b e^{-\lambda^b A_s}\right) \frac{\frac{\partial \Pi(T_2(t),t)}{\partial T_2(t)} - \frac{\partial c(T_2(t),t)}{\partial T_2(t)} - \frac{\partial v_t(A_t)}{\partial T_2(t)}}{\frac{\partial v_t(A_t)}{\partial T_2(t)}} ds \\ &= \frac{e^{-rt}\left(p_0 \lambda^g e^{-\lambda^g A_t} + (1-p_0)\lambda^b e^{-\lambda^b A_t}\right)\left(\frac{\partial \Pi(T_2(t),t)}{\partial T_2(t)} - \frac{\partial c(T_2(t),t)}{\partial T_2(t)}\right)}{\frac{\partial \dot{w}_t(A_t)}{\partial T_2(t)}} \end{aligned} \tag{2.41}$$

Note that this is equivalent to a first-order condition for the first-stage deadline if $\frac{\frac{\partial \Pi(T_2(t),t)}{\partial T_2(t)} - \frac{\partial c(T_2(t),t)}{\partial T_2(t)} - \frac{\partial v_t(A_t)}{\partial T_2(t)}}{\frac{\partial v_t(A_t)}{\partial T_2(t)}}$ is replaced by $-\left(\frac{\partial \Pi(T_2(t),t)}{\partial T_2(t)} - \frac{\partial c(T_2(t),t)}{\partial T_2(t)}\right)$ and bonus transfers could not be used. However, the latter can be shown to be greater than the former and as the other term is increasing in t, $\hat{t}$ less than the deadline if no bonus payment is used. By giving the principal the additional possibility of a bonus payment she is at least weakly better off and hence, $\hat{t} < T_1$.

Sufficiency of the Necessary Conditions Because we have established the existence of a solution previously, we can conclude that if $\{a_t\}$ is the only effort path that satisfies the necessary conditions, these are also sufficient. Recall that necessity requires an effort path $\{\hat{a}_t\}$ together with a costate $\hat{\gamma}_t$ such that they satisfy (2.8). Recall that we have from (2.15)

(2.42)
$$\frac{\dot{\gamma}_t}{p_0\lambda^g e^{-\lambda^g A_t} + (1-p_0)\lambda^b e^{-\lambda^b A_t}} = -r\left(b_t + v(t, A_t) - \frac{p_0 e^{-\lambda^g A_t} + (1-p_0)e^{-\lambda^b A_t}}{p_0\lambda^g e^{-\lambda^g A_t} + (1-p_0)\lambda^b e^{-\lambda^b A_t}}c\right) + \dot{b}_t + \dot{v}(t, A_t) - \frac{\partial v(t, A_t)}{\partial A_t}a_t$$

which we can, using the on-path and off-path values, rewrite as

$$\begin{aligned} &\frac{\dot{\hat{\gamma}}_t}{p_0\lambda^g e^{-\lambda^g \hat{A}_t} + (1-p_0)\lambda^b e^{-\lambda^b \hat{A}_t}} \\ &= r\left(\frac{p_0 e^{-\lambda^g A_t} + (1-p_0)e^{-\lambda^b A_t}}{p_0\lambda^g e^{-\lambda^g A_t} + (1-p_0)\lambda^b e^{-\lambda^b A_t}} - \frac{p_0 e^{-\lambda^g \hat{A}_t} + (1-p_0)e^{-\lambda^b \hat{A}_t}}{p_0\lambda^g e^{-\lambda^g \hat{A}_t} + (1-p_0)\lambda^b e^{-\lambda^b \hat{A}_t}}\right)c \\ &+ \left(\dot{v}_t(\hat{A}_t) - \dot{v}_t(A_t)\right) + \frac{\partial v(t, A_t)}{\partial A_t}a_t - \frac{\partial v(t, \hat{A}_t)}{\partial \hat{A}_t}\hat{a}_t \end{aligned} \tag{2.43}$$

Define $\tau_0 \equiv \inf\{t|\hat{\gamma}_t \neq 0\}$. Suppose $\tau_0 = 0$ and $\hat{\gamma}_{\tau_0} > 0$. By continuity of $\hat{\gamma}_t$, there is an ε such that for $\hat{\gamma}_t < 0$ for $t \in (0, \varepsilon)$. By optimality, we know that $\hat{a}_t = 0$ for $t \in (0, \varepsilon)$. This implies that $\hat{A}_t \leq A_t$ where $\hat{A}_t$ corresponds to the effort of the hypothetical effort path $\{\hat{a}_t\}$ and A_t to the on-path effort path $\{a_t\}$. I want to show that $\dot{\hat{\gamma}}_t < 0$ if $\hat{A}_t \leq A_t$ for r close to zero. Recall (2.23), then, (2.43) further reduces to

(2.44)
$$\frac{\dot{\hat{\gamma}}_t}{p_0\lambda^g e^{-\lambda^g \hat{A}_t} + (1-p_0)\lambda^b e^{-\lambda^b \hat{A}_t}} = \dot{v}(t)\left(\frac{p_{2,0}(\hat{A}_{1,t})}{p_{2,0}(A_{1,t})} - 1\right) + v(t)\left(\frac{\dot{p}_{2,0}(\hat{A}_{1,t})}{p_{2,0}(A_{1,t})} - \frac{p_{2,0}(\hat{A}_{1,t})}{p_{2,0}(A_{1,t})}\frac{\dot{p}_{2,0}(A_{1,t})}{p_{2,0}(A_{1,t})}\right) + \frac{\partial v_t(A_t)}{\partial A_t}a_t - \frac{\partial v_t(\hat{A}_t)}{\partial \hat{A}_t}\hat{a}_t$$

The right-hand side is now less than zero as $\hat{A}_{1,t} \leq A_{1,t}$ implies $p_{2,0}(\hat{A}_{1,t}) \geq p_{2,0}(A_{1,t})$. Hence, we know that $\dot{\hat{\gamma}}_t < 0$ on $t \in (0, \varepsilon)$ implying that $\hat{\gamma}_t < 0$. Together this implies that $\hat{\gamma}_t < 0$ for all

$t \in [0, T_1]$. Recall that the transversality condition implies that

$$\begin{aligned}
\hat{\gamma}_T &= \left(p_0\lambda^g e^{-\lambda^g \hat{A}_T} + (1-p_0)\lambda^b e^{-\lambda^b \hat{A}_T}\right) v(t, \hat{A}_t) - \left(p_0 e^{-\lambda^g \hat{A}_T} + (1-p_0)e^{-\lambda^b \hat{A}_T}\right) c \\
&\geq \\
\gamma_T &= \left(p_0\lambda^g e^{-\lambda^g A_T} + (1-p_0)\lambda^b e^{-\lambda^b A_T}\right) v(t, A_t) - \left(p_0 e^{-\lambda^g A_T} + (1-p_0)e^{-\lambda^b A_T}\right) c \\
&= 0
\end{aligned} \tag{2.45}$$

where the inequality follows from $\hat{A}_T \leq A_T$. $\hat{\gamma}_T \geq 0$ contradicts $\hat{\gamma}_t < 0$ for all $t \leq T_1$. An analogous argument applies for $\tau_0 > 0$. For all $t < \tau_0$, $A_t = \hat{A}_t$ and $\hat{\gamma}_t = 0$. Following τ_0 with $\hat{A}_t < A_t$, the reasoning from above yields a contradiction with the transversality condition. Note that $a_t = 1$ will be optimal and this direction suffices to guarantee sufficiency of the necessary conditions in the optimal contract.

Costly Incentives. Note that to complete the solution of the optimal control problem, we need to prove that the principal always sets $\gamma_t = 0$ in the agent's problem. This implies, that the agent's incentive constraint is never slack in the optimal contract. I restrict attention to this case, as the main contribution of the paper lies in the case when first-stage incentives are relevant. In the remaining cas es, it occurs that theprincipal wants to increase the agent's value to get closer to the second-stage second-best value. To do that, she increases γ_t above zero. These cases can occur only if the agent's promised value lies below the second-best value in some regions. The problem can analogously solved for cases with $\gamma_t > 0$. Due to a lack of closed-form solutions there is no sharp characterization of the parametric assumptions for the costly incentives case. However, a sufficient condition on contract terms is that: $\dot{w}_t < \dot{v}^{SB}(t)$ for all $t \in [0, T_1]$ and $w_{T_1} \geq v^{SB}(T_1)$. This implies that at the first-stage deadline the value in the contract is higher than the second-best value of the contract. Moreover, the total reward is decreasing steeper than the value of the second-best contract. Hence, the total reward is always higher than the second-best second-stage contract.

Existence of Solution to Principal's Problem. Recall the requirements from Clarke (2013), Theorem 23.11. Denote the control variables by a and the state variables by A. The theorem applies as

- the laws of motion of the state variables, $g(A)$ are measurable in t and continuous in A
- the control set $a \in \mathcal{A}$ is closed and convex
- the running cost is
  - Lesbesgue measurable in t and (A, a)

  - lower semicontinuous in (A, a)
  - convex in a for any (t, A)

- the effort path $a_t = 0$ for all t and $A_t = 0$ for all t is admissible and delivers a finite value.

The running cost is given by

$$\Lambda(t, A, a) = a_t e^{-rt}(p_0 e^{-\lambda^g A_t} + (1-p_0)e^{-\lambda^b A_t})\left(\Pi(t, v(t)) - v(t)\right). \tag{2.46}$$

Convexity in a of the set $\{\Lambda(t, A, \cdot)\}$ has to be established. It suffices to show that Λ is concave in a. Suppose we are in the case with $\gamma_t = 0$. Then, it remains to show that $\Pi(t, v(t))$ is concave in $v(t)$. Recall that

$$\Pi(t, v(t)) = \int_0^{T(v(t))} e^{-rt} p_{2,0}(t) e^{-\lambda s}(p_s \lambda \pi - c) ds \tag{2.47}$$

$$= \frac{p_{2,0}(t)(\pi\lambda - c)}{r+\lambda}\left(1 - e^{-(r+\lambda)T(v(t))}\right) - \frac{1 - p_{2,0}(t)c}{r}\left(1 - e^{-rT(w(t))}\right). \tag{2.48}$$

Hence, we get

$$\frac{d^2\Pi(t, v(t))}{dv(t)^2} = \frac{d^2T(v(t))}{dv(t)^2}\left(p_{2,0}(t)(\pi\lambda - c)e^{-(r+\lambda)T(w(t))} - (1-p_{2,0}(t))ce^{-rT(w(t))}\right)$$
$$- \left(\frac{dT(w(t))}{dw(t)}\right)^2\left(p_{2,0}(t)(\pi\lambda - c)(r+\lambda)e^{-(r+\lambda)T(w(t))} - r(1-p_{2,0}(t))ce^{-rT(w(t))}\right)$$

which is less than zero. Hence, the running cost is concave in the promised utility and we can conclude that the set $\Lambda(t, A, \cdot)$ is convex.

Proof of Theorem 2.2. Suppose the principal considers introducing an additional deadline, T_1' before the initial one, T_1 to replace the agent if he succeeds before the second but after the first. To simplify notation denote by Π the surplus in the second stage. Introducing T_1' alters the principal's profits by

$$\int_0^{T_1'} e^{-rt}(1-p_0)\frac{p_t}{1-p_t}\left(\left(\Pi(t, w(t, T_1', T_1)) - \Pi(t, w(t, T_1, T_1))\right) - \left(w(t, T_1', T_1) - w(t, T_1, T_1)\right)\right) dt$$
$$+ \int_{T_1'}^{T_1} e^{-rt}(1-p_0)\frac{p_t}{1-p_t}\left(\left(\Pi^{SB}(t) - \Pi(t, w(t, T_1, T_1))\right) - \left(w^{SB}(t) - w(t, T_1, T_1, T_1)\right)\right)$$
$$- \int_{T_1'}^{T_1} e^{-rt}(1-p_0)\frac{p_t}{1-p_t}\frac{c}{p_{T_1}\lambda^g + (1-p_{T_1})\lambda^b} dt.$$

Multiplying by $\frac{1}{T_1 - T_1'}$ and taking the limit $T_1' - T_1 \uparrow 0$ we have

$$\int_0^{T_1} e^{-rt}(1-p_0)\frac{p_t}{1-p_t}\frac{\partial w(t, T_1, T_1', T_1)}{\partial T_1'}\left(1 - \frac{\partial \Pi(t)}{\partial w(t, T_1)}\right) dt > 0 \tag{2.49}$$

where the inequality follows from $\frac{\partial \Pi(t)}{\partial w(t, T_1, T_1')} < 1$ if the agent's value is above the second-best value and $\frac{\partial w(t, T_1, T_1')}{\partial T_1'} > 0$. So there is an incentive to introduce entrepreneur replacement. Note that if stages are independent $\frac{\partial w(t, T_1, T_1'}{\partial T_1'}|_{T_1' = T_1} = 0$.
On the other hand, suppose $T_1' = 0$ and consider the choice of introducing a reward with continuation contracts instead of replacement, that is marginally increasing T_1'. This yields as gain

(2.50)
$$-e^{-rT_1'}(1-p_0)\frac{p_{T_1'}}{1-p_{T_1'}}\left(\Pi^{SB}(T_1') - w^{SB}(T_1') - \left(\Pi(T_1', w(T_1', T_1)) - w(T_1', T_1)\right) - \frac{c}{p_{T_1}\lambda^g + (1-p_{T_1})\lambda^b}\right).$$

Note that if $T_1' = 0$, $w(0, T_1) = \frac{c}{p_{T_1}\lambda^g + (1-p_{T_1})\lambda^b}$ and hence, we get

$$-(1-p_0)\frac{p_{T_1'}}{1-p_{T_1'}}\left(\Pi^{SB}(T_1') - w^{SB}(T_1') - \Pi(T_1', w(T_1', T_1))\right). \tag{2.51}$$

If $w^{SB}(T_1') < \frac{c}{p_{T_1}\lambda^g + (1-p_{T_1})\lambda^b}$, then $\Pi(T_1', w(T_1', T_1)) > \Pi^{SB}(T_1')$ and increasing T_1' is profitable. However, if $w^{SB}(T_1') > \frac{c}{p_{T_1}\lambda^g + (1-p_{T_1})\lambda^b}$, then $\Pi(T_1', w(T_1', T_1)) < \Pi^{SB}(T_1')$, however, as the principal's payoff falls in the parameters that would increase the agent's second-best value, $\Pi^{SB}(T_1') - w^{SB}(T_1') < \Pi(T_1', w(T_1', T_1))$ and hence, a period without replacement would be introduced.

Proof of Lemma 2.3. To show that the optimal two-stage contract converges to the optimal one-stage contract I consider first the limits of the second-stage contract's instruments, $T(v(\tau_1))$ and $b^2(T(v(\tau_1))$. Recall that

$$T(v(\tau_1)) = -\frac{v(\tau_1)}{(1-p_{2,0}(\tau_1))} - \frac{1}{\lambda}W_{-1}\left(-e^{-\frac{v(\tau_1)\lambda}{(1-p_{2,0}(\tau_1))}}\right) \tag{2.52}$$

We are now interested in the limit as $\lambda \to \infty$.

$$\lim_{\lambda\to\infty} T(v(\tau_1)) = \lim_{\lambda\to\infty}\left(-\frac{v(\tau_1)}{(1-p_{2,0}(\tau_1))} - \frac{1}{\lambda}W_{-1}\left(-e^{-\frac{v(\tau_1)\lambda}{(1-p_{2,0}(\tau_1))}}\right)\right) \tag{2.53}$$

$$= -\frac{v(\tau_1)}{(1-p_{2,0}(\tau_1))} - \lim_{\lambda\to\infty}\frac{1}{\lambda}W_{-1}\left(-e^{-\frac{v(\tau_1)\lambda}{(1-p_{2,0}(\tau_1))}}\right) \tag{2.54}$$

Applying L'Hôspital's Rule gives

$$= -\frac{v(\tau_1)}{(1-p_{2,0}(\tau_1))} - \lim_{\lambda\to\infty}\left(W_{-1}\left(-e^{-\frac{v(\tau_1)\lambda}{(1-p_{2,0}(\tau_1))}}\right)\right)' \tag{2.55}$$

$$= -\frac{v(\tau_1)}{(1-p_{2,0}(\tau_1))} - \lim_{\lambda\to\infty}\left(-\frac{v(\tau_1)}{(1-p_{2,0}(\tau_1))c}\left(1-\frac{1}{1+W_{-1}\left(-e^{-\frac{v(\tau_1)\lambda}{(1-p_{2,0}(\tau_1))}}\right)}\right)\right) \tag{2.56}$$

$$= -\frac{v(\tau_1)}{(1-p_{2,0}(\tau_1))} - \left(-\frac{v(\tau_1)}{(1-p_{2,0}(\tau_1))c}\left(1-\frac{1}{1+\lim_{\lambda\to\infty}W_{-1}\left(-e^{-\frac{v(\tau_1)\lambda}{(1-p_{2,0}(\tau_1))}}\right)}\right)\right) \tag{2.57}$$

$$\lim_{\lambda\to\infty} T(v(\tau_1)) = 0 \tag{2.58}$$

as $W_{-1}(x \uparrow 0) = \infty$.
Recall that the bonus is given by $b^2(v(\tau_1)) = \frac{c}{\pi\lambda}\left(1+\frac{1-p_{2,0}(\tau_1)}{p_{2,0}(\tau_1)}e^{\lambda T(v(\tau_1))}\right)$. We get for the limit

$$\lim_{\lambda\to\infty} b^2(v(\tau_1)) = \lim_{\lambda\to\infty}\frac{c}{\pi\lambda}\left(1+\frac{1-p_{2,0}(\tau_1)}{p_{2,0}(\tau_1)}e^{\lambda T(v(\tau_1))}\right) \tag{2.59}$$

$$= \frac{c}{\pi}\frac{1-p_{2,0}(\tau_1)}{p_{2,0}(\tau_1)}\lim_{\lambda\to\infty}\frac{e^{\lambda T(v(\tau_1))}}{\lambda} \tag{2.60}$$

and again by L'Hôspital's rule

$$= \frac{c}{\pi}\frac{1-p_{2,0}(\tau_1)}{p_{2,0}(\tau_1)}\frac{v(\tau_1)}{(1-p_{2,0}(\tau_1))c}\lim_{\lambda\to\infty}\left(1-\frac{1}{1+\lim_{\lambda\to\infty}W_{-1}\left(-e^{-\frac{v(\tau_1)\lambda}{(1-p_{2,0}(\tau_1))}}\right)}\right) \tag{2.61}$$

$$\lim_{\lambda\to\infty} b^2(v(\tau_1)) = \frac{v(\tau_1)}{\pi p_{2,0}(\tau_1)}. \tag{2.62}$$

Next, note that we have $\lambda^b \to 0$ and hence

$$\lim_{\lambda^b\to 0} p_{2,0}(\tau_1) = \frac{\lambda^g p_{\tau_1}}{\lambda^g p_{\tau_1} + (1-p_{\tau_1})\lambda^b} = 1. \tag{2.63}$$

So, that the second stage is immediately and successfully completed in the limit and the agent receives $v(\tau_1)$ as a payment while the principal keeps $\pi - v(\tau_1)$.
As a consequence of the limit of the belief after a success, we have that

$$\lim_{\lambda^b\to 0} \dot{v}(t) = 0 \tag{2.64}$$

and no informativeness rent is required in the first stage. Hence, we may conclude that the optimal two-stage contract converges to the optimal one-stage contract if $\lambda^b \to 0$.

3 The Effect of Asymmetric Common Value Uncertainty in Procurement Auctions

joint with Stefan Weiergräber

3.1 Introduction

Public procurement is an important sector of economies: In the OECD countries, public procurement accounted for 12.1% of GDP in 2013.[1] The use of auctions in public procurement aims at creating competition between bidders and at selecting the efficient firm to carry out the service. While auctions perform well in selecting the efficient bidder when participants are symmetric, this is not necessarily the case when participants are asymmetric. Importantly, bidders in procurement auctions are likely to be asymmetric: First, former monopolists or incumbents may be better informed about the common value component due to their experience or have access to superior information. Second, incumbents with a larger network may be either more or less efficient, for example, due to economies of scale or capacity constraints.

Many procurement auctions involve both a private and a common value component. While private value components typically consist of idiosyncratic cost components, typical common values are common cost components or potential revenues from the object. However, empirical studies of asymmetric auctions predominantly study private value auctions.[2] Neglecting potential asymmetries in common value components has substantial implications on the results: if incumbents win systematically more often the theory of asymmetric private value auctions attributes this to a more efficient cost distribution (see Maskin and Riley (2000a)). From an efficiency perspective the incumbent wins too few auctions because its competitors bid more aggressively. We show in a theoretical model with private and common value component that the dominance of a firm can also be explained by asymmetrically precise common value signals. In this case, the incumbent firm wins too many auctions from an efficiency perspective. Hence, it is an important empirical questions to distinguish and to quantify the respective importance of asymmetries in private and common value components.

Asymmetries between firms are particularly important in markets with experienced incumbents and entrants that recently became active in the market. In the 1990s, many European countries

[1]See OECD (2015).

[2]See Athey et al. (2011), Suzuki (2010), Estache (2008) and Tas (2017).

started to liberalize markets that used to be controlled by a state monopolist. The aim was a more efficient provision of publicly subsidized goods due to increased competition. In many markets experiences with privatization have been mixed, however. We exploit a detailed data set on awardings in the German market for short-haul railway passenger services (SRPS) from 1995 to 2011. Since SRPS are generally not profitable, the state procures specific tracks to train operating companies and subsidizes them for the provision of the service. While the aim of the liberalization was to attract competitors, the former state monopolist (DB Regio) still operates the majority of the tracks (71%, FAZ Nr. 281 (2015)). An explicit concern by procurement agencies and industry experts is that entrants are either not participating at all or bidding cautiously. The German market for SRPS is only one example, but given its size of 8 billion EUR in subsidies for 2016, is an important one that shares many features with similar markets in other countries. DB has more experience for the services and as a publicly held firm may have advantages for financing compared to its rivals. In addition, DB Vertrieb, which is integrated with the DB holding has access to all the ticket revenue and passenger data. Entrants and even agencies typically do not have access to this information (Monopolkommission 2013). Considering these asymmetries, the reasons for DB still being the dominant firm are not clear. On the one hand, it could be the efficient firm for most services. On the other hand, entrants might bid very cautiously due to DB's informational advantage about future revenues.

Comparing the winning bids of entrants with those of DB in gross auctions we find that entrants win with significantly lower bids than DB. However, in net auctions there is no significant difference in the winning bids. We take this as first evidence that DB is indeed better informed about the common value component as entrants shade their bids relatively more in net compared to gross contracts. Moreover, Hunold and Wolf (2013) provide reduced-form evidence for the fact that using net contracts makes it more likely that DB Regio wins the awarding.

We take our model that builds on the theoretical work of Goeree and Offerman (2003) to a detailed contract-level data set on German short-haul railway passenger service (SRPS) procurement auctions. With this data set we can disentangle the two asymmetries by making use of a variation in the contract design: local state agencies that procure these services can choose who bears the revenue risk from ticket sales. If the ticket revenues remain with the agency (*gross contract*) the auction is a standard asymmetric independent private values auction. If the train operating company is the claimant of the ticket revenues (*net contract*), the auction is one with a private value (cost) as well as a common value (ticket revenues) component. In a first step, we estimate the cost distributions of DB and the entrants from the winning bids in gross auctions. Identification follows from Athey and Haile (2007) who show that asymmetric independent private value auctions are identified from the winning bid and the winner's identity only. We use several contract characteristics to control for the specifics of the respective contracts and obtain the cost distribution conditional on contract characteristics. In a second step, we make use of the net auction data. Given the first-step results, we know the cost distribution for each of the awarded tracks. Hence, differences in bidding behavior that are not explained by the differences in cost distributions can be attributed to the common value component. We can estimate the bid distribution as functions of a *net cost signal* that consists of the private and common value signal as well as the informativeness of the common value signal.

The results of our structural analysis show no systematic cost advantage of DB over its rivals. Importantly, they are not as large as one may initially expect - under a pure private value assumption - given DB's dominance in the market for SRPS. The estimation of the informational advantage over its competitors reveals that indeed in most auctions DB holds significantly more precise information about future ticket revenues. This highlights the concerns in Monopolkommission (2015) that DB's dominance is at least partially due to its informational advantage which may call for regulatory interventions that symmetrize the information across the bidders. Alternatively, efficiency could be increased by awarding more gross contracts which eliminates the common value component from the auction. We study this intervention in a counterfactual analysis and find that entrants shade their bids less than in net contract auctions, making bid distributions more symmetric. This increases ex ante efficiency of the auctions from 64% to 75%.

The assumption that the choice between net and gross contracts is exogenous is important for our analysis. This choice is typically strongly agency-dependent and there is only very little variation within an agency over time, while track characteristics differ within and across agencies. Therefore, we believe that the contract type (gross vs. net) is indeed not driven by fundamental characteristics of a track but rather exogenously determined by the preferences of the agency.[3]

Related literature The SRPS industry is characterized by asymmetries of bidders due to the presence of a former state monopolist and incumbent on many tracks, DB. Therefore, our methodology builds on the theory of first-price asymmetric auctions.[4] For example, theory predicts that stochastically weaker firms bid more aggressively and stochastically stronger firms win with higher profits (Maskin and Riley 2000a). Moreover, the release of public information in a symmetric model implies more aggressive bids by all firms (reduced information rents). Our application is also reminiscent of the theoretical literature on auctions of fixed price vs. cost-plus contract as in Laffont and Tirole (1986) and McAfee and McMillan (1986). While they focus on asymmetric information between the procurement agency and bidders and moral hazard after a contract has been awarded, we abstract from the latter and focus on informational asymmetries between competing bidders during the auction stage.

There is relatively little empirical literature on asymmetric common value auctions due to known difficulties with identification in common value auctions (see Athey and Haile (2002). Li and Philips (2012) analyze the predictions of the theoretical asymmetric common value auction model in Engelbrecht-Wiggans et al. (1983) in a reduced-form analysis. They find evidence for private information of neighbor firms in drainage lease auctions. Hong and Shum (2002)

[3]A comparison of track characteristics between the two different contract modes shows no significant difference in either of the observed contract characteristics. In Bahn-Report (2007) it is also argued that the choice is rather agency-dependent than an endogenous choice due to contract characteristics.

[4]See, for example, Parreiras (2006), Kirkegaard (2009), Maskin and Riley (2000a), Reny and Zamir (2004).

investigate the effect of competition in a model with both private and common value components and symmetric bidders. They find that the winner's curse effect can outweigh the competition effect so that more bidders can results in less aggressive bidding. In addition, Hong and Shum (2002) estimate the relative importance of private and common value components in procurement contracts in New Jersey. In contrast to their study, we have relatively precise information about which parts of the contracts correspond to private and which to common value components. Furthermore, we observe exogenous differences in the design of different auctions that eliminate or add specific parts of risk for the bidding firms. This allows us to focus on the effect of asymmetric information about the common value across incumbent and entrant bidders.

De Silva et al. (2003) analyze an asymmetric procurement model and confirm the theoretical predictions of Maskin and Riley (2000a) with reduced form regressions using data on highway procurement in Oklahoma. In a follow up paper, De Silva et al. (2009) argue that asymmetric information about contract characteristics is a particularly important problem for new entrants. However, their application is quite different from ours: We analyze a setting in which the incumbent is usually more cost-efficient, but also faces less uncertainty about ticket revenues than the entrants. In order to estimate our auction model, we rely on the literature on the structural estimation of asymmetric auctions. In particular, we borrow elements from Brendstrup and Paarsch (2006), Brendstrup and Paarsch (2003), Athey et al. (2011) and Hendricks et al. (2003) and adapt them to our application.

While recent research (Lalive et al. 2015) analyzes the respective benefits of auctions and negotiations in the context of our application, to the best of our knowledge, we are the first to analyze the role of auction designs and asymmetric information in this market using structural econometric methods. Lalive et al. (2015) analyze how the agency's choice of whether to engage in direct negotiations or to run an auction affects procurement outcomes. While they focus on the trade-off between competitive auctions and non-competitive negotiations, we focus on the specific auction design, in particular the implications of procuring net or gross contracts.

3.2 Auction Model and the Effect of Asymmetries

In this section, we present our model for procurement auctions of gross and net contracts and study the effect of two asymmetries: (i) the effect of asymmetric private value distributions, and (ii) the effect of asymmetric precision of the common value signals. All auctions are standard first-price sealed-bid auctions. The valuation for a contract as well as the bidding behavior of firms crucially depend on whether a gross or a net contract is tendered. For a

- *gross* contract, the valuation consists solely of the firm-specific costs of the contract (c_i) since the firm's revenue is fully determined by the winning bid.
- *net* contract, the valuation consists of the firm-specific costs of the contract (c_i) and additionally the ticket revenues R, which are unknown to all firms when bidding for a

contract.

We index bidding firms by i, its bid by b_i and denote the number of bidders by N. The cost component c_i is a private value drawn for each firm i from F_{c_i} and is observed by i only. The ticket revenue, R, is an unknown common value for which firms observe only a private signal, r_i drawn from F_r. We allow F_{c_i} to differ across firms to model differences across incumbent and entrants in cost efficiency. The differential information about expected revenues comes from the reliability of the own signal drawn from the common distribution F_r as discussed later on. All signals are independent across firms and cost signals are independent of revenue signals within firms. We assume that bidders are risk neutral.

Gross contract auctions Firms compete for a single indivisible item (one track) by submitting bids b_i (the requested subsidy). Firm i's ex-post value of winning and the expected value of a bid is given by the formulas for an independent private values (IPV) auction:

$$v_i = b_i - c_i \tag{3.1}$$

$$E[v_i(b_i)] = (b_i - c_i) \cdot \Pr(b_i < min_{i \neq j} b_j | c_i, b_i) \tag{3.2}$$

Ties are broken randomly. For reasons to be discussed in the next subsection, we assume that F_{c_i} is logconcave. As the incumbent is vertically integrated with the network operator, DB Netz, and is the former state monopolist as well as still publicly held by the Federal Republic of Germany, we assume that the incumbent draws its costs from a different distribution than the entrants. We further assume that entrants are symmetric for simplicity. As this gives rise to an asymmetric auction, we build on the theoretical work on asymmetric IPV auctions, in particular, we build on the predictions in Maskin and Riley (2000a).

In a gross contract, there is only ex ante uncertainty about the operating costs c_i. Before bidding, a firm receives private information about its costs which is distributed according to $F_{c^i} \in \mathcal{C}^2$ with strictly positive density on support $[c_{i,L}, c_{i,H}]$. After having received the signal, firm i knows its cost perfectly. However, it does not know its rivals' cost realizations.

Therefore, firm i chooses b to maximize expected profit:

$$\pi_i(b, c_i) = (b - c_i) \prod_{j \neq i} (1 - F_j(\phi_j(b))$$

where $\phi_j(b)$ is bidder j's inverse bid function. Rodriguez (2000) and Reny and Zamir (2004) establish that a unique equilibrium in pure strategies with strictly increasing and differentiable bid functions exists. The equilibrium is implicitly defined by a system of differential equations in inverse bid functions with boundary conditions. The solution to that system gives equilibrium existence. Denote by G_i the distribution of the opponents' maximum bid given own bid b being pivotal and a set of bidders N. Inverse bid functions have to satisfy:

$$b_i = c_i + \frac{1 - G_i(b, b, N)}{g_i(b, b, N)}. \tag{3.3}$$

We borrow the following Lemma and definition of conditional stochastic dominance[5] both adapted to the procurement setting from Maskin and Riley (2000a).

Lemma 3.1 (Maskin and Riley (2000a), Proposition 3.3 and Proposition 3.5.)**.** *If the private value distribution of i conditionally stochastically dominates the private value distribution bidder j, then i is the weak bidder and bids more aggressively than bidder j. The bid distribution of i is stochastically dominates the bid distribution of j.*

Lemma 3.1 shows that the weaker bidder bids more aggressively. As a result, the auction may be inefficient and the strong bidder wins too few auctions from an efficiency perspective. This result has been generalized by De Silva et al. (2003) to also hold in the presence of an additional common value component and therefore, also holds for the net auction case.

Net contract auctions When net contracts are procured, the bidders' value of a contract consist of a private cost and a common value component. We develop an asymmetric first-price auction model with both, private and common value components. The value of the item differs among bidders and consists of two components: (i) a private component, which is the cost of fulfilling the contract c_i drawn from distribution F_{c_i} (same as in gross auctions), and (ii) a common component, the ticket revenues, R. In addition to the private value signal, firms receive an additional signal r_i drawn from F_r on the common value, i.e. the expected ticket revenues R. Revenue signals, r_i, are conditionally independent given R. Because of the additional revenue component, the ex-post value of winning and the expected value of a bid is:

$$\pi_i = R - c_i + b_i \tag{3.4}$$

$$\mathbb{E}[\pi_i(b)|b, c_i, r_i] = \left(b - c_i + \alpha_i r_i + \sum_{j \neq i} \alpha_j \mathbb{E}\left[r_j | \rho_i \geq B_j^{-1}(b) \right] \right) \left(1 - F_{\rho_j}^{1:N-1}(B_j^{-1}(b) \right) \tag{3.5}$$

Both, F_{c_i} and F_r, are assumed to be logconcave. From a theoretical perspective, this model is a modified version of Goeree and Offerman (2003) which studies the symmetric case and the extension by De Silva et al. (2003) which allows for asymmetric private-value distributions.

We employ a standard but important assumption: the common value component is given by the weighted average of the signals received, i.e., $R = \sum_{i=1}^{N} \alpha_i r_i$. This assumption is crucial due to the following reason: The strategic variable for a bidding firm is its bid, i.e. a scalar. However,

[5]Conditional Stochastic Dominance is defined as follows: There exists $\lambda \in (0, 1)$ and $\gamma \in [c_{j,L}, c_{i,H}]$ such that $1 - F_i(x) = \lambda(1 - F_j(x))$ for all $x \in [\gamma, c_{j,H}]$ and $\frac{d}{dx}\frac{1-F_i(x)}{1-F_j(x)} > 0$ for all $x \in [c_{j,L}, \gamma]$.

the valuation for the success is two-dimensional, consisting of the private and common value component. There is typically no straightforward mapping from two-dimensional signals into a one-dimensional variable. However, with the linear specification and logconcavity of the signal distribution this is possible as shown in Goeree and Offerman (2003): the expected value of winning can be rewritten as a linear composition of the private signals, r_i and c_i as $\rho_i \equiv c_i - \alpha_i r_i$, and terms independent of the private information. This scalar statistic is sufficient to capture the private information in one dimension. Therefore, the standard auction theory methods as in (Milgrom and Weber 1982) can be applied.

We extend the model of Goeree and Offerman (2003) and its extension in De Silva et al. (2003) by allowing for asymmetries not only in the private value component and but also in the common value component. Denote $R = \sum_{i=1}^{N} \alpha_i r_i$ with $\sum_{i=1}^{N} \alpha_i = 1$. While every firm draws its signal r_i from the same distribution, the asymmetry between incumbent and entrants is captured by α_i. We denote the incumbent's and entrants' weights by α_I and α_E, respectively. Intuitively, α_i measures informational value of a bidder's signal. A higher α_i indicates a more reliable revenue signal for bidder i. The variance $var[R] = \sum_{i=1}^{N} \alpha_i^2 \sigma_r$ and conditional on a signal r_i, we get:

$$\mathbb{E}[R|r_i = r] = \alpha_i r + \sum_{j \neq i} \alpha_j \mathbb{E}[r_j] = \alpha_i r + \sum_{j \neq i} \alpha_i R$$

due to independence of the revenue signals $\{r_j\}_{j=1}^{N}$, the variance is given by:

$$var[R|r_i = r] = \sum_{j \neq i} \alpha_i^2 \sigma_r$$

As $\alpha_i = \alpha_E$ for all entrants and $\alpha_i = \alpha_I$ for the incumbent, we get:

$$var[R|r_E = r] = ((N-2)\alpha_E^2 + \alpha_I^2)\sigma_r \text{ for the entrant} \tag{3.6}$$

$$var[R|r_I = r] = (N-1)\alpha_E^2 \sigma_r \text{ for the incumbent} \tag{3.7}$$

and hence $var[R|r_E = r] > var[R|r_I = r]$ if $\alpha_I > \alpha_E$. Note that the vector $\alpha = (\alpha_I, \alpha_E)$ effectively consists only of one parameter since we can normalize $\alpha_I + (N-1)\alpha_E = 1$. For now, we assume that the asymmetry is constant across all auctions.[6]

A bidder maximizes expected utility conditional on the observed signals. The structure of the common component allows us to write this as $\alpha_i r_i - c_i + b_i + \sum_{j \neq i} \alpha_j r_j$ where the last term is independent of the own signals. We can summarize the bidder's private information as $\rho_i = c_i - \alpha_i r_i$ which should be interpreted as a *net cost signal* or *negative profitability signal*. Bidding behavior is described by a system of differential equations as derived in the following Lemma. The standard derivation is carried out in the Appendix.

Lemma 3.2. *The following system of differential equations constitutes a Bayesian Nash equilib-*

[6] At the expense of having to estimate additional parameters, we can model σ_r and α as a function of track characteristics X.

rium of the first-price auction with asymmetric cost distribution and asymmetric signal precision:

$$b = \left(c_i - \alpha_i r_i - \sum_{j \neq i} \alpha_j \mathbb{E}\left[r_j | \rho_i = B_j^{-1}(b) \right] \right) + \frac{(1 - F_\rho^{1:N\backslash i}(B_j^{-1}(b)))}{f_\rho^{1:N\backslash i}(B_j^{-1})(b) B_j^{\prime -1}(b)} \tag{3.8}$$

where $f_\rho^{1:N\backslash i}$ and $F_\rho^{1:N\backslash i}$ denote the density and distribution function of the first-order statistic of other players' signals. $B_j^{-1}(\cdot)$ denotes the inverse bid function of bidder j.

The intuition is analogous to bidding in the gross auction. Players bid their expected valuation of winning the auction plus a bid-shading term. However, in the net auction case the expected valuation also depends on the other players' revenue signals. Because this is a common value setting, the bidder faces a winner's curse motif. This can be seen in the conditioning set of the expectation of the other players' revenue signals $\mathbb{E}\left[r_j | \rho_i = B_j^{-1}(b) \right]$. If bidder i wins with bid b, then it must be the case that other players' signals were not too good. Hence, when computing that expectation, the player has to take that into account.

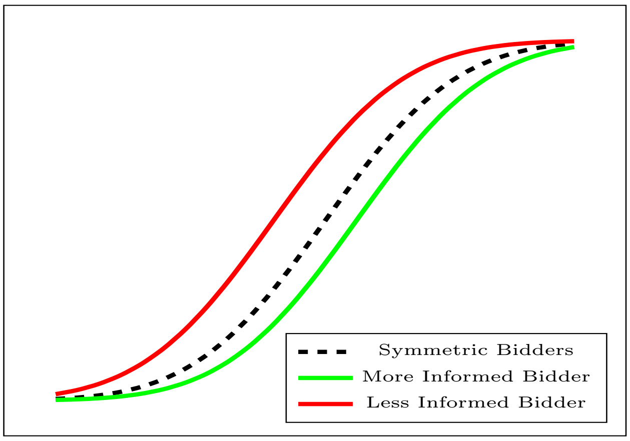
While theory gives strong predictions about how winning bids by the incumbent and the entrants compare in private value auctions, this is much less clear in net contracts because of the additional common revenue component. Especially if the revenue signal firms receive have asymmetric precision. We give an intuition on the effect of the common value asymmetry in the following Lemma that assumes a symmetric and known cost component.

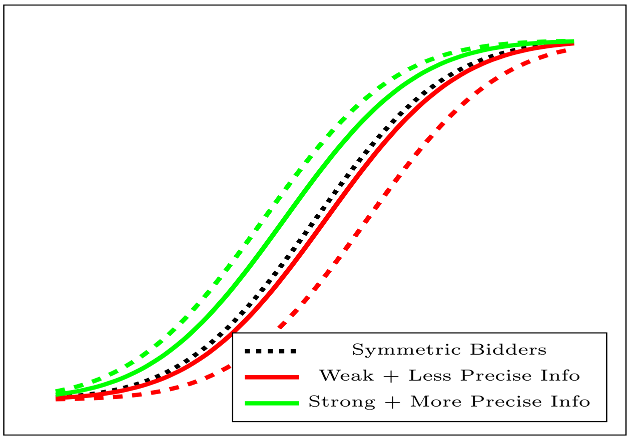
Lemma 3.3. *Assume there are two firms that have the same cost c. Then, if $\alpha_1 > \alpha_2$ and the distributions of the compound common value signals $\alpha_i r_i$ satisfy conditional stochastic dominance, bidder 2 shades her bid more than bidder 1.*

Lemma 3.3 shows that a less precisely informed bidder is affected more by the winner's curse and will shade its equilibrium bid more than a more precisely informed bidder.

In our setting, allowing for both asymmetries, there are two effects in place that determine bidding behavior: (i) one bidderis (potentially) more efficient than the other bidders on average. Hence, the less efficient bidders bid more aggressively than the more efficient bidder. (ii) One bidder is (potentially) more precisely informed about the common value component. This implies that the less informed bidders shade their bid more than the more informed bidder due to a stronger winner's curse effect. Taken together, the effects work can amplify each other or work against each other depending on the identity of the advantages.

In particular, each of the asymmetries can be a source of inefficiency in the auction. If the bidders are symmetric in the common value component, but one bidder is on average more efficient than the competitors, these bid more aggressive. Hence, the auction outcome may be inefficient, when the efficient realized cost advantage is not too big. If the bidders' private





value distributions are symmetric, but their common value signals asymmetrically precise, the auction may be inefficient as well due to an asymmetric winner's curse effect. If the less precisely informed bidder draws a lower cost but both the same common value signal, she wins the auction only if the cost advantage is sufficiently big, because she shades her bid more than the competitor. If both asymmetries are present, there are two possibilities: (i) The more efficient bidder is also more precisely informed. In this case, the more aggressive bidding by the disadvantaged bidder due to the weaker private value distribution is mitigated by the stronger winner's curse effect. (ii) The more efficient bidder is less precisely informed. In this case, the less aggressive bidding by the stronger bidder is amplified due to the winner's curse effect. This is illustrated in Section 3.2.

3.3 Application: Short Haul Railway Passenger Services in Germany

3.3.1 Industry description

As many other industries, the German railway sector was liberalized in the 1990s. This liberalization followed the EU Directive 91/440 *Development of the community's railways* implemented through the *Eisenbahnneuverordnungsgesetz* in 1993. One of the main objectives was to induce competition in the railway sector. Towards this, the *regionalisation* was carried out. Short hail railway passenger services are part of the universal service obligation and not profitable for operators. Therefore, procurement agencies on behalf of the federal states were assigned the task to choose an operator that provides this service. As these services require high subsidies (around 7 billion EUR in 2016, Monopolkommission (2015)), the procurement agencies aim at competition *for* the tracks to keep the required subsidies at a low level.

In another part of the reform, the former state monopolist Deutsche Bundesbahn in West Germany and Deutsche Reichsbahn in East Germany merged into Deutsche Bahn AG which is still publicly owned by the Federal Republic of Germany. As a consequence, entrants into the market for German SRPS compete with a publicly held operator, Deutsche Bahn AG (DB), that formerly was the state monpolist.

When procuring these services, the procurement agencies have a high degree of freedom in designing the contract as well as the rules of the awarding. The agencies specifies the basic components of the contract: for example, how frequent a company has to run services on a certain line, the duration of the contract and the type of vehicles to be used. One important additional feature is that it also chooses who obtains the ticket revenues, the agency itself or the train-operating company. When the agency receives ticket revenues the contract is called a gross contract, while the contract is called a net contract, when the operating company receives the ticket revenues.

While the market share of competitors has been rising over the years since the liberalization, in 2013 DB still had a market share of 73.6% measured in train-kilometers (see Monopolkommission (2015)). This raised a debate about the underlying reasons: are features in the procurement process reinforcing the dominance of DB or is it due to DB being the efficient firm in the market? We assess this question in the empirical implementation of our auction model.

3.3.2 Data description

Our data set consists of (almost) all procurement contracts from the German market for SRPS from 1995 to 2011. The data contain detailed information on the awarding procedure, contract characteristics, the number of participating firms, the winning bid and the identity of the winning firm. Table 3.1 displays an overview of our sample size for different subset of awardings. While this data set contains relatively few observations, it is to our knowledge the most comprehensive data set on the German market for SRPS out there and we plan to supplement it with the most

Table 3.1: Number of observed train line awardings by winning firm and auction mode

Auctions	gross contracts	net contracts
Incumbent wins	22	39
Entrant wins	55	51
$\sum$	77	90

recent awardings from 2012-2016. Moreover, we collected data on demographic characteristics of the track region and data on track access charges and frequency of service from the German Federal Statistical Office and additional publicly available sources. Currently, the estimation of gross (net) auctions is based on 77 (90) awardings respectively.

3.3.3 Relating the theory to the application

The procurement agencies choose, when procuring a contract, whether the agency or the operator receives the ticket revenues. We assume that this choice is exogenous. In general, one might be worried that these differences across gross auctions and net auctions are driven by selection issues and endogenous procurement decisions by the agencies. This is a potential problem if agencies decide the contract mode (net vs. gross) based on unobservable contract characteristics that inherently favor either the incumbent or the entrants. We argue that the role of endogenous contract mode is negligible in our application for two reasons. First, we do not find systematic differences in the most important track characteristics across our two groups of auctions from which we conclude that the two sets of tracks are very similar. Second and more importantly, industry experts also proclaim that the main procurement features are mostly determined by agency preferences that generally are orthogonal to the structural cost and revenue characteristics of a track, cf. the extensive discussion in Bahn-Report (2007).

Theoretically, the difference between net and gross contracts is the presence of a common value component, the ticket revenues. As most features of the contract that affect demand are pre-specified by the agency, for example, the frequency of the service, the type of vehicle to be used, we consider the demand to be a common value for all firms. Moreover, we consider the costs to be a private value as the firms have different access to vehicles, funding opportunities, and can apply different wages.[7]

While we expect entrants to be symmetric with respect to their cost distribution, we expect the cost of DB to be potentially different from the entrants' cost. First, DB owns a large pool of vehicles that it can easily reuse for various services, entrants typically have to buy or lease vehicles. The cost for vehicles is a significant component of the costs of serving a contract. Also, DB is likely to have cheaper access to funds as a publicly held firm. Altogether, we expect DB

[7] There certainly are common components in the cost like electricity and infrastructure charges. However, these can be anticipated by the firms in advance and involve relatively little uncertainty.

to have a cost advantage.

In net auctions, there is additional uncertainty about future demand and therefore about ticket revenues. Again, we expect systematic differences between DB and its competitors. DB Regio (the branch of DB that operates in the SRPS sector) is vertically integrated with DB Vertrieb GmbH. Most tickets - even when DB is not operating the track - are sold through DB Vertrieb GmbH. Therefore, DB possesses an informational advantage about demand as competitors cannot access the information that DB Vertrieb GmbH has (see Monopolkommission (2015)).

Given these observations, we model gross auctions as an asymmetric independent private value auction and net auctions as an auction with private and common values in which we allow for asymmetries in the private value component and asymmetric precisions of the common value signal.[8]

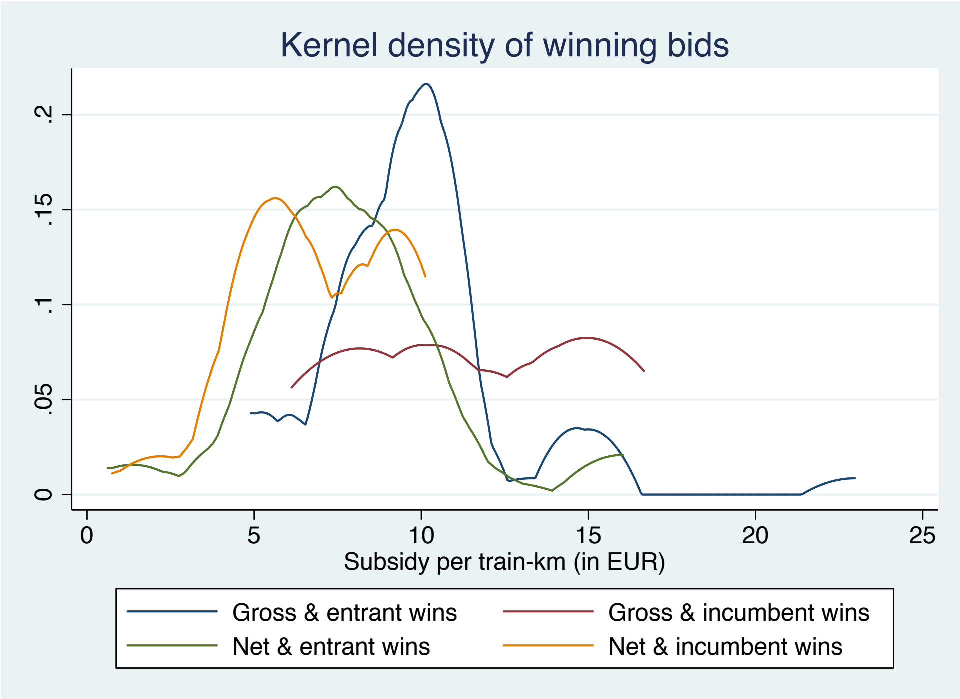
3.3.4 Reduced-form evidence and descriptive statistics

An analysis of the raw data provides support for our initial conjecture that DB has an informational advantage over its competitors. However, we do not find strong evidence that DB is more cost-efficient than its competitors (in a conditional stochastic dominance sense). The theoretical model predicts that if a bidder has a cost advantage over a competitor, than his bid distribution is also shifted to the left. As a consequence, the winning bid of the more cost-efficient bidder should be systematically lower than the winning bids of its competitors. If the common value is added to the model and the precision of the common value signal is asymmetric, the model predicts bids of the informationally disadvantaged bidder to be systematically higher, if the cost are symmetric.

Figure 3.1 show that when entrants win with lower bids than DB in gross auctions, which provides evidence that DB might not be more efficient than its competitors. Comparing the winning bids in net auctions, the winning bids are closer for DB and the entrants. This is partial evidence that the bid function of the entrants is stronger affected by the common value component than the bid function of DB. Also, Hunold and Wolf (2013) show that DB wins significantly more when the contract that is auctioned is a net and not a gross contract.

[8] The main criticism of the model by Goeree and Offerman (2003) is that the uncertainty about the common value increases in the number of bidders N. We believe that is less of a concern for our model. First, we expect α_I to be much larger than the entrants' α_E. Since the incumbent bids in all auctions, variation in N across auctions comes only from a different number of less informed entrants. Second, increased uncertainty about ticket revenues could well be consistent with our data. Tracks that attract more bidders are usually contracts with a higher expected profit, but higher expected profits usually come also with higher demand risk. While we are aware of the shortcomings of this model, we believe it nevertheless provides a framework that fits our application well. With a sufficiently large sample, we could avoid this problem, by simply estimating the asymmetry parameters separately for each N.

Figure 3.1: Kernel density of winning bids



3.4 Identification & Estimation

3.4.1 Identification arguments

The cost distributions in an asymmetric IPV model are non-parametrically identified from the winning bid, the number of bidders and the identity of the winner (Athey and Haile 2002). In contrast, the non-parametric identification of a common value component is much more complicated. Identification of the joint distribution of the common value and all the signals requires observing the full bid distribution and either exogenous variation in the number of bidders or the ex post value of the auctioned object. In principle, the realized ticket revenues are observable. Unfortunately, currently we do not have access to it. Identification of just the joint distribution of all common value signals fails if some bids are not observed. In principle, the full bid distribution is recorded by the agencies. Unfortunately, we do not have access to these data at this point. Therefore, we cannot provide a formal identification argument for our common value component. Similarly to Hong and Shum (2002), we rely on an intuitive argument to identify the distribution of the common value.

Intuitively, identification of the revenue risk parameters comes from comparing differences between the incumbent's and the entrants' bidding strategies across gross and net auctions. Our key idea is to compare similar tracks under different procurement mechanisms (net vs. gross). Since the procurement mode is assumed to be orthogonal to unobserved contract characteristics, any systematic difference in bidding behavior should be attributed to the revenue uncertainty in net auctions. In addition, we exploit some arguably mild functional form assumptions that help us in identifying the common value component. For example, we assume independence of bidders' revenue signals instead of trying to identify their joint distribution from the data.

3.4.2 Estimation strategy

Our estimation proceeds in two steps. First, we estimate the asymmetric IPV model using data on auctions of gross contracts. This allows us to compute the distribution of costs for a track with given characteristics. Second, we estimate or model with private (cost) and common value (ticket revenue) components using data on net auctions. Since we extrapolate the cost distributions from the first step, we can isolate the effect of the common value signal in the second step.

As in Athey et al. (2011) we assume that there are two types of bidders: DB as the incumbent who participates in all auctions and $N-1$ symmetric entrants. Asymmetry complicates the estimation since in general the differential equations in the first order conditions do not have a closed-form solution anymore. An additional complication is that under asymmetry the markup term has to be computed for each bidder configuration, i.e. for each number of bidders, separately.[9] With a sufficiently large sample, we can follow the non-parametric approach of Brendstrup and Paarsch (2003) who generalize Guerre et al. (2000) to asymmetric IPV auctions.

Since the total number of procured tracks is still relatively small, a fully non-parametric estimation will be very imprecise. Therefore, we employ a parametric approach. As in Lalive et al. (2015) and Athey et al. (2011), we assume that the bid functions $G()$ follow a Weibull distribution:

$$G(b_i|X,N) = 1 - \exp\left[-\left(\frac{b_i}{\lambda(X,N)}\right)^{\nu(X,N)}\right] \tag{3.9}$$

where λ and ν are the scale and shape parameters. Both vary across incumbent and entrants and are modeled as a function of observed contract characteristics:

$$\begin{aligned}
\log(\lambda^I(X,N)) &= \lambda_0^I + \lambda_X^I X + \lambda_N^I N \\
\log(\lambda^E(X,N)) &= \lambda_0^E + \lambda_X^E X + \lambda_N^E N \\
\log(\nu^I(X,N)) &= \nu_0^I + \nu_X^I X + \nu_N^I N \\
\log(\nu^E(X,N)) &= \nu_0^E + \nu_X^E X + \nu_N^E N
\end{aligned}$$

where I and E denote the incumbent and entrants respectively. In order to keep the number of parameters reasonably low, we include only the number of train kilometers and the infrastructure access costs associated with using the corresponding track sections in the contract characteristics

[9] Campo et al. (2003) develop a non-parametric estimation technique that is appropriate for this setting.

X. We believe, that the admission costs charged by DB Netz are a good proxy for the type of track that is procured. Moreover, the total number of train kilometers is a good proxy for the complexity of a project. Finally, we include the contract's specified frequency-of-service as an additional regressor as a proxy for demand conditions.[10]

Since we only observe the winning bids, our estimation relies on the first order statistic, i.e. the lowest realization of N random variables where $N-1$ bids are drawn from the entrants' distribution and one is drawn from the incumbent's distribution. With one incumbent and $N-1$ entrants, the density of the first order statistic conditional on the incumbent or an entrant winning are given by (see Appendix XXX for the derivation):

$$h(x^{(1:N)}, I) = g^I(x)(1 - G^E(x))^{N-1} \tag{3.10}$$

$$h(x^{(1:N)}, E) = (N-1)g^E(x)(1 - G^E(x))^{N-2}(1 - G^I(x)). \tag{3.11}$$

The likelihood function is then based on equations (3.10) and (3.11):

$$LL(\lambda, \nu) = \sum_{j=1}^{T_G} \log(h_{b^{1:N}}(b_j)) \tag{3.12}$$

where b_j denotes the winning bid in auction j and T_G is the total number of gross auctions in our sample. Given the estimated parameters of the bid distributions, we can back out the cost distribution of each track with characteristics X by inverting bidders' FOCs. Following Athey et al. (2011), we compute the cost distribution for a given track without imposing any additional parametric assumptions as follows:

1. Draw a pseudo-sample of bids for both incumbent and entrant from the estimated bid distributions. $G^I(b|X, N)$ and $G^E(b|X, N)$.
2. The pseudo-sample of bids has to satisfy the the first-order conditions 3.13 and 3.14:

$$\hat{c}^I = b_i^I - \frac{1 - \hat{G}_{M,B}^I(b_i^I, b_i^I, N)}{\hat{g}_{M,B}^I(b_i^I, b_i^I, N)} \tag{3.13}$$

$$\hat{c}^E = b_i^E - \frac{1 - \hat{G}_{M,B}^E(b_i^E, b_i^E, N)}{\hat{g}_{M,B}^E(b_i^E, b_i^E, N)} \tag{3.14}$$

In our procurement application, the markup terms can be computed as follows:

$$\begin{aligned} \hat{G}(b_i) &= G_{M_i|B_i}(b_i|b_i, X, N) \\ &= Pr(\min_{j \neq i} B_j \geq b_i | b_i, X, N) \\ &= (1 - G^E(b_i))^{N-2}(1 - G^I(b_i)) \text{ (for an entrant)} \\ &= (1 - G^E(b_i))^{N-1} \text{ (for the incumbent)} \end{aligned} \tag{3.15}$$

[10] Experimenting with different regressors yields qualitatively similar results which are available upon request.

where in the last 2 lines G^E and G^I denote the estimated bid distributions for incumbent and entrants. Intuitively, $\hat{G}(b_i)$ describes the CDF of the lowest rival bid evaluated at the actual winning bid b_i, i.e. conditioning on the event that bid b_i was pivotal. The denominator of the markup term $\hat{g}$, is simply the derivative of $\hat{G}$:

$$\begin{aligned} \hat{g} &= \frac{\partial \hat{G}(b_i)}{\partial b_i} \\ &= -(N-1)(1-G^E(b_i|X,N))^{N-2} g^E(b_i|X,N) \text{ (for the incumbent)} \\ &= -(N-2)(1-G^E(b_i|X,N))^{N-3} g^E(b_i|X,N)(1-G^I(b_i|X,N)) \\ &- g^I(b_i|X,N)(1-G^E(b_i|X,N))^{N-1} \text{ (for entrants)} \end{aligned} \tag{3.16}$$

3. This results in a pseudo-sample of cost realizations for each track. Now, kernel smoothing treating $\hat{c}$ as a draw from the cost distribution can be used to compute the cost distribution non-parametrically.

Using the gross auction estimates, we can compute the cost distribution for each track and each bidder type. In our second step we use these to extrapolate costs to the net auction contracts. This allows us to focus on the effects of the common value signals on incumbent's and entrants' bidding behavior.

Recall that we assume that firms receive a pair of signals (c_i, r_i) for private costs and common revenues respectively. We assume that revenue signals r_i are drawn from a logconcave distribution $F(R, \sigma_r)$ with mean R and variance σ_r. As discussed in Section 3.2 the structure of our net auction model allows us to combine the two signals into one *net cost* signal: $\rho_i = c_i - \alpha_i r_i$ that completely determines bidding behavior. Moreover, we denote the expected valuation of the contract conditional on winning the auction with bid b by $\mathcal{P}_i \equiv c_i - \alpha_i r_i - \sum_{j \neq i} \alpha_j \mathbb{E}\left[r_j | \rho_i = B_j^{-1}(b)\right]$ given inverse bid functions B_j^{-1}. Then, according to Lemma 3.2 bidding behavior is determined by the system

$$\mathcal{P}^I = b^I - \frac{1 - G^I_{M,B}(b^I, b^I, N)}{g^I_{M,B}(b^I, b^I, N)} \tag{3.17}$$

$$\mathcal{P}^E = b^E - \frac{1 - G^E_{M,B}(b^E, b^E, N)}{g^E_{M,B}(b^E, b^E, N)}. \tag{3.18}$$

Our goal in this section is to estimate the additional parameters contained in the common value component. In particular, we are interested in the parameter vector α that describes the different precision of the players' information. Our net auction estimation proceeds in two steps:

1. Since we have relatively few observations, we continue to follow a parametric estimation approach. We assume that bid functions follow a Weibull distribution whose parameters are functions of track and contract characteristics (analogous to the gross auction estimation). After having estimated the net bid function parameters, we can back out the combined cost-revenue signal (net cost signal) $\mathcal{P}$ based on the first-order conditions 3.17.

2. Afterwards, we can treat $\mathcal{P}_i$ as known and transform the sample of winning bids into a sample of (winner's) expected valuations given the winning bid b. Moreover, from the gross auction step we know the cost distributions from which c is drawn. This allows us to isolate the revenue signal part of $\mathcal{P}$ via

$$\mathcal{P}^i \equiv \rho_i - \sum_{j \neq i} \alpha_j \mathbb{E}\left[r_j | \rho_i = B_j^{-1}(b)\right] \tag{3.19}$$

$$\mathcal{P}^i - c_i = -\alpha_i r_i - \sum_{j \neq i} \alpha_j \mathbb{E}\left[r_j | \rho_i = B_j^{-1}(b)\right] \tag{3.20}$$

We know $\mathcal{P}$ from the first step and the distribution of c from the gross auction step. Therefore, the LHS "is known" (in expectation). The distribution of the RHS is based on $r \sim F(R, \sigma_r)$ and can be computed up to a vector of parameters (R, σ, α). Thus, we can estimate the parameters using maximum likelihood.

Derivation of the conditional expectation. To carry out the second step, we need to compute the conditional expectation in the expected valuation of winning the line with bid b. The expectation term conditions on the bid being pivotal, i.e. $\rho_i = B_j^{-1}(b)$. However, from the first step we only know the compound expected valuation conditional on winning with bid b. As a consequence, we have to decompose $\mathcal{P}^i$ into ρ_i (i's own signal) and the expectation about rivals' revenue signals. This is a non-trivial exercise as we have to do this consistently with the first-order conditions for equilibrium bidding. We make use of the fact that in equilibrium given the signal ρ_i, the conditional expectation term is a deterministic number. Intuitively, it describes i's expectation about the opponents' revenue signal conditioning on the event that i won with bid b and that b was a pivotal bid.

Given the first step of the estimation procedure we can compute for every winning bid b_w the corresponding (compound) signal that induces opponents to bid b_w, i.e. the opponents' signal that makes b_w pivotal. If i is the winning bidder, denote this signal by $\bar{\mathcal{P}}_{-i}(b_w)$ and note that if an entrant wins, this is immediately given by the winning $\mathcal{P}$ of this line for the other entrants. For any arbitrary player $-i$ this can be computed by inverting bidder $-i$'s bid function at the observed winning bid:

$$\bar{\mathcal{P}}_{-i}(b_w) = b_w - \frac{1 - G_{M,B}^{-i}(b_w, b_w, N)}{g_{M,B}^{-i}(b_w, b_w, N)}. \tag{3.21}$$

This gives us for every line with corresponding winning bid a sample of N expected valuations conditional on winning with a bid b_w. These have to be consistent with each other due to the following observation: In the expected value of i's opponents' signals, the conditional expectation of i's revenue signal appears again. Hence, we have for each auction N equations in N unknowns conditional on the corresponding ρ_i. The equation system is given by assuming that i wins the auction with bid b

$$\bar{\mathcal{P}}^i(b) = c_i - \alpha_i r_i - \sum_{j\neq i} \alpha_j \mathbb{E}\left[r_j | \mathcal{P}^i = \bar{\mathcal{P}}^j(b)\right] \text{ (for winner)} \tag{3.22}$$

$$\bar{\mathcal{P}}^j(b) = c_j - \alpha_j r_j - \sum_{k\neq j} \alpha_k \mathbb{E}\left[r_k | \bar{\mathcal{P}}^j(b) = \bar{\mathcal{P}}^k(b)\right] \text{ (for } N-1 \text{ rival bidders)} \tag{3.23}$$

with $\bar{\mathcal{P}}^i(b) = \mathcal{P}^i(b)$. This is a fixed-point problem in N unknowns conditional on a set of parameters α, R, σ_r. These unknowns are the conditional expectations about the opponents' revenue signals. $\bar{\mathcal{P}}^j(b)$ can be computed from our estimation in the first step. Then, the expectation term for every j is, observing that r_j has to satisfy (by simply rearranging the FOC)

$$r_j(c_j) = \frac{1}{\alpha_j}\left(c_j - \bar{\mathcal{P}}^j(b) - \sum_{k\neq j} \alpha_k \mathbb{E}\left[r_k | \bar{\mathcal{P}}_j(b) = \bar{\mathcal{P}}^k(b)\right]\right), \tag{3.24}$$

which we can use to compute

$$\mathbb{E}\left[r_j | \bar{\mathcal{P}}_j(b) = \bar{\mathcal{P}}^j(b)\right] = \int_{\underline{c}}^{\bar{c}} r_j(c) f_r(r_j(c)) f_{c,j}(c) dc \tag{3.25}$$

where the joint density $f(c_j, r_j) = f_{c,j}(c_j) f_r(r_j(c_j))$ follows from the independence of the revenue and cost signals. Then, applying (3.24) in (3.25) and using this in (3.22) delivers us a system of N equations in N unknowns for any combination of parameters ρ_i, c_i, α_i for every i. However, as entrants are symmetric this reduces to a two-dimensional system with unknowns X^I and X_E[11]

(3.26)

$$\bar{\mathcal{P}}^I(b) = c_I - \alpha_I r_I - (N-1)\alpha_E \int_{\underline{c}}^{\bar{c}} \frac{1}{\alpha_E}\left(c - \bar{\mathcal{P}}^E(b) - \underbrace{\sum_{k\neq j} \alpha_k \mathbb{E}\left[r_k | \bar{\mathcal{P}}_j(b) = \bar{\mathcal{P}}^k(b)\right]}_{X_E}\right) f_r(r_E(c)) f_{c,E}(c) dc$$

(3.27)

$$\bar{\mathcal{P}}^E(b) = c_E - \alpha_E r_E - (N-2)\alpha_E \int_{\underline{c}}^{\bar{c}} \frac{1}{\alpha_E}\left(c - \bar{\mathcal{P}}^E(b) - \underbrace{\sum_{k\neq j} \alpha_k \mathbb{E}\left[r_k | \bar{\mathcal{P}}_j(b) = \bar{\mathcal{P}}^k(b)\right]}_{X_E}\right) f_r(r_E(c)) f_{c,E}(c) dc$$

$$- \alpha_I \int_{\underline{c}}^{\bar{c}} \frac{1}{\alpha_I}\left(c - \bar{\mathcal{P}}^I(b) - \underbrace{\sum_{k\neq j} \alpha_k \mathbb{E}\left[r_k | \bar{\mathcal{P}}_j(b) = \bar{\mathcal{P}}^k(b)\right]}_{X_I}\right) f_r(r_I(c)) f_{c,I}(c) dc$$

[11] X_E and X_I differ only in the composition of the firms over which the summation is taken.

For now, we are only interested in the conditional expectation terms and hence, we can reduce this system using $X_i = \sum_{j\neq i} \alpha_j \mathbb{E}[r_j|\overline{\mathcal{P}}^j(b) = \overline{\mathcal{P}}^j(b)]$ further to:

$$X_I = (N-1)\int_{\underline{c}}^{\overline{c}} \big(c - \bar{\mathcal{P}}^E(b) - (N-2)\alpha_E X_E - \alpha_I X_I\big) \tag{3.28}$$

$$f_r(\frac{1}{\alpha_E}\big(c - \bar{\mathcal{P}}^E(b) - (N-2)\alpha_E X_E - \alpha_I X_I\big)) f_{E,c}(c) dc \tag{3.29}$$

(3.30)

$$X_E = (N-2)\int_{\underline{c}}^{\overline{c}} \big(c - \bar{\mathcal{P}}^E(b) - (N-2)\alpha_E X_E - \alpha_I X_I\big)$$

$$f_r(\frac{1}{\alpha_E}\big(c - \bar{\mathcal{P}}^E(b) - (N-2)\alpha_E X_E - \alpha_I X_I\big)) f_{E,c}(c) dc \tag{3.31}$$

$$+ \int_{\underline{c}}^{\overline{c}} \big(c - \bar{\mathcal{P}}^I(b) - (N-1)\alpha_E X_E\big)\, f_r(\frac{1}{\alpha_I}\big(c - \bar{\mathcal{P}}^I(b) - c - (N-1)\alpha_E X_E\big)) f_{I,c}(c) dc.$$

This system can be solved (numerically) for (X_I, X_E) for any given set of parameters (α, R, σ_r) using the estimated distributions f_{c_i} from the gross auctions and the distributional assumptions on f_{r_i} being a truncated normal distribution with mean R, variance σ_r and truncation thresholds $\underline{r}$ and $\overline{r}$. Existence of a solution to the system follows directly from Brouwer's fixed point theorem as it is a continuous mapping from a convex and compact set to itself. Formally proving uniqueness of the fixed point is much harder. Therefore, we rely on extensive robustness checks in which we initiate the solver at different starting values to check that the results are likely to constitute the unique fixed point.

Derivation of the Likelihood Function. Given the values of the conditional expectation terms, X_I, X_E, for any vector of parameters (R, σ_r, α), we can construct a likelihood function from the first-order conditions for equilibrium bidding using the estimated values $\mathcal{P}^i$:

$$\mathcal{P}^I = c_I - \alpha_I r_I - \underbrace{(N-1)\alpha_E X_E(R, \sigma_r, \alpha)}_{E_I} \tag{3.32}$$

$$\mathcal{P}^E = c_E - \alpha_E r_E - \underbrace{(N-2)\alpha_E X_E(R, \sigma_r, \alpha) - \alpha_I X_I(R, \sigma_r, \alpha)}_{E_E} \tag{3.33}$$

where the left-hand side is the "dependent variable" $\mathcal{P}^i$ that we back out in the first stage. The right-hand side depends on the parameters (R, σ_r, α) and is the sum of two independent random variables. We can compute their density using the convolution of their distributions. The pdf of $\alpha_i r_i$ is $f_{\alpha_i r_i} = \frac{1}{\alpha_i}\mathcal{N}(r_i/\alpha_i; R, \sigma_r, \underline{r}, \overline{r})$. c_i is distributed according to $f_c(c_i)$. Hence, the density of $c_i - \alpha_i r_i$ is:

$$f_{c_i - \alpha_i r_i}(x) = \int_{-\infty}^{\infty} f_{-(\alpha_i r_i)}(y - x) f_c(y) dy \tag{3.34}$$

where x is the right-hand side of Equation (3.32).

Finally to capture revenue heterogeneity across tracks, we model the mean of the revenue distribution (R) as a function of the frequency of service and the total number of train kilometers as a sufficient statistic for demand, so that $R = \gamma_0 + \gamma_1 fs + \gamma_2 tkm$. Similarly, we model the variance of the revenue signal distribution as a function of the contract length: $\sigma_r = \bar{\sigma}_r + \gamma_3 cl$. In principle, one can also parametrize α in a variety of ways. For example, we could estimate the asymmetry parameters as a continuous function of the number of bidders or a function of time which would enable us to allow for entrants learning about the common value over time.

3.5 Estimation results

Table 3.5 displays the results for the estimation of bid functions in gross and net auctions for the incumbent and the entrants. In a highly non-linear model it is difficult to interpret the magnitude of the coefficients.[12] Therefore, we focus on the shape of the implied bid functions and cost distribution estimates. We provide graphs for bid functions and cost distributions for both incumbent an entrant for several representative gross and net auction lines in Appendix (3.8).

Generally bid functions in gross auctions are close for incumbent and entrants and entrants' dominate the incumbent's bid function in the lower tail. This is consistent with the theory of asymmetric auctions which prescribes weaker bidders to bid more aggressively. Our cost distribution estimates are mostly as expected: Generally, the incumbent's cost distribution dominates the entrants' distribution, i.e. incumbents cost are shifted to the left. However, for most lines this difference is smaller than what one would expect and on a several lines entrants even seem to have a cost advantage.

When comparing a typical bid function in a gross auction with one in a net auctions, we find striking differences. Overall, in net auctions the incumbent is much more aggressive compared to the entrants. This is line with our theoretical model that prescribes that entrants who are at a higher risk of the winner's curse will shade their bids more.

Having estimated bidding behavior in both gross and net auctions allows us to predict firms' hypothetical bids if net auction tracks would have been procured in a gross auction. Moving from net to gross contracts makes the bidding functions for the two types much more similar and often results in the familiar picture of the entrants bidding more aggressively than the incumbent in the left tail of the distribution. Analogously, estimating the cost distributions associated with our net contracts reveals positive but only small cost advantages for the incumbent.

Table 3.3 displays the estimation results for the revenue signal and asymmetry parameters. As

[12]One striking feature of our estimates are the high standard errors which are mostly due to our small sample size. This problem is well-known in the literature and shared with many other studies estimating parametric bid functions, for example Athey et al. (2011). One explanation is that the asymptotic MLE formula provides a bad description of the behavior of the estimator in small finite samples. In future versions, we plan to compute bootstrap standard errors instead of the currently provided asymptotic MLE standard errors.

Table 3.2: Estimation results: Bid function parameters

	Gross auctions	Net auctions
λ_0^I	0.8107	-1.5602***
	(0.9570)	(0.3893)
λ_X^I	2.5096***	23.5200
	(0.3706)	(29.2770)
	0.0121	0.6044
	(2.8902)	(5.0368)
	0.8513	-2.9337
	(4.0425)	(5.9271)
λ_N^I	0.1554	1.5298
	(0.7100)	(2.2238)
λ_0^E	2.2926**	-3.2925**
	(1.0765)	(1.4675)
λ_X^E	2.8809	66.7780
	(2.7107)	(100.6200)
	-0.7412	-3.2713
	(10.5870)	(27.6500)
	-1.1131	2.5225
	(11.1180)	(11.1990)
λ_N^E	-0.3706	0.5490
	(2.1102)	(2.8292)
ν_0^I	10.4300***	-2.8274
	(0.3564)	(19.0480)
ν_X^I	0.2138	-32.4190
	(4.3872)	(74.5280)
	0.8138	-1.7658
	(31.6310)	(42.4940)
	-13.2610	18.7310
	(43.7320)	(31.6820)
ν_N^I	-2.8742	-0.4042
	(13.3170)	(16.2530)
ν_0^E	-0.3185	2.4473
	(6.2125)	(13.6530)
ν_X^E	0.1677	-22.5010
	(4.6607)	(64.9760)
	1.1258	4.1126
	(23.2190)	(21.5850)
	2.0834	-0.7447
	(14.1550)	(36.1390)
ν_N^E	0.6701	-0.7632
	(5.2758)	(3.3997)

Table 3.3: Estimation results: Revenue parameters and asymmetry parameters

	Parameter estimates
α_2^I	0.6224*** (0.0140)
α_{3+}^I	0.5480*** (0.1943)
σ_{r0}	2.8665*** (0.2855)
σ_{r1}	0.8292*** (0.1441)
β_{R0}	2.0355 (1.5010)
β_{R1}	8.1742 (6.0590)
β_{R2}	4.9362*** (0.7209)

expected the expected revenue is increasing in the size of the contract with a highly significant coefficient. The expected revenue is also increasing in the specified frequency of service although, due to a high standard error, not significant. The variance of the revenue distribution is highly significant and positive and, not surprisingly, increasing in the length of the contract.

Most importantly, our estimates for the asymmetry parameters reveal that the incumbent has a substantial information advantage. In our main specification, we estimate α_I separately for auctions with 2 bidders and 3 or more bidders. For $N = 2$, we get an estimated α_2^I of 0.62 implying $\alpha_2^E = 0.38$. Put differently, in auctions with only 2 bidders the incumbent typically has almost 66% more information about the ticket revenues than an entrant. An even more asymmetric pattern persists for auctions with more than 2 bidders. For example, in auctions with 3 bidders, we get $\alpha_3^I = 0.55$ implying $\alpha_3^E = 0.23$.

3.6 Counterfactuals

In this section, we consider a series of counterfactuals and analyze the effects of procurement design on efficiency and agency revenues. First, we define an ex ante efficiency measure in our setup and then compare the ex ante probability of selecting the efficient bidder for three scenarios: first, the actual gross auction sample, second the actual net auction sample and finally we analyze the efficiency effects of procuring the net auction sample as gross auctions. Afterwards, we propose several additional counterfactuals for future research. In particular, we plan to consider the symmetrization of the information between incumbent and entrants in net auctions; that is, enforcing an $\alpha = \frac{1}{N}$, as could be obtained by requiring DB to make its information public.

3.6.1 Efficiency

Consider bidder i winning with bid b resulting from cost realization c. The probability that bidder i winning with bid b is the efficient bidder is given by

$$\Pr(c \leq \min_{j \neq i} c_j | b \leq \min_{j \neq i} b_j). \tag{3.35}$$

According to the definition of conditional probabilities this is given by

$$\Pr(c \leq \min_{j \neq i} c_j | b \leq \min_{j \neq i} b_j) = \frac{\Pr(b \leq \min_{j \neq i} b_j \cap c \leq \min_{j \neq i} c_j)}{\Pr(b \leq \min_{j \neq i} b_j)}. \tag{3.36}$$

We can rewrite the second event in terms of the bids as follows: the cost $c = b_i^{-1}(b)$ of bidder i corresponding to the winning bid b is lower than the minimum cost of all opponents, $\min_{j \neq i} c_j$, then, the every other bidder j has to have bid more than the bid that corresponds to the same cost realization, i.e. $b_j(c) = b_j(b_i^{-1}(b))$. That is, the second event corresponds to the condition

$$b_j \geq b_j(b_i^{-1}(b)) \, \forall j \neq i. \tag{3.37}$$

Therefore, we get the new condition

$$\Pr(c \leq \min_{j \neq i} c_j | b \leq \min_{j \neq i} b_j) = \frac{\Pr(b \leq b_j \, \forall j \neq i \cap b_j \geq b_j(b_i^{-1}(b)) \, \forall j \neq i)}{\Pr(b \leq \min_{j \neq i} b_j)} \tag{3.38}$$

that only depends on the bid functions. Note that if bidders were symmetric, the first event trivially implies the second event and the ex ante probability of selecting the efficient bidder is equal to one. We can rewrite this condition further to

$$\Pr(c \leq \min_{j \neq i} c_j | b \leq \min_{j \neq i} b_j) = \frac{\Pr(b_j \geq b \cap b_j \geq b_j(b_i^{-1}(b)) \, \forall \neq i)}{\Pr(b \leq \min_{j \neq i} b_j)} \tag{3.39}$$

$$= \frac{\Pr(b_j \geq \max\{b, b_j(b_i^{-1}(b))\} \, \forall j \neq i)}{\Pr(b \leq \min_{j \neq i} b_j)}. \tag{3.40}$$

The max operator can be solved for each of the bidders directly from the bid functions for each b. Then, we can compute this probability directly from the bid functions estimated in the previous sections. The denominator is again given by the first-order statistics of the bid functions.

What remains to be done is to aggregate over all possible cases that can occur: all winning bids and the corresponding winner's identity. Therefore, the ex ante probability of selecting the efficient bidder is given by

$$\int_{\underline{b}}^{\overline{b}} \Pr(\text{incumbent } i \text{ wins with bid } b) \frac{\Pr(b_j \geq b \cap b_j \geq b_j(b_i^{-1}(b)) \, \forall j \neq i)}{\Pr(b \leq \min_{j \neq i} b_j)} \tag{3.41}$$

$$+ \Pr(\text{entrant } i \text{ wins with bid } b) \frac{\Pr(b_j \geq b \cap b_j \geq b_j(b_i^{-1}(b)) \, \forall j \neq i)}{\Pr(b \leq \min_{j \neq i} b_j)} dF(b) \tag{3.42}$$

where $F(b)$ is the distribution of the winning bid. The probability of the incumbent and an entrant winning given bid b is given by

$$\text{Pr(incumbent wins with } b) = \text{Pr(incumbent bids } b \text{ and all entrants bid } b_e \geq b) \tag{3.43}$$

$$= g_I(b)(1 - G_E(b))^{N-1} \tag{3.44}$$

$$\text{Pr(incumbent wins with } b) = \text{Pr(an entrant bids } b \text{ and all other bidders bid } b_i \geq b) \tag{3.45}$$

$$= (N-1)g_E(b)(1 - G_E(b))^{N-2}(1 - G_I(b)). \tag{3.46}$$

This yields the following ex ante probability of selecting the efficient bidder:

$$\int_{\underline{b}}^{\overline{b}} g_I(b)(1 - G_E(b))^{N-1} \frac{\Pr(b_j \geq b \cap b_j \geq b_j(b_i^{-1}(b)) \, \forall j \neq i)}{\Pr(b \leq \min_{j \neq i} b_j)} + \tag{3.47}$$

$$(N-1)g_E(b)(1 - G_E(b))^{N-2}(1 - G_I(b)) \frac{\Pr(b_j \geq b \cap b_j \geq b_j(b_i^{-1}(b)) \, \forall j \neq i)}{\Pr(b \leq \min_{j \neq i} b_j)} db \tag{3.48}$$

Table 3.4: Efficiency comparison for different auction formats

	Gross Auctions	Net Auctions	Net → Gross
Pr(selecting efficient firm)	0.7984	0.6363	0.7528

Table 3.4 displays the computed probabilities of selecting the efficient firm for various procurement modes. All auction modes are fairly efficient with average probabilities of selecting the efficient firm between 64% and 80%. Our gross auction sample exhibits the highest efficiency measure with an average probability of almost 80%. In contrast, the efficiency probability is substantially lower in our net auction sample (around 64%). One interpretation of this large difference is that the gross auction sample consists of lines that are somewhat easier to procure efficiently than the lines in the net auction sample. However, the difference could also be a direct effect of the different procurement modes. In order to investigate the latter effect, we compute the counterfactual efficiency probability when procuring the net auction sample as gross auctions.

We find a significant increase in the probability of selecting the efficient bidder (from 64% to 75%) brining the efficiency of the net auction sample almost to the efficiency level of the gross auction sample. One take-away message from this exercise is that in our application, the asymmetry introduced by potential cost asymmetries is relatively small compared to the inefficiency introduced by asymmetric information about the common value. Our policy implications are somewhat similar to the ones by Hong and Shum (2002): More competition, which is often put forward as an argument for net auctions, need not always be desirable, especially if the winner's curse is strong. In our application, letting train operating companies bear the revenue risks can be detrimental for procurement efficiency since net auctions are likely to put the incumbent at a large advantage.

3.6.2 Additional counterfactuals

Resulting subsidies While looking at efficiency probabilities is arguably the most important property of an auction design, the procurer might also care about the expected subsidy to be paid. Our estimates allow us to predict the subsidy that the agency has to pay to the winning firm. Since we have estimated the bid functions for all bidder types and all auction formats, we can compute the expected winning bid via:

$$\int_{\underline{b}}^{\overline{b}} b(g_I(b)(1-G_I(b))^{N-1} + (N-1)g_E(b)(1-G_E(b))^{N-2}(1-G_I(b))db. \tag{3.49}$$

When comparing the expected subsidy from gross and net auctions it has to be kept in mind that in gross auctions, the agency also obtains the ticket revenues and, therefore, this has to be substracted from the subsidy paid to the winning firm.

Gross as net auctions A straightforward extension is to analyze the effects for efficiency and revenues when procuring the gross auction sample as net auctions. In light of the above results for procuring net as gross auctions, we expect average efficiency to decrease.

Eliminating the informational asymmetry Completely abandoning net contracts might not be desirable. Therefore, we simulate how procurement outcomes would change if entrants and incumbents had the same information on the common value component, i.e. if their revenue signals have equal weight. In our setting this is done by setting α_i equal for every firm and appropriately adjusting the expected value of the signal. For example, one could simulate auction outcomes when the entrants and incumbent become equally informed by setting $\alpha_I = \alpha_E$. This will not affect the variance of the incumbent's signal, but will decrease the uncertainty for the entrants (see the theoretical discussion above). In practice, this could be easily achieved by mandating the incumbent to share its information on ticket sales with the procurement agencies and rival bidders. We expect that going all the way from net to gross auctions is not necessary to increase the ex ante efficiency and suspect that information-symmetric net auctions could well result in the highest efficiency probability.

3.7 Conclusion

We develop and study a model of procurement auctions that allows for asymmetries in the private value component and asymmetrically precise information on the common value component. Theory predicts that if a bidder is on average more efficient than his competitors, he will bid less agressively while the less efficient bidders bid more aggressively. Moreover, if a bidder is more precisely informed about the common value component, he is less affected by the winner's curse than the competitor and will shade his bid less than the competitors. Observing a dominant firm in the market can be explained by both asymmetries: the dominant firm can have on average lower costs than the competitors or be more precisely informed both allowing it to submit on average lower bids than the competitors.

We take this model to a data set on short haul railway passenger auctions in Germany. With this data set we can disentangle the two asymmetries by making use of a variation in the contract design: local state agencies that procure these services can choose who bears the revenue risk from ticket sales. If the ticket revenues remain with the agency (*gross contract*) the auction is a standard asymmetric independent private values auction. If the train operating company is the claimant of the ticket revenues (*net contract*), the auction is one with a private value (cost) as well as a common value (ticket revenues) component. In a first step, we estimate the cost distributions of DB and the entrants from the winning bids in gross auctions. Given the first-step results, differences in bidding behavior that are not explained by the differences in cost distributions can be attributed to the common value component.

The results of our structural analysis show no systematic cost advantage of DB over its rivals. Importantly, they are not as large as one may initially expect - under a pure private value assumption - given DB's dominance in the market for SRPS. The estimation of the informational advantage over its competitors reveals that indeed in most auctions DB holds significantly more precise information about future ticket revenues. This highlights the concerns in Monopolkommission (2015) that DB's dominance is at least partially due to its informational advantage which may call for regulatory interventions that symmetrize the information across the bidders. Alternatively, efficiency could be increased by awarding more gross contracts which eliminates the common value component from the auction. We study this intervention in a counterfactual analysis and find that entrants shade their bids less than in net contract auctions, making bid distributions more symmetric. This increases ex ante efficiency of the auctions from 64% to 75%.

3.8 Appendix

In this appendix, we provide bid functions and estimated cost distributions for several representative lines for both gross and net auctions.

3.8.1 Exemplary bid and cost distributions

Bid distributions in gross auctions

The following graphs display a comparison of incumbent and entrant bid functions for gross auction, i.e. auctions in which the bidders do not face any revenue risk.

Figure 3.2: Bid distribution for gross auction line 18

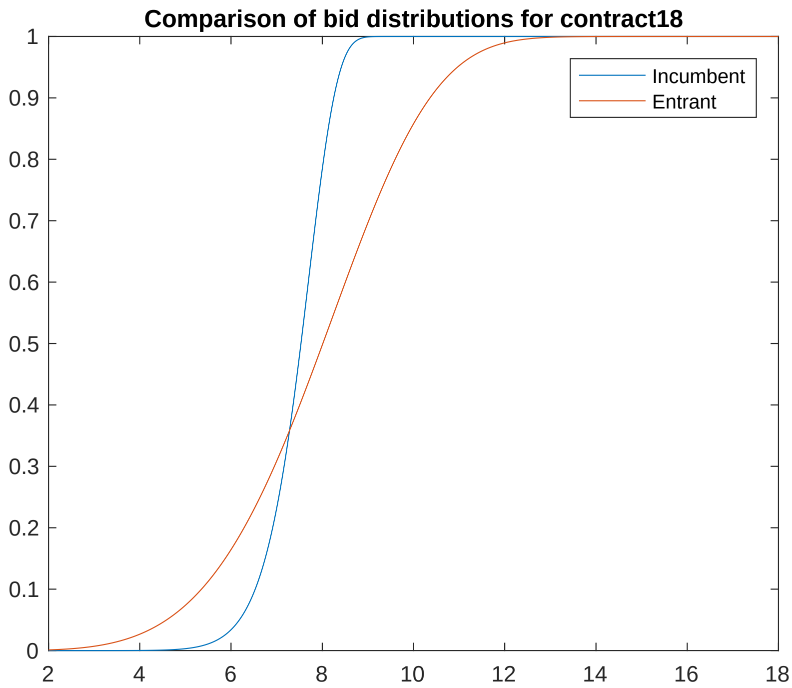


Figure 3.3: Bid distribution in gross auction line 20

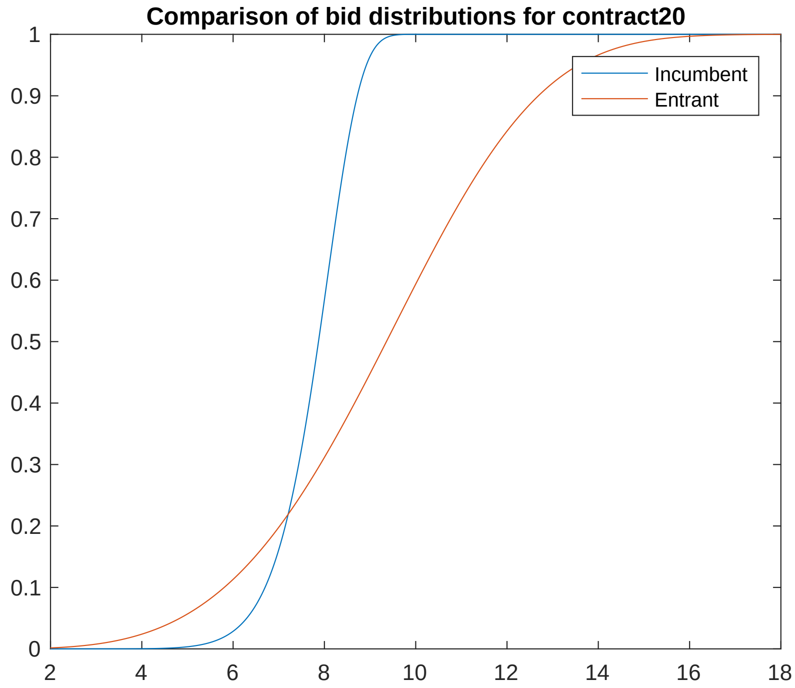
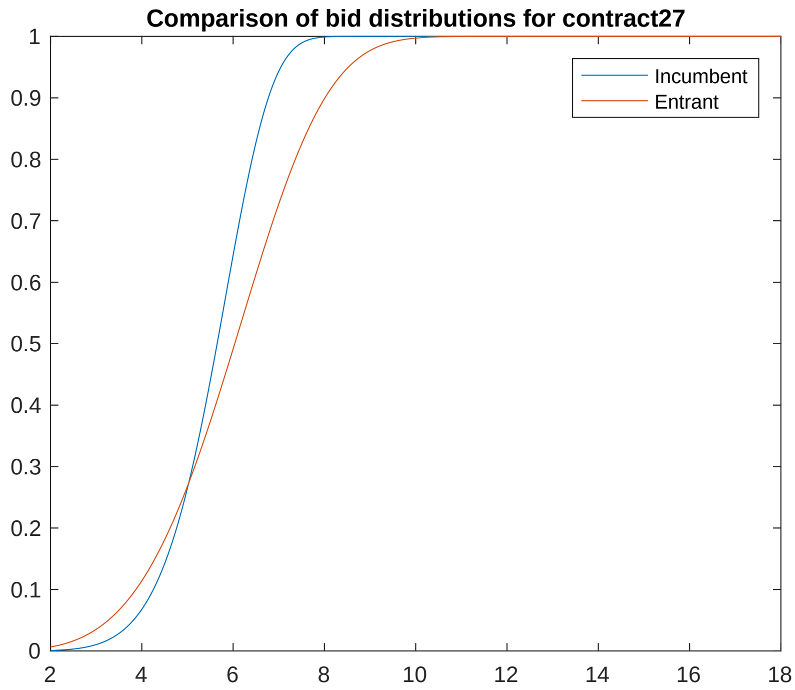


Figure 3.4: Bid distribution in gross auction line 27



Estimated cost distributions in gross auctions

In this section, we graph our estimates of the cost distributions associated with the bid functions provided in the previous section.

Figure 3.5: Cost distribution for gross auction line 18

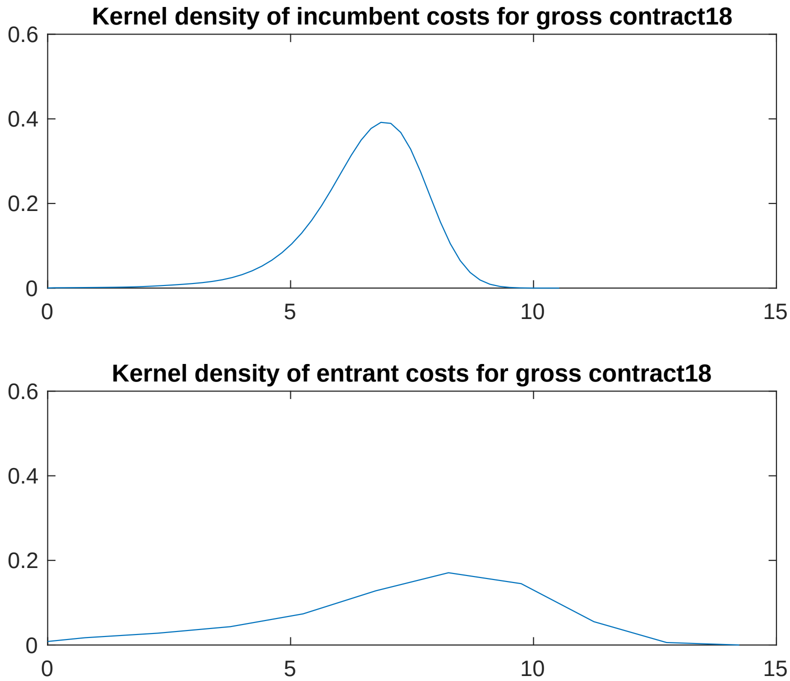


Figure 3.6: Cost distribution for gross auction line 20

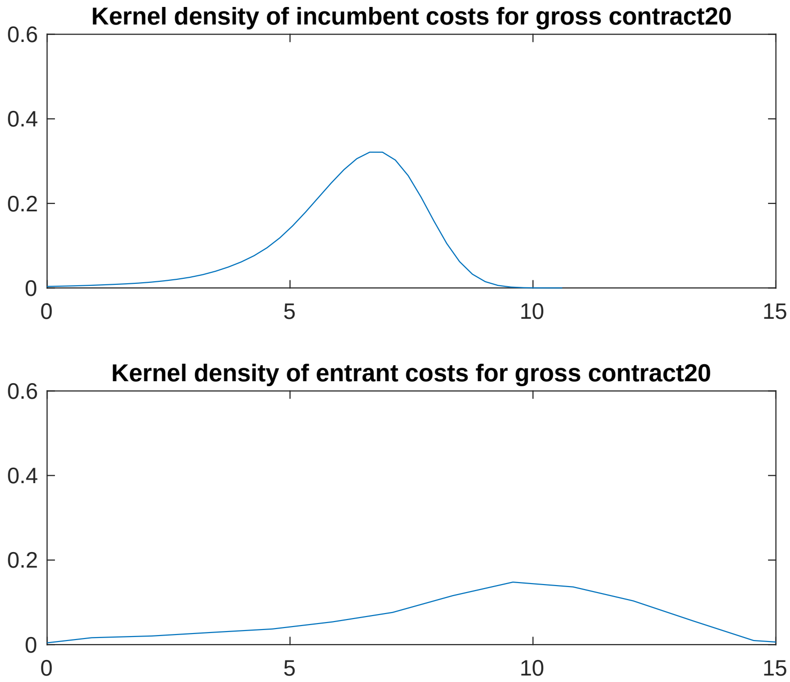
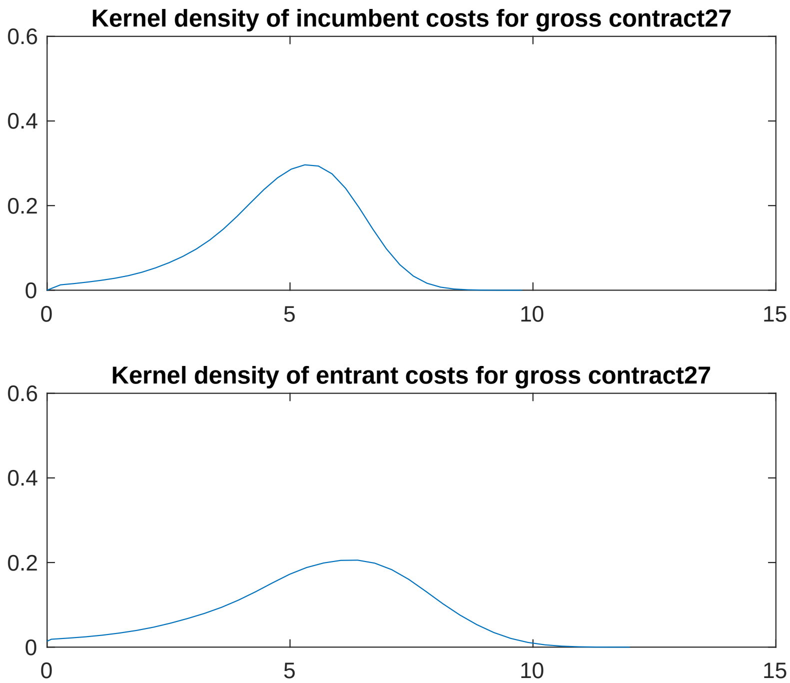


Figure 3.7: Cost distribution for gross auction line 27



Bid distributions in net auctions

In this section, we provide graphs of bid functions for the incumbent and the entrants for several representative net auctions, i.e. auctions in which the firm bears the revenue risk.

Figure 3.8: Bid distribution for net auction line 26

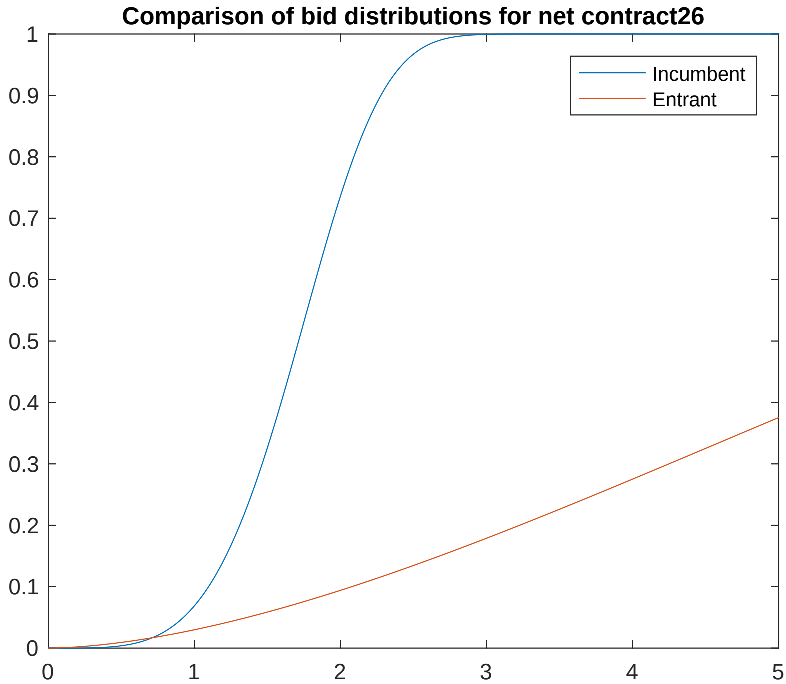


Figure 3.9: Bid distribution for net auction line 46

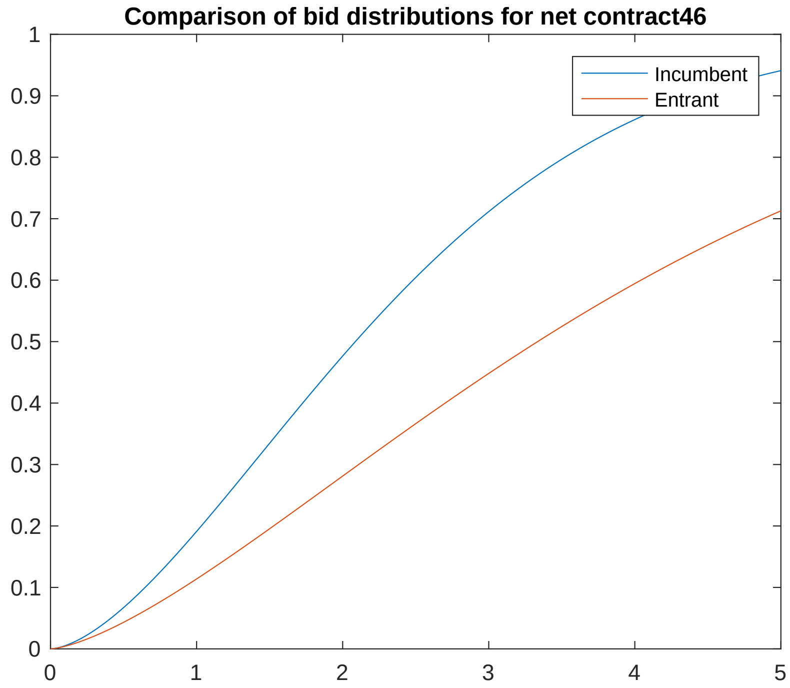
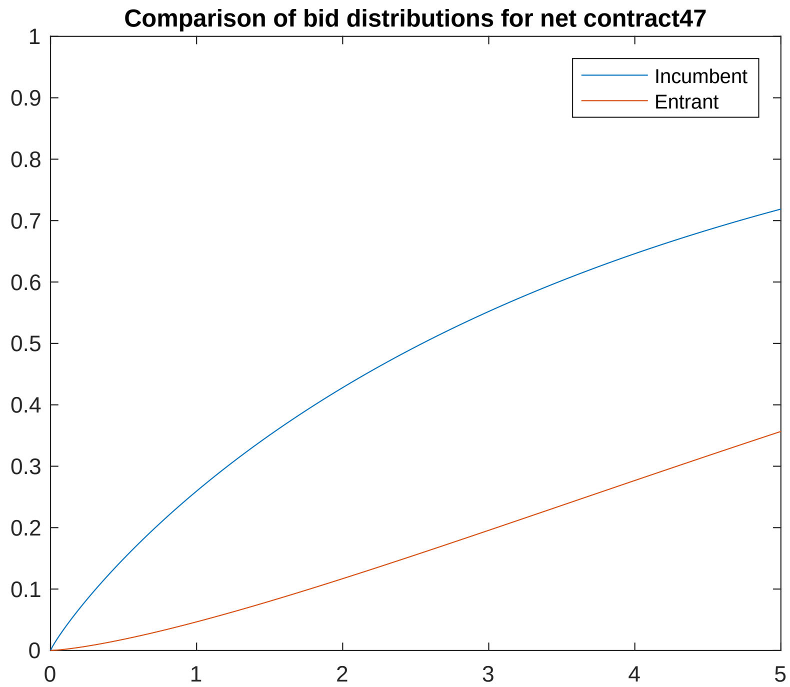


Figure 3.10: Bid distribution for net auction line 47



Hypothetical bid distributions in net auctions

In this section, we display hypothetical bid functions for the three net auctions presented previously. They illustrate how incumbent and entrant would have bid if the same train track would have been awarded as a gross instead of a net contract.

Figure 3.11: Hypothetical bid distribution for net auction line 26

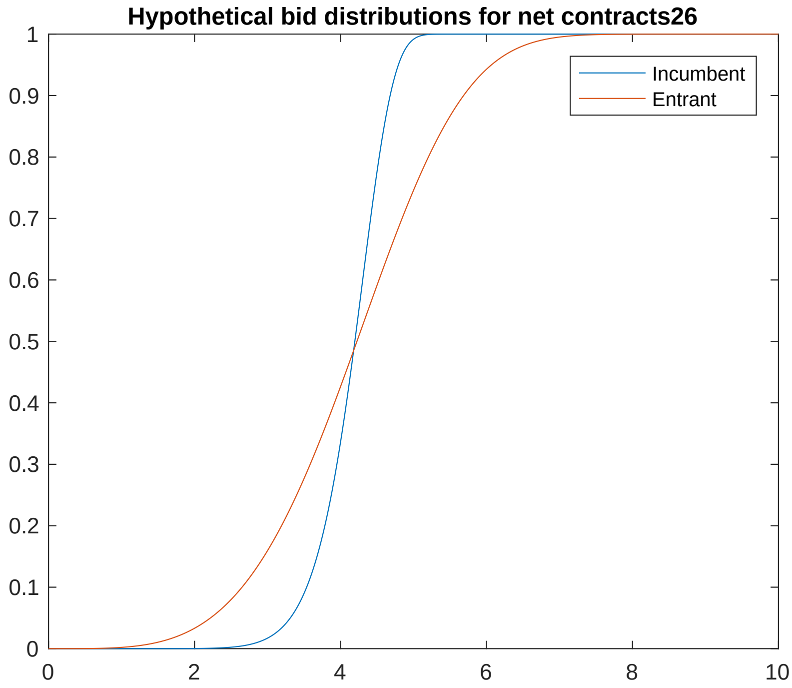


Figure 3.12: Hypothetical bid distribution for net auction line 46

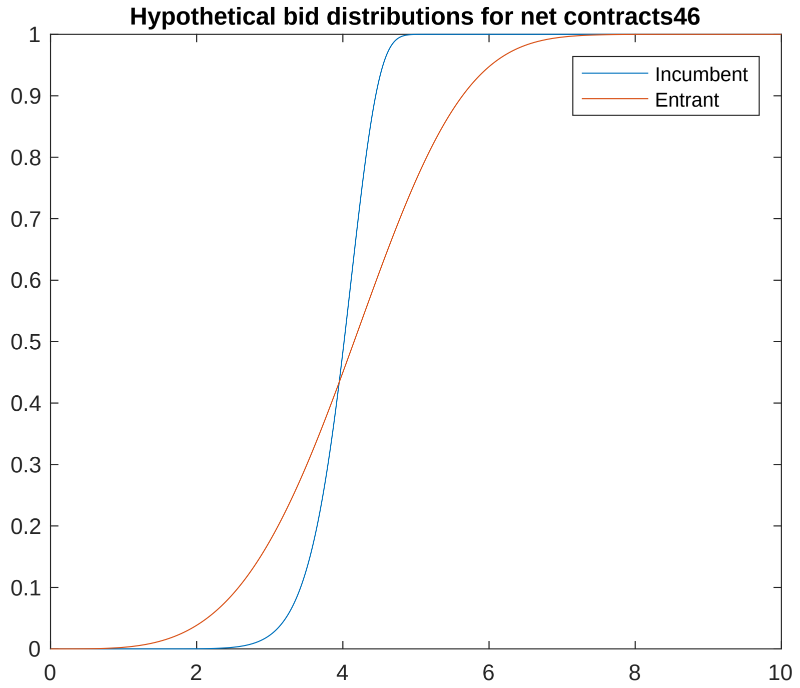
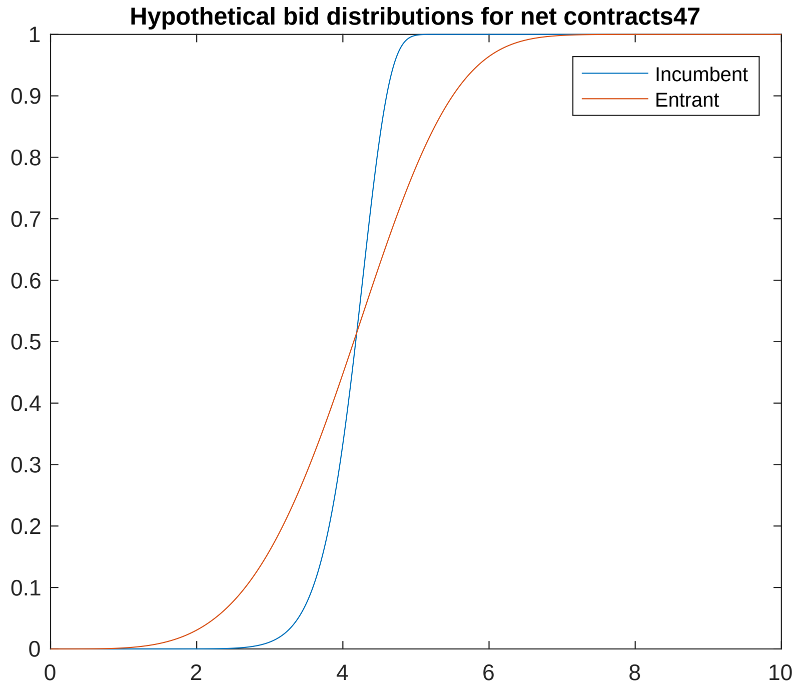


Figure 3.13: Hypothetical bid distribution for net auction line 47



Estimated cost distributions in net auctions

In this section, we present our estimates for the cost distributions for the three net auction tracks displayed previously.

Figure 3.14: Cost distribution for net auction line 26

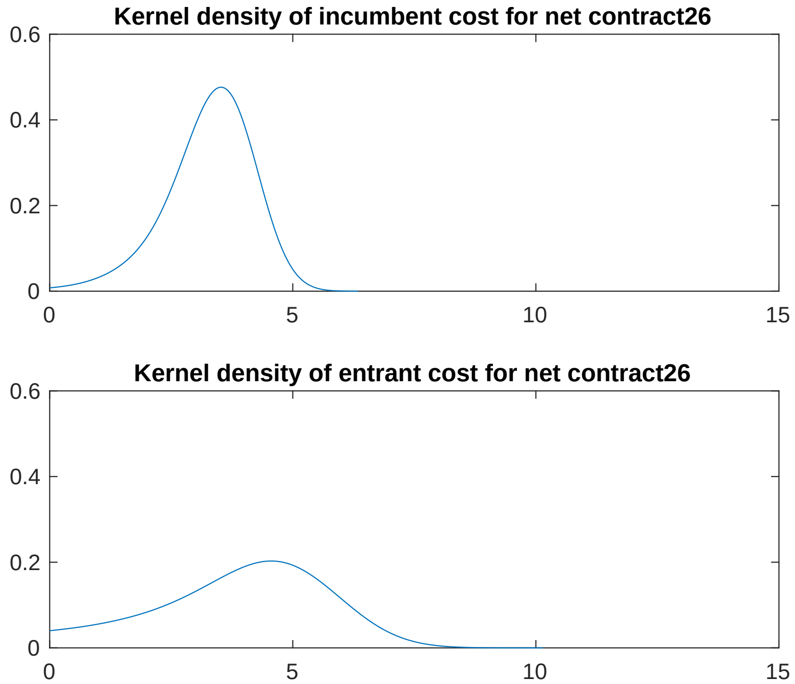


Figure 3.15: Cost distribution for net auction line 46

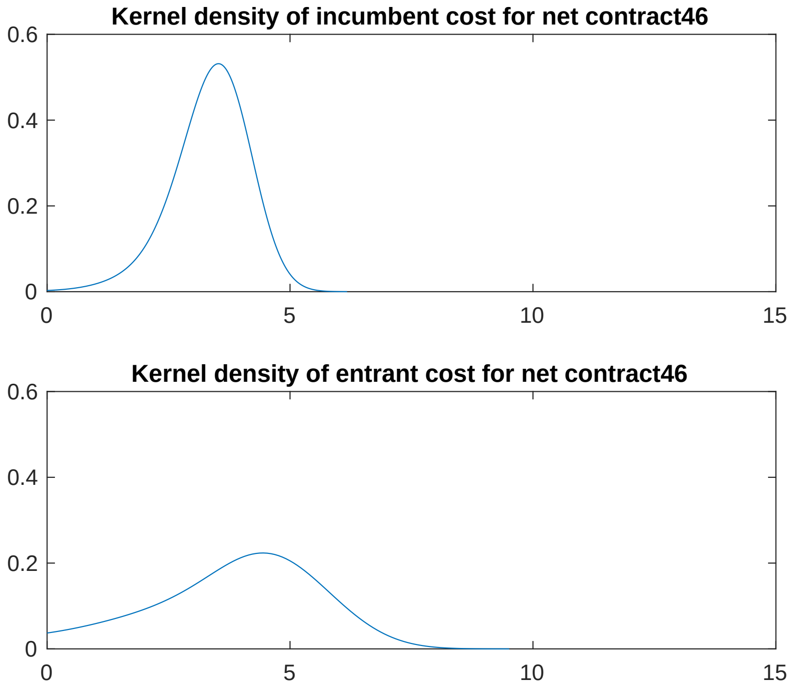
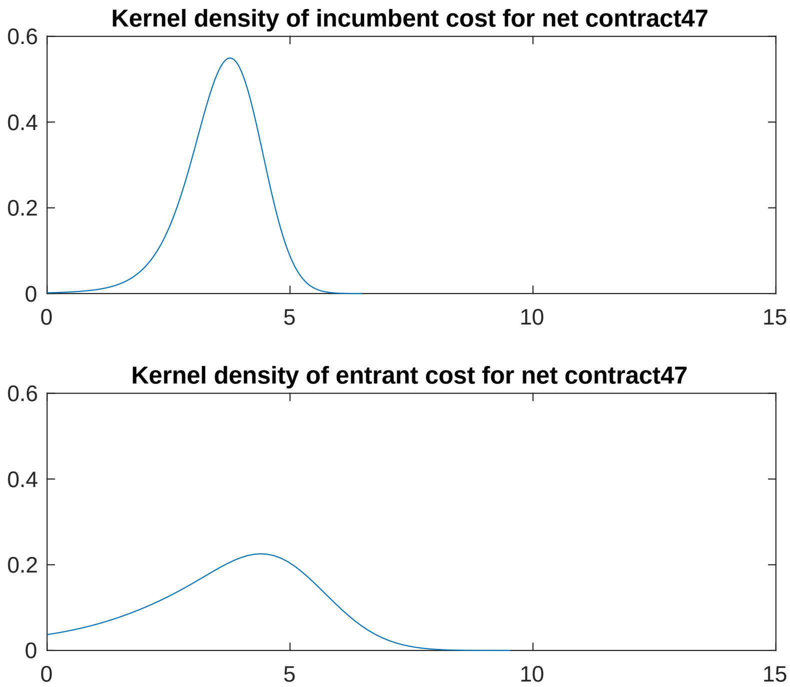


Figure 3.16: Cost distribution for net auction line 47



3.8.2 Detailed estimation results

Table 3.5: Estimation results: bid functions in gross auctions

	Point Estimates	Standard Errors	t-Statistics	p-Values
λ_0^I	0.8107	0.9570	0.8472	0.3969
λ_X^I	2.5096	0.3706	6.7722	0.0000
	0.0121	2.8902	0.0042	0.9966
	0.8513	4.0425	0.2106	0.8332
λ_N^I	0.1554	0.7100	0.2189	0.8267
λ_0^E	2.2926	1.0765	2.1296	0.0332
λ_X^E	2.8809	2.7107	1.0628	0.2879
	-0.7412	10.5867	-0.0700	0.9442
	-1.1131	11.1176	-0.1001	0.9203
λ_N^E	-0.3706	2.1102	-0.1756	0.8606
ν_0^I	10.4298	0.3564	29.2657	0.0000
ν_X^I	0.2138	4.3872	0.0487	0.9611
	0.8138	31.6314	0.0257	0.9795
	-13.2611	43.7319	-0.3032	0.7617
ν_N^I	-2.8742	13.3168	-0.2158	0.8291
ν_0^E	-0.3185	6.2125	-0.0513	0.9591
ν_X^E	0.1677	4.6607	0.0360	0.9713
	1.1258	23.2194	0.0485	0.9613
	2.0834	14.1552	0.1472	0.8830
ν_N^E	0.6701	5.2758	0.1270	0.8989

.

3.8.3 Derivation of Likelihood Function.

The likelihood function derives from the first-order statistic of the winning bid. That is, the probability that the outcome of the auction is that bidder U wins the auction with bid x given the other bidders $\mathcal{N}$. Introduce the following general notation for bidder type U given our parametric Weibull assumption on the bid function:

$$\exp_U = \exp\left(-(\frac{x}{\lambda_U})^{\rho_U}\right) \tag{3.50}$$

$$g_U = \exp\left(-(\frac{x}{\lambda_U})^{\rho_U}\right)(\frac{x}{\lambda^U})^{\rho^U-1} = \exp_U(\frac{x}{\lambda^U})^{\rho^U-1} \tag{3.51}$$

$$G_U = 1 - \exp\left(-(\frac{x}{\lambda_U})^{\rho_U}\right) = 1 - exp_U \tag{3.52}$$

Denote the density function of the first-order statistic of winning bid x from winner U given number of bidders N by $h^U_{b1:N}$. In case the incumbent wins, the likelihood function is derived

Table 3.6: Estimation results: bid functions in net auctions

	Point Estimates	Standard Errors	t-Statistics	p-Values
λ_0^I	-1.5602	0.3893	-4.0082	0.0001
λ_X^I	23.5199	29.2769	0.8034	0.4218
	0.6044	5.0368	0.1200	0.9045
	-2.9337	5.9271	-0.4950	0.6206
λ_N^I	1.5298	2.2238	0.6879	0.4915
λ_0^E	-3.2925	1.4675	-2.2437	0.0249
λ_X^E	66.7781	100.6160	0.6637	0.5069
	-3.2713	27.6500	-0.1183	0.9058
	2.5225	11.1989	0.2252	0.8218
λ_N^E	0.5490	2.8292	0.1941	0.8461
ν_0^I	-2.8274	19.0483	-0.1484	0.8820
ν_X^I	-32.4191	74.5279	-0.4350	0.6636
	-1.7658	42.4943	-0.0416	0.9669
	18.7309	31.6825	0.5912	0.5544
ν_N^I	-0.4042	16.2532	-0.0249	0.9802
ν_0^E	2.4473	13.6525	0.1793	0.8577
ν_X^E	-22.5005	64.9759	-0.3463	0.7291
	4.1126	21.5850	0.1905	0.8489
	-0.7447	36.1394	-0.0206	0.9836
ν_N^E	-0.7632	3.3997	-0.2245	0.8224

from

$$h^I_{b^{1:N}}(x, \text{I wins}) = \Pr(b^I = x, b^{E_1} \geq x, \ldots, b^{E_{N-1}} \geq x) \tag{3.53}$$
$$= \Pr(b^I = x) \Pr(b^{E_1}, \ldots, b^{E_{N-1}}) \tag{3.54}$$
$$= g^I(x)(1 - G^E(x))^{N-1} \tag{3.55}$$

and for an entrant it is given by

$$h^E_{b^{1:N}}(x, \text{ one E wins}) = (N-1)\Pr(b_i^E = x, b^E_{j\neq i} > x, b^I > x) \tag{3.56}$$
$$= (N-1)\Pr(b_i^E = x)\Pr(b^E_{j\neq i} > x, b^I > x) \tag{3.57}$$
$$= (N-1) g^E(x)(1 - G^E(x))^{N-2}(1 - G^I(x)) \tag{3.58}$$

3.8.4 Proof of Lemma 3.2.

The expected payoff of winning with bid b given signal $-c_i + \alpha_i r_i$ is given by

$$\pi_i(b) = \left(b - c_i + \alpha_i r_i + \sum_{j\neq i} \alpha_j \mathbb{E}\left[r_j | \rho_i \geq B_j^{-1}(b) \right] \right) \left(1 - F^{1:N-1}_{\rho_j}(B_j^{-1}(b)\right) \tag{3.59}$$

Table 3.7: Estimation results: asymmetry parameters in net auctions

	Point estimates	Standard errors	t-statistics	p-values
α_2^I	0.6224	0.0140	44.4813	0.0000
α_{3+}^I	0.5480	0.1943	2.8213	0.0048
σ_{r0}	2.8665	0.2855	10.0403	0.0000
σ_{r1}	0.8292	0.1441	5.7530	0.0000
β_{R0}	2.0355	1.5010	1.3561	0.1751
β_{R1}	8.1742	6.0590	1.3491	0.1773
β_{R2}	4.9362	0.7209	6.8476	0.0000

where $F_{\rho_j}^{1:N\backslash i}$ denotes the first-order statistic of opponents' signals. The first-order condition yields

$$0 = -\left(b - c_i + \alpha_i r_i + \sum_{j \neq i} \alpha_j \mathbb{E}\left[r_j | \rho_i \geq B_j^{-1}(b)\right]\right) f_{\rho_j}^{1:N\backslash i}(B_j^{-1})(b) B_j^{'-1}(b)$$

$$+ (1 - F_j^{1:N\backslash i}(B_j^{-1}(b))) \left(1 + \sum_{j \neq i} \alpha_j \frac{f_{\rho_j}^{1:N\backslash i}(B_j^{-1})(b)}{1 - F_j^{1:N\backslash i}(B_j^{-1}(b))} B_j^{'-1}(n)\right.$$

$$\left.\left(\mathbb{E}\left[r_j | \rho_i = B_j^{-1}(b)\right] - \mathbb{E}\left[r_j | \rho_i \geq B_j^{-1}(b)\right]\right)\right)$$

$$0 = -\left(b - c_i + \alpha_i r_i + \sum_{j \neq i} \alpha_j \mathbb{E}\left[r_j | \rho_i = B_j^{-1}(b)\right]\right) f_{\rho_j}^{1:N\backslash i}(B_j^{-1})(b) B_j^{'-1}(b) + (1 - F_j^{1:N\backslash i}(B_j^{-1}(b))).$$

3.8.5 Proof of Lemma 3.3.

We assume that both firms have the same cost level c. Thus, the auction is a pure common value auction with asymmetric signal distributions and symmetric value functions. The signal of firm i is $x_i = c - \alpha_i r_i$ where r_i is independently and identically distributed according to a cdf F_r with support $[c - \alpha_i \overline{r}, c]$. Hence, the signals x_i are independently but not identically distributed with $F_{x_i}(x) = \frac{1}{\alpha_i} F_r(\frac{x}{\alpha_i})$ on $[c - \alpha_i \overline{r}, c]$. It follows from the definition of conditional stochastic dominance that $\frac{1-F_{x_2}(x)}{f_{x_2}(x)} > \frac{1-F_{x_1}(x)}{f_{x_1}(x)}$.[13] To simplify notation we write F_j for F_{x_j}.

We now proceed by deriving the tying function for procurement auctions following the proof in Parreiras (2006) for standard first-price common value auctions. The expected value of bidder i conditional on winning with bid b given signal x_i is given by

$$\int_{\phi_j(b)}^{\infty} (b - x_i - \phi_j(b) + c) \; d(1 - F_j(\phi_j(b)))$$

[13] We cannot guarantee that conditional stochastic dominance is satisfied under our parametric assumptions. However, conditional stochastic is sufficient but not necessary for the result in this Lemma to hold, while FOSD is necessary.

where $\phi_j(b)$ is the inverse bid function. As in Parreiras (2006), an equilibrium in monotone strategies exists and therefore the derivate of the inverse bid function is differentiable almost everywhere. Uniqueness of the equilibrium also follows as in Parreiras (2006). The first-order condition implicitly characterizing the equilibrium is given by

$$\frac{1}{\dot{\phi}_j(b)} = (b - x_i - \phi_j(b) + c)\,\frac{f_j(\phi_j(b))}{1 - F_j(\phi_j(b))}. \tag{3.60}$$

Taking the ratio of the two first-order conditions and applying the function $Q(\phi_1(b)) = \phi_2(b)$ which has derivative $\dot{Q}(\phi_1(b)) = \frac{\dot{\phi}_2(b)}{\dot{\phi}_1(b)}$ yields together with the definition $\phi_1(b) = x$

$$\dot{Q}(x) = \frac{1 - F_2(Q(x))}{f_2(Q(x))} \Big/ \frac{1 - F_1(x)}{f_1(x)} \tag{3.61}$$

with standard boundary condition $Q(\underline{x}) = \underline{x}$. The interpretation of function $Q(x)$ is that it gives the signal of player 2 that places the same bid as player 1 given signal x. Hence, when $Q(x) \geq x$, bidder 2 shades the bid more than player 1. Given the assumption of conditional stochastic dominance, we have that $\dot{Q}(x)|_{Q(\underline{x})=\underline{x}} > 1$. Moreover, whenever $Q(x)$ approaches x, that is, whenever similar signals yield similar bids, $\dot{Q(x)} > 1$ and pushes $Q(x)$ above x. Formally, $\lim_{Q(x)\searrow x} \dot{Q}(x) > 1$.

4 Collaborating under Asymmetric Information

joint with Sinem Hidir

4.1 Introduction

Many work and research activities are carried out in teams and require uncertain time to be completed. Since Holmström (1982), it is well known that partners in a team rely on their peers when their reward depends on the output of the team and not on their individual effort. This is the well-known freeriding result in collaboration games. The extent to which individuals can rely on their peers depends, however, on the information they hold about their productivity. If a player is aware that the others are not productive in completing the project, he can rely less on their progress. If a player is aware that the others are very productive, he has stronger freeriding incentives. However, partnerships are often formed in presence of incomplete information about the ability of a partner. When collaborating, the progress of the project is informative about the ability of the partners. Therefore, the players' knowledge about their peers evolves over time.

A second aspect is that relationships and information about coworkers are typically not symmetric. Asymmetries of contributors to a project are widely observed. In research, coauthors form teams to write papers together to benefit from their joint effort. Although there may be initial information about the coauthor's ability, it might not be perfect information about the productivity in the current project. In particular, there is likely to be little uncertainty about an established researcher who has already published in the area of the project. However, there is only limited information and therefore higher uncertainty about the ability of more junior researchers. In firms, teams consist of several workers some of which have been present for many years and have known ability while others have just recently joined the firm and coworkers are therefore still uncertain about their ability.

We consider a setting in which a player of certain ability (senior) owns a project and forms a partnership with a player of uncertain productivity (junior) to achieve a one time breakthrough before a deadline. The junior knows his own as well as the other's ability, while the senior only knows his own ability. We study this dynamic collaboration game to understand the interaction of dynamic freeriding and learning about the collaborators. With unobservable effort levels, the senior cannot distinguish a productive but shirking junior from an unproductive junior if the project does not progress. Hence, the absence of a success is inconclusive news for the senior: he adjusts his belief about the junior's ability downwards and becomes more pessimistic. Increasing pessimism implies that the senior's freeriding incentives are decreasing over time and, therefore, he will increase his effort over time. This, however, affects the junior's incentive to exert effort: if the senior is expected to exert more effort, he can freeride on this additional effort and reduce his own. When the deadline for project completion approaches, both, the senior and the productive junior, will increase their efforts. We show that the junior's effort path can be non-monotone over time: initially he exerts high effort because the effort increase due to the more pessimistic senior in the future is discounted. When the time of high effort by the senior approaches, the junior reduces his effort and freerides more. Finally, a deadline effect kicks in and both players will increase their effort levels.

In equilibrium, as a consequence of the uncertainty of the senior, the junior benefits from higher contributions of the senior. Considering the tradeoff before a team is built, the productive junior would want the senior to be sufficiently optimistic to form a team with the junior. Conditional on being in the team, the junior prefers the senior to be as pessimistic as possible. A more pessimistic initial belief about the junior reduces the incentives of the senior to freeride at each instant and therefore increases his total effort level. Taking this to the extreme, it would be optimal not to tell workers that there are others working on the same project. By means of numerical examples, we show that it can be beneficial from a planner's perspective to keep the senior uncertain about his collaborator. In particular, it can be better to draw a junior from a distribution that has a positive probability of selecting an unproductive junior, than choosing with certainty a productive junior. However, this logic ignores any form of synergies in the player's efforts and therefore, we plan to extend the model to allow for synergies. In a weaker sense, higher effort levels can be obtained by changing the composition of groups in sequential projects to reduce the information about the peers within the team.

Literature. Our paper contributes to the literature on dynamics in teams. While it relates to the literature on team experimentation as in the seminal papers Bolton and Harris (1999) and Keller, Rady and Cripps (2005), learning in the present paper is not about the project's quality but rather about the productivity of the collaborators. The most closely related paper is Bonatti and Hörner (2011) which studies a dynamic collaboration game in which both players are productive but there is common uncertainty about the project's feasibility. In the present paper, effort is also unobservable but the uncertainty is about the value of the contribution of one of the players. In contrast to Bonatti and Hörner (2011) this yields to asymmetric freeriding incentives that change over time. Our paper is also related to Guo and Roesler (2016) who study dynamic collaboration on a project of uncertain feasibility. In their paper, players can privately learn that the project is of bad quality and therefore learn privately the value of their effort. However, as learning is about the project quality, there is no asymmetry in the benefit of effort by each player but on the private belief about the benefit of individual and collective effort.

Georgiadis (2014) considers a team problem without asymmetries across players. In his setting, the project progresses gradually at the rate of players' effort levels. Gradual progress induces efforts - in contrast to our setup - to be strategic complements instead of strategic substitutes over time. All of these papers, including ours, relate to the large literature on moral hazard in teams.[1] In contrast to this older literature, we consider a stochastic production technology and asymmetric information about the value of each player's contributions to the team.

There is a small literature on static contribution games which addresses the question of the effect of uncertainty on individual contributions, e.g. Sandler, Sterbenz and Posnett (1987), Bramoull and Treich (2009) These papers show that uncertainty can increase efficiency. To the best of our knowledge, we are the first ones to consider how dynamic freeriding, asymmetric information about team members and learning interact.

Georgiadis (2015) studies the optimal composition and incentive contracts of a team. This is an extension we are planning to consider in our model in the presence of asymmetric information and efforts being strategic substitutes over time as in Bonatti and Hörner (2011).

4.2 Two-Period Example

To build intuition, consider first a two-period example in discrete time. There are two juniors, a senior and a junior, $i \in \{s, j\}$, that collaborate in two periods, $t \in \{1, 2\}$. The project's success depends on the players' effort levels, u_{it} and their productivity. The junior can either be productive or unproductive. If he is unproductive, his effort never increases the success probability and therefore, this type never puts effort. The junior's productivity is his private information and the senior believes that the junior is productive with probability μ_1 in period $t = 1$. The success probability is given by the sum of the effort levels, $\sum_{i\in\{s,j\}} u_{it}$. The effort levels of the players are unobservable. Completing the project gives a payoff v to the players and ends the game. Effort costs are quadratic, cu_{it}^2. Second stage payoffs are discounted at a common factor $\delta \leq 1$.

The absence of a success makes the senior more pessimistic about the junior's productivity. This follows from the probability of not observing a success being higher if the junior's productivity-weighted effort is lower which is the case for the low productivity type. As effort levels are not directly observed, players hold beliefs about the other player's effort choices, $\hat{u}_{it}$. The expected payoffs for the senior is given by

$$V^s = \underbrace{v\left(u_{s1} + \mu_1 \hat{u}_{j1}\right) - \frac{cu_{s1}^2)}{2}}_{\text{Payoff in } t=1} + \delta(1 - u_{s1} - \mu_1 \hat{u}_{j1}) \underbrace{\left((u_{s2} + \mu_2(u_{s1}, \hat{u}_{j1})\hat{u}_{j2})v - \frac{cu_{s2}^2}{2}\right)}_{\text{Payoff in } t=2} \tag{4.1}$$

[1]See, for example, Holmstrom (1982), Legros and Matthews (1993), Strausz (1999).

and for the junior by

$$V^j = \underbrace{v\left(\hat{u}_{s1} + u_{j1}\right) - \frac{cu_{j1}^2)}{2}}_{\text{Payoff in } t=1} + \delta(1 - \hat{u}_{s1} - u_{j1}) \underbrace{\left((\hat{u}_{s2} + u_{j2})v - \frac{cu_{j2}^2}{2}\right)}_{\text{Payoff in } t=2} \tag{4.2}$$

Note that the first-stage actions affect the second stage through their effect on the belief. Bayes' rule gives for the second period belief of the senior

$$\mu_2(\hat{u}_{j1}, \hat{u}_{s1}) = \frac{(1 - \hat{u}_{j1} - u_{s1})\mu}{1 - \mu_1 \hat{u}_{j1} - u_{s1}} \tag{4.3}$$

which is (weakly) less than μ_1. This depends on the senior's beliefs about the junior's effort and off the equilibrium path this belief may be misspecified.

We solve for the perfect Bayesian equilibrium of this game. In equilibrium, the effort levels of the players must be optimal given the beliefs and the expectation about the player's choices. Moreover, these expectations are correct; hence, $u_{it} = \hat{u}_{it}$. Then, the first-order conditions in the second period give

$$u_{it}^* = \frac{v}{c}. \tag{4.4}$$

In the last period, the junior's chooses the same effort level which is the static optimum. In the first stage, however, the first-order conditions deliver for the senior

$$\frac{\partial V^s}{\partial u_{1s}} = v - cu_{1s} - \delta \left((u_{s2}^* + u_{j2}^* \mu_1)v - \frac{cu_{s2}^{*2}}{2} \right) = 0 \tag{4.5}$$

$$\Leftrightarrow u_{1s}^* = \frac{v}{c} \left(1 - \delta(\frac{1}{2} + \mu_1)\frac{v}{c} \right) \tag{4.6}$$

and for the junior

$$\frac{\partial V^j}{\partial u_{1j}} = v - cu_{1j} - \delta \left((u_{2j}^* + u_{2s}^*)v - \frac{cu_{j2}^{*2}}{2} \right) \tag{4.7}$$

$$\Leftrightarrow u_{1j}^* = \frac{v}{c} \left(1 - \delta \frac{v}{c} \frac{3}{2} \right) < u_{1s}^*. \tag{4.8}$$

We find that, due to the asymmetric information about the junior's type, the senior puts in more effort into the project in the first period because he fears that if no success is obtained in the current period, he is more likely to face an unproductive junior. This difference in effort levels increases the more pessimistic the senior is. In addition to this effect of asymmetric information, there is the standard procrastination effect:[2] The effort level in period one is lower than the static optimum which is equal to the effort level in the second period. One interesting feature is

[2] See Bonatti and Hörner (2011).

that compared to the full information case, the total effort is higher because the senior faces the risk of working alone on the project.

It is evident that the junior benefits from the senior being more pessimistic about his type: the more pessimistic the senior is, the more effort the senior will exert in the first period. A two-period model does not suffice to analyze the effect of a potentially more pessimistic senior in the future on the incentives of the junior today. The junior knows that - if he exerts effort today - with a positive probability he faces a more pessimistic senior tomorrow who will exert more effort. This, however, reduces incentives to exert effort today in order to freeride on the higher effort level of the senior tomorrow. To address these dynamics, we consider a continuous-time model in the following sections.

4.3 Model

We study a dynamic collaboration game under asymmetric information. There are two players, $i \in \{s, j\}$, a senior and a junior. They collaborate in continuous time, $t \in [0, T]$, to complete a project before a deadline T. The probability that the project is completed in time interval $[t, t + dt)$ depends on the players' unobserved effort levels $u^{\omega}_{i,t}$ and their ability $\lambda_{i\omega}$. While the senior's productivity is known to be $\lambda_s = 1$, the junior's productivity is private information and $\lambda_\omega \in \{\lambda_h, \lambda_l\}$. The senior's intial belief about the junior being of high productivity is denoted by μ_0. Given a player's identity and type, the instantaneous success probability of a success is given by $\sum_{i\in\{s,j\}} u_{it}\lambda_{i\omega}$. In the absence of a success, the senior's belief evolves according to the following differential equation

$$\dot{\mu}_t = -\mu_t(1 - \mu_t)(\hat{u}^h_{jt}\lambda_{jh} - \hat{u}^l_{jt}\lambda_{jl}). \tag{4.9}$$

where $\hat{u}^{\omega}_{jt}$ refers to the senior's belief about the junior's unobserved effort choices.

If the project is completed, the senior receives value v^s and the junior receives value v^j. If the project is not completed before the deadline, both players receive a value of zero. In the current version, we restrict attention to the case where only the high productivity type of the junior member of the team can succeed. Hence, we assume that $\lambda_{jh} > \lambda_{jl} = 0$. This implies that the unproductive type of the junior will never exert effort, $u^l_{jt} = 0$. We simplify notation and write $u^h_{jt} = u_{jt}$ and can neglect the unproductive junior's problem in the analysis as it becomes trivial. As a consequence, when we refer to the junior, the productive type of the junior is meant. Moreover, we normalize the productivities to be equal to 1 $\lambda_s = \lambda_{jh} = 1$. We assume that effort cost are convex with $c_i u^2_{it}$ for player $i \in \{s, j\}$. Both players discount the future at the common discount rate r.

Hence, the payoff of player i is given by

$$V^i = \int_0^T e^{-rt}\, \Pr(\text{no success before } t)\, (\Pr(\text{success at } t)v^s - \text{cost of effort})\, dt. \tag{4.10}$$

The probability that no success has been obtained before time t is given by $e^{-\int_0^t (u_{s\tau}+u_{j\tau})d\tau}$ from the productive junior's perspective. The senior only holds a belief about the junior's type. Therefore, from his perspective, the probability that no success has been obtained before time t is given by $e^{-\int_0^t (u_{s\tau}+\mu_\tau u_{j\tau})d\tau}$. The probability of obtaining a success in $[t, t+dt)$ is given by $(u_{st} + u_{jt})$ from the junior's perspective and $(u_{st} + \mu_t u_{jt})$ from the senior's perspective. Denote for now the cost of effort simply by $c(u_{it})$. The senior's objective function is then given by

$$V^s = \int_0^T e^{-rt} e^{-\int_0^t (u_{s\tau}+\mu_\tau u_{j\tau})d\tau} \left((u_{st} + \mu_t u_{jt})v^s - c^s \frac{u_{st}^2}{2} \right) dt \tag{4.11}$$

while the junior's objective function is given by

$$V^j = \int_0^T e^{-rt} e^{-\int_0^t (u_{s\tau}+u_{j\tau})d\tau} \left((u_{st} + u_{jt})v^j - c^j \frac{u_{jt}^2}{2} \right) dt. \tag{4.12}$$

The players maximize their total discounted payoff, that is they choose an effort path $\{u_{it}\}_{t\geq 0}$ to maximize their objective function.

4.4 Analysis

4.4.1 Complete-Information Benchmark

As a benchmark, we consider first the outcome under complete information. In this case, the senior knows whether he is working by himself (i.e. that the junior team member is unproductive) or whether he is working together with a productive junior team member.

Cooperative Solution. To build a first intuition, suppose that both players were to cooperatively maximize their payoffs. If the junior is unproductive, there is only one active player and the cooperative solution coincides with the noncooperative solution in Lemma 2. If the junior is productive, both players exert effort to maximize the sum of their payoffs. The following Lemma describes the equilibrium behavior in this case.

Lemma 4.1. *The cooperative solution is given by the following systems of differential equations*

$$\begin{aligned} \dot{u}_{it} &= r(u_{it} - \frac{v^i}{c^i}) + \frac{u_{it}^2}{2}(1 + \frac{c^{-i}}{c^i}) \\ u_{iT} &= \frac{v^i}{c^i}. \end{aligned}$$

We find that even in the cooperative solution players do not perfectly smooth their effort over time as might be expected with convex costs. The delay of effort can be understood as follows:[3] The success time given total effort is an exponentially distributed random variable. However, ex ante players are uncertain about the realization of this random variable and it might very well be possible that a low effort level generates a success. Hence, players initially exert low effort to save effort cost with a positive probability. As time passes without a success, player update their beliefs such that it is more likely that the realization of the random variable is high. As a result, they have to increase their effort levels due to the approaching deadline. This effect is present in all the variants of the model that we consider.

Noncooperative solution. Assume first that the senior knows that that the junior is unproductive, i.e. he is working by himself. In this case, he solves the following problem[4]

$$\max_{\{u_t\}_{t\geq 0}} \int_0^T e^{-rt-\int_0^t u_\tau d\tau}(u_t v^s - c^s u_t^2)dt. \tag{4.13}$$

The senior's incentives in this case are driven by an incentive to smooth effort over time due to the convexity of cost, discounting and the option value of succeeding in the next instant.

Lemma 4.2. *The senior's effort path if he faces an unproductive junior is described by the following differential equation*

$$\dot{u}_t = r(u_t - \frac{v^s}{c^s}) + \frac{u_t^2}{2} \tag{4.14}$$

with boundary condition $u_T = \frac{v^s}{c^s}$.

The senior's effort in this case is increasing over time until it hits the myopically optimal effort level at the deadline. That effort is increasing over time is due to the reasons discussed in the cooperative solution. Discounting reduces the incentive to delay effort and makes the effort flatter. If the senior is fully myopic, he will choose $u_t = \frac{v^s}{c^s}$ at each instant.

If the senior is certain to be collaborating with a productive junior, both players play a dynamic contribution game. In this case, both solve a similar problem only being different in the value of a success and their cost level. The maximization problem is given by

$$\max_{u_{it}} \int_0^T e^{-rt-\int_0^t (u_{j\tau}+u_{s\tau})d\tau}((u_{jt}+u_{st})v^i - \frac{c^i u_{it}}{2})dt. \tag{4.15}$$

In this case, there is an additional motif to procrastinate present. A success could also be obtained by the other team member and a player could save his own cost of effort. The presence of a team member therefore creates the familiar freeriding incentive.

[3]This logic has been presented in Kuelpmann (2016).

[4]Abusing notation we write $u_{st} = u_t$ when there is no confusion.

Lemma 4.3. *Equilibrium effort levels in the complete information two player collaboration game are described by the following differential equations with boundary conditions*

$$\dot{u}_{it} = r(u_{it} - \frac{v^i}{c^i}) + u_{it}\left(u_{-it} + \frac{u_{it}}{2}\right)$$
$$u_{iT} = \frac{v^i}{c^i}.$$

We see that the senior's effort level in this case is everywhere lower than if he is working alone. This is driven by the opportunity to freeride on the productive junior. The senior has an incentive to delay effort to potentially save on the cost of effort because the junior's effort might suffice to complete the project. Moreover, the effort is, exactly for the same reason, lower than in the cooperative solution. If players maximize their payoffs cooperatively they internalize the negative externality on the collaborator's payoffs and hence do not freeride on their successes.

4.4.2 Asymmetric Information

In the case of asymmetric information the senior is not certain about the productivity of his junior team member. However, he holds a belief about his productivity denoted by μ_t. As the absence of a success is weak evidence for a low productivity junior, the belief is continuously decreases when the project is not completed and the productive junior is expected to exert effort. The belief follows the law of motion

$$\dot{\mu}_t = -\mu_t(1 - \mu_t)u_{jt}. \tag{4.16}$$

Note that the junior's actions in the past determine the current belief of the senior and therefore his current effort. The current effort of the senior, however, affects the junior's choices today. Hence, there is a non-trivial intertemporal linkage of the junior's effort choices. The payoff of the senior in the setting with asymmetric information is given by

$$\int_0^T e^{-rt-\int_0^t u_{s\tau}d\tau-\int_0^t \mu_\tau u j\tau d\tau}\left((u_{st} + \mu_t u_{jt})v^s - \frac{c^s u_{st}^2}{2}\right)dt \tag{4.17}$$

and the payoff of the junior by

$$\int_0^T e^{-rt-\int_0^t (u_{s\tau}+u_{j\tau})d\tau}\left((u_{st} + u_{jt})v^j - \frac{c^j u_{jt}^2}{2}\right)dt. \tag{4.18}$$

In a perfect Bayesian equilibrium, both players maximize their payoffs given their conjectures about the other player's actions and the belief of the senior which is upated according to Bayes rule. The following Proposition that describes the unique equilibrium is proved using Pontryagin's principle. The proof is carried out in the Appendix.

Proposition 4.1. *The unique perfect Bayesian equilibrium of the game with asymmetric infor-*

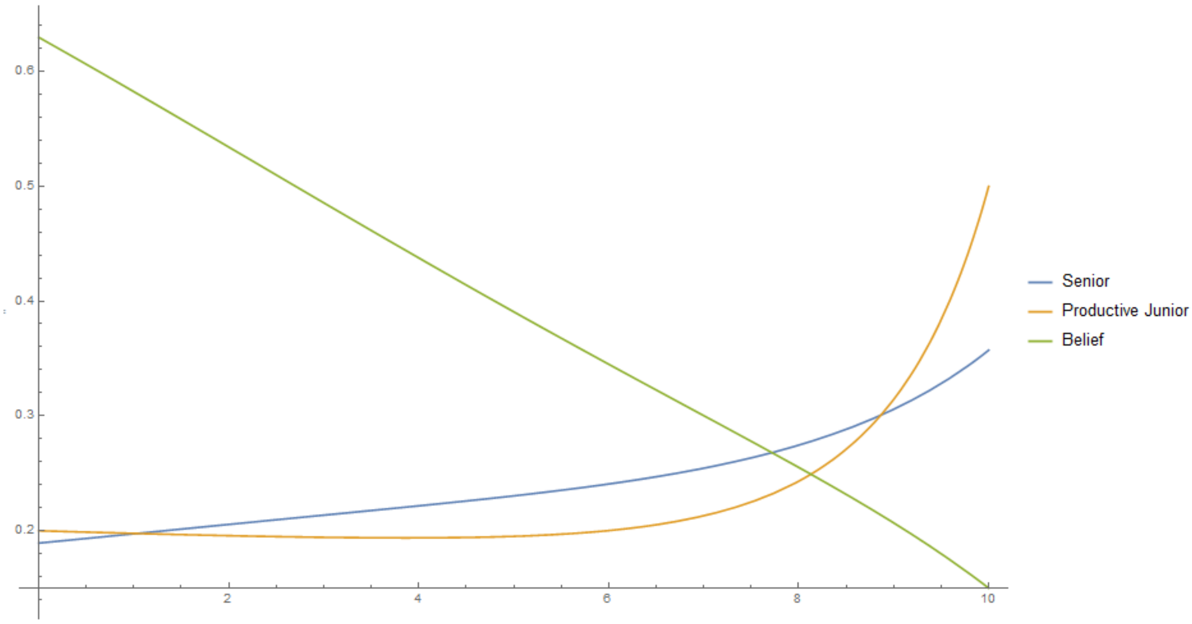


Figure 4.1: Equilibrium effort paths of Senior and Productive Junior. *Parameter values:* $r = 0.2,\ v^s = v^j = 2.5,\ c^s = 7,\ c^s = 5,\ \mu_0 = 0.63,\ T = 10.$

mation is described by the following system of differential equations with boundary conditions

$$\begin{aligned}
\dot{u}_{jt} &= r(u_{jt} - \frac{v^j}{c^j}) + u_{jt}\left(\frac{u_{jt}}{2} + u_{st}\right) \\
u_{jT} &= \frac{v^j}{c^j} \\
\dot{u}_{st} &= r(u_{st} - \frac{v^s}{c^s}) + u_{st}\left(\mu_t u_{jt} + \frac{u_{st}}{2}\right) \\
u_{sT} &= \frac{v^s}{c^s}.
\end{aligned}$$

It follows from the Proposition that the uncertainty puts upward pressure on the senior's effort. That is, compared to full information, the effort path is flatter and the senior exerts more effort. In return, this makes the junior's effort path steeper and therefore reduces his effort. Hence, the junior benefits from uncertainty while the senior suffers from uncertainty.

One interesting feature of the equilibrium is that the effort path of the junior can be non-monotone while the senior's effort is increasing over time as illustrated in Figure 1. In the example of the figure, the productive junior is the more efficient contributor to the project as his cost is lower. Therefore, he will exert more effort close to the deadline than the senior. Initially, when the senior is still relatively optimistic, the junior exerts more effort than the senior. However, the junior knows that the senior becomes increasingly pessimistic and therefore will increase his effort. This increase in the senior's effort reduces the junior's incentives to exert effort today. Therefore, the junior's effort can be decreasing when the discount factor is strictly above zero. Without discounting, the junior will smooth the effort over time and not exert a higher effort level initially because the senior's high future effort is discounted.

The senior's effort level increases if he is initially more pessimistic about the junior's ability. In return, the junior decreases his effort even further and has an increased freeriding incentive. Hence, the junior benefits from a more pessimistic senior team member. That is, the junior faces a *ratchet effect*-like incentive when entering the team: he wants the senior to be sufficiently optimistic to have him in the project. However, conditional on being in the team, the junior wants the senior to be pessimistic about his productivity.

In the case of a more productive junior, as illustrated in Figure 4.1, one can interpret the equilibrium as follows: the senior, who has high effort cost, for example, due to opportunity cost, lets the junior work on a project initially. However, if he does not observe progress, he becomes more and more involved in the project to make up for the (in expectation) less productive junior team member. This benefits the junior and he can reduce his effort in return and (partially) freerides on the senior's effort. As the project deadline approaches, both increase their efforts substantially to complete the project.

We can show numerically that it can be beneficial for a social planner to select an unproductive junior to collaborate with the senior with positive probability compared to selecting an unproductive junior always. This is exactly due to the reduced freeriding incentive of the senior if he is uncertain about the junior's productivity.

Both Players Uninformed. If both players are uninformed about their types a similar intuition as before now also applies to the junior team member. With some, and over time increasing, probability he is facing an unproductive senior. Therefore, his freeriding incentives are reduced an he puts in more effort than under certainty.

Lemma 4.4. *The equilibrium when both players are uninformed about their team member's ability is defined by the following system of differential equations together with boundary conditions*

$$
\begin{aligned}
\dot{u}_{jt} &= r(u_{jt} - \frac{v^j}{c^j}) + u_{jt}\left(\mu_t^j u_{st} + \frac{u_{jt}}{2}\right) \\
u_{jT} &= \frac{v^j}{c^j} \\
\dot{u}_{st} &= r(u_{st} - \frac{v^s}{c^s}) + u_{st}\left(\mu_t^s u_{jt} + \frac{u_{st}}{2}\right) \\
u_{sT} &= \frac{v^s}{c^s}.
\end{aligned}
$$

Hence, also the senior member benefits from keeping the junior uninformed about his type because this reduces freeriding incentives.

4.4.3 Implications of the Different Information Structures

The previous subsections have shown that the player whose identity is unknown benefits from this asymmetric information. This is because in the presence of perfect information his collaborator has an incentive to freeride on his effort. However, if the uninformed player is uncertain whether he faces a productive team member, his freeriding incentive is reduced and he exerts more effort himself.

It follows that to maximize overall effort it is optimal to have two players that believe to be working by themselves to complete the project. In this case, they do not freeride on the other collaborator's effort and will therefore behave as if they were working by themselves. Hence, in this model, uncertainty about team members can actually be welfare enhancing.

4.5 Discussion

In addition to the results that have already been established, there are multiple interesting ways to extend the setup.

Deciding on a Collaboration and Delegation. It seems natural to consider a situation in which the senior can decide whether to start a collaboration, to delegate the project to a junior or to leave the relationship when he gets too pessimistic. We plan to carry out these comparisons and give conditions under which each of these scenarios is optimal. The value of delegating the task to the junior is that the junior cannot freeride on a pessimistic senior and will therefore increase the own effort. However, this comes at the cost of reducing the total potential effort. The value of working alone can be that the benefits of the project do not have to be shared with the junior.

Continuing Relationships. In the current setup there is no value for the senior of learning the junior's type because the game ends after a success. However, it might be that players work on sequential projects. This introduces additional incentives to learn about the junior's type, for example, by delegating one task with a deadline to learn his type. As successes do not perfectly reveal the junior's type, the junior faces the tradeoff to choose his actions in a way to keep the senior sufficiently optimistic to continue collaborating with him, while conditional on continuation he wants him to be more pessimistic so that he exerts more effort.

Collaboration Design. Similar to Georgiadis (2015) it would be interesting to study the optimal contract design from the senior's point of view. He could determine a value-sharing rule as function of the success time. Also, the optimal composition of a team with multiple members would be an interesting variant of the model. So far, we have assumed that the low productivity type of the junior never exerts effort. However, if he would, the senior might be able to screen juniors with different sharing rules and deadlines as in Halac, Kartik and Liu (2016) which studies a principal-agent experimentation problem with adverse selection.

Additional Signal Technology. We are considering a very special learning technology. However, in practice, workers have multiple ways to learn about a collaborator through unrelated work, advise or potentially even breakdowns. We plan to consider the presence of non-terminal signals, which can arrive as a result of effort throughout the project and are informative about the junior's type. This would allow more active learning of the senior about the junior. One could also consider this to be an additional effort investment. The senior's incentives to invest into active learning might be changing over time with limited resources: initially, he is more optimistic and can prevent exerting high effort at low beliefs by investing early in the relationship. Later in the relationship, when he is pessimistic and the deadline approaches, the opportunity cost of investing into learning may become too high. Additionally, this may be a way for the junior to signal his type if the senior can decide to leave the relationship. If the senior's belief is very low the junior might reduce his effort in the project and instead try to generate information to prove his ability and make the senior stick with the team.

Synergies One important assumption in our model is that we do not allow for synergies in the player's efforts. This is realistic in many team settings. However, the interaction of synergies with asymmetric information is not straightforward. If the senior believes to be interacting with a high productivity junior, he will believe that his effort has a higher value due to the synergies. In contrast, if he thinks that he is facing a low productivity worker, his effort is valuable due to lower expected synergies. Hence, synergies may mitigate the effect of asymmetric information on freeriding incentives and change the dynamics of the interaction.

4.6 Conclusion

Asymmetric information about the productivity of peers in a dynamic collaboration setting are frequent. Often, the productivity of new team members is not known and there is only some initial information about the new collaborator. However, if the partnership evolves, players learn about their team members and may therefore adjust their actions to the new information.

We study a dynamic game between two players in which one player's productivity is common knowledge while the other player's productivity is his private information. We show that the asymmetric information affects the freeriding incentives of the players: the uninformed player will increase his effort because he faces the risk of a low productivity collaborator. This in turn increases the freeriding motif of the informed player. He knows that if the project is not completed today that the uninformed player will be more pessimistic and therefore work harder tomorrow. Hence, the informed player has an incentive to reduce his effort. By means of numerical examples we can show that introducing asymmetric information into a dynamic partnership game can increase the overall ex ante expected effort levels by reducing freeriding incentives.

4.7 Appendix

4.7.1 Proof of Lemma 4.1

The Hamiltonian is given by

$$\mathcal{H}^W = e^{-rt}e^{-U_t}\left((u_{st}+u_{jt})v - \frac{u_{st}^2 c^s}{2} - \frac{u_{jt}^2 c^j}{2}\right) + \eta_t(u_{st}+u_{jt}) \tag{4.19}$$

where, again, $U_t = \int_0^t (u_{s\tau}+u_{j\tau})d\tau$. This gives the following necessary conditions

$$\eta_t = -e^{-rt}e^{-U_t}\left(v - u_{st}c^s\right) \tag{4.20}$$

$$\eta_t = -e^{-rt}e^{-U_t}\left(v - u_{jt}c^j\right) \tag{4.21}$$

$$\dot{\eta}_t = e^{-rt}e^{-U_t}\left((u_{st}+u_{jt})v - \frac{u_{st}^2 c^s}{2} - \frac{u_t^2 c^j}{2}\right) \tag{4.22}$$

$$\eta_T = 0. \tag{4.23}$$

This implies that $u_{st} = \frac{c^j}{c^s}u_{jt}$ Differentiation of the first-order conditions with respect to time delivers

$$\dot{\eta}_t = e^{-rt}e^{-U_t}\left((v^s - u_{st}c^s)(r+u_{st}+u_{jt}) + c^s\dot{u}_{st}\right) \tag{4.24}$$

$$\dot{\eta}_t = e^{-rt}e^{-U_t}\left(\left(v^j - u_{st}c^s\right)(r+u_{st}+u_{jt}) + c^j\dot{u}_{jt}\right) \tag{4.25}$$

applying the law for the costate gives

$$\left((u_{st}+u_{jt})v - \frac{u_{st}^2 c^s}{2} - \frac{u_{jt}^2 c^j}{2}\right) = \left(\left(v^i - u_{it}c^i\right)(r+u_{st}+u_{jt}) + c^i\dot{u}_{it}\right) \tag{4.26}$$

$$\dot{u}_{st} = r(u_{st} - \frac{v^s}{c^s}) + u_{st}(u_{st}+u_{jt}) - \frac{u_{st}^2}{2} - \frac{u_{jt}^2 c^j}{2c^s} \tag{4.27}$$

$$\dot{u}_{jt} = r(u_{jt} - \frac{v^j}{c^j}) + u_{jt}(u_{st}+u_{jt}) - \frac{u_{st}^2 c^s}{2c^j} - \frac{u_{jt}^2}{2}. \tag{4.28}$$

Now recall that $u_{it} = \frac{c^{-i}}{c^i}u_{-it}$. Applying this to the differential equations delivers

$$\dot{u}_{it} = r(u_{it} - \frac{v^i}{c^i}) + \frac{u_{it}^2}{2}(1 + \frac{c^{-i}}{c^i}) \tag{4.29}$$

$$u_{iT} = \frac{v^i}{c^i}. \tag{4.30}$$

4.7.2 Proof of Lemma 4.2

Consider the case that the senior works alone. Then, his Hamiltonian is given by

$$\mathcal{H}^s = e^{-rt}e^{-U_t^s}\left(u_{st}v^s - \frac{c^s u_{st}^2}{2}\right) + \xi_t^s u_{st} \tag{4.31}$$

where $U_t^s \equiv \int_0^t u_{s\tau}d\tau$. The co-state evolution is given by

$$\dot{\xi}_t = e^{-rt}e^{-U_t^s}\left(u_{st}v^s - \frac{c^s u_{st}^2}{2}\right) \tag{4.32}$$

with boundary condition $\xi_T = 0$. The first-order conditon for effort is given by

$$-e^{-rt}e^{-U_t^s}\left(v^s - c^s u_{st}\right) = \xi_t \tag{4.33}$$

Differentiating this with respect to time delivers

$$\dot{\xi}_t = -e^{-rt-U_t^s}\left(r\left(v^s - c^s u_{st}\right) + u_{st}\left(v^s - c^s u_{st}\right) - c^s \dot{u}_{st}\right) \tag{4.34}$$

and equating this with the co-state evolution yields

$$\dot{u}_{st} = r\left(u_{st} - \frac{v^s}{c^s}\right) + \frac{u_{st}^2}{2}. \tag{4.35}$$

The boundary condition is given by $u_{sT} = \frac{v^s}{c^s}$ due to the transversality condition that $\xi_T = 0$.

4.7.3 Proof of Lemma 4.3

Consider the case when the senior and the productive junior work together and the senior knows that he is collaborating with the productive junior. Then, the Hamiltonian of senior and junior are given by

$$\mathcal{H}^i = e^{-rt}e^{-U_t}\left((u_{jt} + u_{st})v^i - \frac{c^i u_{it}^2}{2}\right) + \eta_t^i(u_{st} + u_{jt}) \tag{4.36}$$

where $U_t \equiv= \int_0^t (u_{j\tau} + u_{s\tau})d\tau$. The necessary conditions are then given by

$$\dot{\eta}_t^i = e^{-rt}e^{-U_t}(u_{jt} + u_{st})\left(v^i - \frac{c^i u_{it}^2}{2}\right) \tag{4.37}$$

$$\eta_t^i = -e^{-rt}e^{-U_t}\left(v^i - c^i u_{it}\right) \tag{4.38}$$

$$\eta_T^i = 0 \tag{4.39}$$

Differentiating the latter condition with respect to time delivers

$$\dot{\eta}_t^i = e^{-rt}e^{-U_t}\left(r(v^i - c^i u_{it}) + (u_{jt} + u_{st})\left(v^i - c^i u_{it}\right) + c_i \dot{u}_{it}\right) \tag{4.40}$$

Equating this with the co-state evolution gives the equilibrium

$$\dot{u}_{it} = r(u_{it} - \frac{v^i}{c^i}) + u_{it}\left(u_{-it} + \frac{u_{it}}{2}\right) \tag{4.41}$$

together with the boundary condition $u_{iT} = \frac{v^i}{c^i}$ following from the transversality condition $\eta_T^i = 0$.

4.7.4 Proof of Proposition 4.1

The junior's Hamiltonian is given by

$$\mathcal{H}^j = e^{-rt}e^{-\int_0^t (u_{j\tau}+u_{s\tau})d\tau}\left((u_{jt} + u_{st})v^j - \frac{c^j u_{jt}^2}{2}\right) + \eta_t(u_{st} + u_{jt}) \tag{4.42}$$

denoting the state as follows $U_t \equiv \int_0^t (u_{j\tau} + u_{s\tau})d\tau$ with law of motion $\dot{U}_t = (u_{st} + u_{jt})$.

The junior's necessary conditions are then given by:

$$\eta_t = -e^{-rt}e^{-U_t}\left(v^j - c_j u_{jt}\right) \tag{4.43}$$

$$\dot{\eta}_t = e^{-rt}e^{-U_t}\left((u_{st} + u_{jt})v^j - \frac{c^j u_{jt}^2}{2}\right) \tag{4.44}$$

$$U_0 = 0 \tag{4.45}$$

$$\eta_T = 0 \tag{4.46}$$

The senior's Hamiltonian is given by

$$e^{-rt}e^{-\int_0^t (u_{s\tau}+\mu_\tau u_{j\tau})d\tau}\left((u_{st} + \mu_t u_{jt})v^s - \frac{c^s u_{st}^2}{2}\right) \tag{4.47}$$

$$+ \xi_t^s u_{st} + \xi_t^j u_{jt} \tag{4.48}$$

Using the belief evolution $\dot{\mu}_t = -\mu_t(1-\mu_t)u_{jt}$ yields $\mu_t u_{jt} = \frac{-\dot{\mu}_{jt}}{1-\mu_t}$. Noting that $\int_0^t \frac{-\dot{\mu}_{js}}{1-\mu_s}ds = \ln(1-\mu_t)$ yields after using the formula for the posterior $\mu_t = \frac{\mu_0 e^{-\int_0^t u_{js}ds}}{\mu_0 e^{-\int_0^t u_{js}ds}+1-\mu_0}$

(4.49)

$$\mathcal{H}^s = e^{-rt}e^{-\int_0^t u_{s\tau}d\tau}\left(\mu_0 e^{-\int_0^t u_{j\tau}d\tau}\left((u_{st} + u_{jt})v^s - \frac{c^s u_{st}^2}{2}\right) + (1-\mu_0)\left(u_{st}v^s - \frac{c^s u_{st}^2}{2}\right)\right) + \xi_t^s u_{st} + \xi_t^j u_{jt}$$

with states $U_t^j \equiv \int_0^t u_{j\tau}d\tau$, $U_t^s \equiv \int_0^t u_{s\tau}d\tau$ following $\dot{U}_t^j = u_{jt}$, $\dot{U}_t^s = u_{st}$.

and the senior's necessary conditions by

$$\xi_t^s = -e^{-rt}e^{-U_t^s}\left(\left(\mu_0 e^{-U_t^j} + (1-\mu_0)\right)(v^s - c^s u_{st})\right) \tag{4.50}$$

$$\dot{\xi}_t^j = e^{-rt}e^{-U_t^j}\mu_0((u_{st}+u_{jt}=)v^s - \frac{c^s u_{st}^2}{2}) \tag{4.51}$$

$$\dot{\xi}_t^s = e^{-rt}e^{-U_t^s}\left(\mu_0 e^{-U_t^j}\left((u_{st}+u_{jt})v^s - \frac{c^s u_{st}^2}{2}\right) + (1-\mu_0)\left(u_{st}v^s - \frac{c^s u_{st}^2}{2}\right)\right) \tag{4.52}$$

$$U_0^j = 0, U_0^s = 0 \tag{4.53}$$

$$\xi_T^j = 0, \xi_T^s = 0. \tag{4.54}$$

Differentiating (4.43) with respect to time yields

$$\dot{\eta}_t = e^{-rt}e^{-U_t}\left(r(v^j - c^j u_{jt}) + (u_{jt}+u_{st})\left(v^j - c_j u_{jt}\right) + c^j \dot{u}_{jt}\right) \tag{4.55}$$

and after using (4.44)

$$\dot{u}_{jt} = r\left(u_{jt} - \frac{v^j}{c^j}\right) + u_{jt}\left(u_{st} + \frac{u_{jt}}{2}\right). \tag{4.56}$$

For the senior we get by differentiating (4.50)

$$\dot{\xi}_t^s = e^{-rt-U_t^s}\Big(r(\mu_0 e^{-U_t^j} + 1 - \mu_0)(v^s - c^s u_{st}) + (u_{st}+u_{jt})\mu_0 e^{-U_t^j}(v^s - c^s u_{st}) \tag{4.57}$$

$$+u_{st}(1-\mu_0)(v^s - c^s u_{st}) - (\mu_0 e^{-U_t^j} + 1 - \mu_0)c^s \dot{u}_{st}\Big) \tag{4.58}$$

and after using (4.52)

$$\dot{u}_{st} = r(u_{st} - \frac{v^s}{c^s}) + u_{st}\left(\mu_t u_{jt} + \frac{u_{st}}{2}\right). \tag{4.59}$$

Together with the boundary conditions $U_0^j = 0, U_0^s = 0, u_{jT} = \frac{v^j}{c^j}, u_{sT} = \frac{v^s}{c^s}$ equations (4.56) and (4.59) constitute the equilibrium of the game.

The necessary conditions are also sufficient as the running costs are strictly concave in the state variables U_t and weakly concave in the controls in both problems.

4.7.5 Proof of Lemma 4.4

Denote by μ_t^i the belief of player i about player $-i$'s productivity. The belief evolves as before according to $\dot{\mu}_t^i = -\mu_t^i(1-\mu_t^i)u_{-it}$. Rewriting the objectives as in the proof for Proposition 1 yields as Hamiltonian for player i

$$\mathcal{H}^i = e^{-rt} e^{-\int_0^t u_{i\tau} d\tau} \left(\mu_0^i e^{-\int_0^t u_{-i\tau} d\tau} \left((u_{it} + u_{-it}) v^i - \frac{c^i u_{it}^2}{2} \right) + (1 - \mu_0^i) \left(u_{it} v^i - \frac{c^i u_{it}^2}{2} \right) \right) \tag{4.60}$$

$$+ \xi_t^i u_{it} + \xi_t^{-i} u_{-it} \tag{4.61}$$

where the states are $U_t^i = \int_0^t u_{i\tau} d\tau$ and $U_t^{-i} = \int_0^t u_{-i\tau} d\tau$. The necessary conditions for player i are then given by

$$\dot{\xi}_t^i = e^{-rt} e^{-U_t^i} \left(\mu_0 e^{-U_t^{-i}} \left((u_{it} + u_{-it}) v^i - \frac{c^i u_{it}^2}{2} \right) + (1 - \mu_0) \left(u_{it} v^i - \frac{c^i u_{it}^2}{2} \right) \right) \tag{4.62}$$

$$\xi_t^i = -e^{-rt} e^{-U_t^i} \left(\left(\mu_0 e^{-U_t^{-i}} + (1 - \mu_0) \right) (v^i - c^i u_{it}) \right) \tag{4.63}$$

$$\dot{\xi}_t^{-i} = e^{-rt} e^{-U_t^{-i}} \mu_0 ((u_{it} + u_{-it} =) v^i - \frac{c^i u_{it}^2}{2}) \tag{4.64}$$

$$U_0^{-i} = 0, U_0^i = 0 \tag{4.65}$$

$$\xi_T^{-i} = 0, \xi_T^i = 0. \tag{4.66}$$

Again, we differentiate the first-order condition with respect to time and equate it with the law for the co-state evolution to obtain the equilibrium

$$\dot{u}_{it} = r(u_{it} - \frac{v^i}{c^i}) + u_{it} \left(\mu_t^i u_{-it} + \frac{u_{it}}{2} \right) \tag{4.67}$$

together with boundary conditions $u_{iT} = \frac{v^i}{c^i}$.

Bibliography

Aghion, P. and J. Tirole (1994): "The Management of Innovation," *The Quarterly Journal of Economics*, 1185–1209.

Akerlof, G. A. (1970): "The market for" lemons": Quality uncertainty and the market mechanism," *The quarterly journal of economics*, 488–500.

Athey, S. and P. A. Haile (2002): "Identification of standard auction models," *Econometrica*, 70, 2107–2140.

——— (2007): "Nonparametric approaches to auctions," *Handbook of econometrics*, 6, 3847–3965.

Athey, S., J. Levin, and E. Seira (2011): "Comparing open and sealed bid auctions: Evidence from timber auctions," Tech. rep., National Bureau of Economic Research.

Bahn-Report, . (2007): "Licht und Schatten im Bestellerdschungel," *Bahn-Report*, 2007, 24–29.

Bajari, P., S. Houghton, and S. Tadelis (2006): "Bidding for incomplete contracts: An empirical analysis," Tech. rep., National Bureau of Economic Research.

——— (2014): "Bidding for incomplete contracts: an empirical analysis of adaptation costs," *The American Economic Review*, 104, 1288–1319.

Bajari, P. and S. Tadelis (2001): "Incentives versus transaction costs: A theory of procurement contracts," *RAND Journal of Economics*, 387–407.

Bergemann, D. and U. Hege (1998): "Venture capital financing, moral hazard and learning," *Journal of Banking & Finance*, 22, 703–735.

——— (2006): "The Financing of Innovation: Learning and Stopping," *RAND Journal of Economics*, 36, 719–752.

Berman, S. M. (1963): "Note on extreme values, competing risks and semi-Markov processes," *The Annals of Mathematical Statistics*, 34, 1104–1106.

Bhaskar, V. (2014): "The Ratchet Effect Re-examined: A Learning Perspective," .

Bienz, C. and J. Hirsch (2011): "The Dynamics of Venture Capital Contracts," *Review of Finance*.

Bolton, P. and D. S. Scharfstein (1990): "A Theory of Predation Based on Agency Problems in Financial Contracting," *The American Economic Review*, 80, 93–106.

Bonatti, A. and J. Hörner (2011): "Collaborating," *The American Economic Review*, 101, 632–663.

——— (2016): "Career Concerns with Exponential Learning," *Working Paper.*

Bonatti, A. and J. Hrner (2017): "Career concerns with exponential learning," *Theoretical Economics*, 12, 425–475.

Booth, G. G., O. M. Dalgic, and A. Young (2004): "The Staging of Venture Equity Capital and Venture Capitalist Bargaining Power," *The Journal of Entrepreneurial Finance*, 9, 29–40.

Börgers, T., I. Cox, M. Pesendorfer, and V. Petricek (2013): "Equilibrium bids in sponsored search auctions: Theory and evidence," *American economic Journal: microeconomics*, 5, 163–187.

Bramoull, Y. and N. Treich (2009): "Can uncertainty alleviate the commons problem?" *Journal of the European Economic Association*, 7, 1042–1067.

Brendstrup, B. and H. Paarsch (2003): "Nonparametric estimation of Dutch and first-price, sealed-bid auction models with asymmetric bidders," Tech. rep., Working Paper, University of Iowa Department of Economics.

Brendstrup, B. and H. J. Paarsch (2006): "Identification and estimation in sequential, asymmetric, english auctions," *Journal of Econometrics*, 134, 69–94.

Campbell, A., F. Ederer, and J. Spinnewijn (2014): "Delay and Deadlines: Freeriding and Information Revelation in Partnerships," *American Economic Journal: Microeconomics*, 6, 163–204.

Campo, S., I. Perrigne, and Q. Vuong (2003): "Asymmetry in first-price auctions with affiliated private values," *Journal of Applied Econometrics*, 18, 179–207.

Che, Y.-K., E. Iossa, and P. Rey (2016): "Prizes versus Contracts as Incentives for Innovation," *TSE Working Paper, n. 16-695.*

Clarke, F. (2013): *Functional Analysis, Calculus of Variations and Optimal Control*, vol. 264, Springer Science & Business Media.

Cumming, D. (2012): *The Oxford Handbook of Venture Capital*, Oxford University Press.

Cuny, C. J. and E. Talmor (2005): "The Staging of Venture Capital Financing : Milestone vs Rounds," *EFA 2005 Moscow Meetings Paper.*

De Silva, D. G., T. Dunne, and G. Kosmopoulou (2003): "An empirical analysis of entrant and incumbent bidding in road construction auctions," *The Journal of Industrial Economics*, 51, 295–316.

De Silva, D. G., G. Kosmopoulou, and C. Lamarche (2009): "The effect of information on the bidding and survival of entrants in procurement auctions," *Journal of Public Economics*, 93, 56–72.

DeMarzo, P. M. and Y. Sannikov (2016): "Learning, Termination, and Payout Policy in Dynamic Incentive Contracts," *Working Paper*.

Edelman, B., M. Ostrovsky, and M. Schwarz (2005): "Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords," Tech. rep., National Bureau of Economic Research.

Engelbrecht-Wiggans, R., P. R. Milgrom, and R. J. Weber (1983): "Competitive bidding and proprietary information," *Journal of Mathematical Economics*, 11, 161–169.

Estache, Antonio Iimi, A. (2008): *Bidder Asymmetry In Infrastructure Procurement: Are There Any Fringe Bidders?*, The World Bank.

FAZ Nr. 281, . (2015): "Die Bahn kmpft gegen die Netzvergabe an Konkurrenten," *Frankfurter Allgemeine Zeitung*, 281, 23.

Gage, D. (2012): "The Venture Capital Secret: 3 Out of 4 Start-Ups Fail," *The Wall Street Journal*.

Garrett, D. F. and A. Pavan (2012): "Managerial Turnover in a Changing World," *Journal of Political Economy*, 120, 879–925.

Georgiadis, G. (2015): "Projects and Team Dynamics," *The Review of Economic Studies*, 82, 187.

Gerardi, D. and L. Maestri (2012): "A principal–agent model of sequential testing," *Theoretical Economics*, 7, 425–463.

Goeree, J. K. and T. Offerman (2003): "Competitive bidding in auctions with private and common values*," *The Economic Journal*, 113, 598–613.

Gomes, R., D. Gottlieb, and L. Maestri (2016): "Experimentation and project selection: Screening and learning," *Games and Economic Behavior*, 96, 145 – 169.

Gompers, P. A. (1995): "Optimal Investment, Monitoring, and the Staging of Venture Capital," *The Journal of Finance*, 50, 1461–1489.

Gompers, P. A. and J. Lerner (2004): *The Venture Capital Cycle*, MIT press.

Gornall, W. and I. A. Strebulaev (2015): "The economic impact of venture capital: Evidence from public companies," *Available at SSRN 2681841*.

Green, B. and C. R. Taylor (2016): "Breakthroughs, Deadlines, and Self-Reported Progress: Contracting for Multistage Projects," *American Economic Review*, 106, 3660–99.

Guerre, E., I. Perrigne, and Q. Vuong (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525–574.

Gugler, K., M. Weichselbaumer, and C. Zulehner (2014): "Analysis of mergers in first-price auctions," .

——— (2015): "Competition in the economic crisis: Analysis of procurement auctions," *European economic review*, 73, 35–57.

Halac, M., N. Kartik, and Q. Liu (2016): "Optimal Contracts for Experimentation," *The Review of Economic Studies*.

Hannan, M. T., M. D. Burton, and J. N. Baron (1996): "Inertia and Change in the Early Years: Employment Relations in Young, High Technology Firms," *Industrial and Corporate Change*, 5, 503–536.

Hansen, R. G. (1986): "Sealed-Bid versus Open Auctions: The Evidence," *Economic Inquiry*, 24, 125–142.

Hart, O. and J. Moore (1994): "A Theory of Debt Based on the Inalienability of Human Capital," *The Quarterly Journal of Economics*, 109, 841–879.

Hege, U., D. Bergemann, and L. Peng (2012): "Venture Capital and Sequential Investments," *Working Paper*.

Hendricks, K., J. Pinkse, and R. H. Porter (2003): "Empirical implications of equilibrium bidding in first-price, symmetric, common value auctions," *The Review of Economic Studies*, 70, 115–145.

Hendricks, K. and R. H. Porter (1988): "An Empirical Study of an Auction with Asymmetric Information," *The American Economic Review*, 78, 865–883.

Holmstrom, B. (1982): "Moral Hazard in Teams," *The Bell Journal of Economics*, 13, 324–340.

Hong, H. and M. Shum (2000): "Structural estimation of auction models," in *Game Practice: Contributions from Applied Game Theory*, Springer, 123–142.

——— (2002): "Increasing competition and the winner's curse: Evidence from procurement," *The Review of Economic Studies*, 69, 871–898.

——— (2003): "Econometric models of asymmetric ascending auctions," *Journal of Econometrics*, 112, 327–358.

Hörner, J. and L. Samuelson (2013): "Incentives for experimenting agents," *RAND Journal of Economics*, 44, 632–663.

Hu, Z. (2014): "Financing Innovation with Unobserved Progress," .

Hunold, M. and C. Wolf (2013): "Competitive procurement design: Evidence from regional passenger railway services in Germany," *ZEW-Centre for European Economic Research Discussion Paper.*

Kaplan, S. N. and P. Stroemberg (2004): "Characteristics , Contracts , and Actions: Evidence from Venture Capitalist Analyses," *The Journal of Finance.*

Keller, G. and S. Rady (2010): "Strategic experimentation with Poisson bandits," *Theoretical Economics*, 5, 275–311.

Keller, G., S. Rady, and M. Cripps (2005): "Strategic Experimentation with Exponential Bandits," *Econometrica*, 73, 39–68.

Kirkegaard, R. (2009): "Asymmetric first price auctions," *Journal of Economic Theory*, 144, 1617–1635.

Krasnokutskaya, E. (2011): "Identification and estimation of auction models with unobserved heterogeneity," *The Review of Economic Studies*, 78, 293–327.

Külpmann, P. (2016): "Rational Delay of Effort in Projects with Uncertain Requirements," *working paper.*

Laffont, J.-J. (1997): "Game theory and empirical economics: The case of auction data," *European Economic Review*, 41, 1–35.

Laffont, J.-J. and J. Tirole (1986): "Using cost observation to regulate firms," *The Journal of Political Economy*, 614–641.

——— (1988): "The Dynamics of Incentive Contracts," *Econometrica*, 56, 1153–1175.

Lalive, R. and A. Schmutzler (2008a): "Entry in liberalized railway markets: The German experience," *Review of Network Economics*, 7.

——— (2008b): "Exploring the effects of competition for railway markets," *International Journal of Industrial Organization*, 26, 443–458.

Lalive, R., A. Schmutzler, and C. Zulehner (2015): "Auctions vs negotiations in public procurement: which works better?" *Unpublished Manuscript.*

Lebrun, B. (1999): "First price auctions in the asymmetric N bidder case," *International Economic Review*, 40, 125–142.

Legros, P. and S. A. Matthews (1993): "Efficient and Nearly-Efficient Partnerships," *The Review of Economic Studies*, 60, 599.

Li, S. and P. Philips (2012): "Construction Procurement Auctions: Do Entrant Bidders Employ More Aggressive Strategies than Incumbent Bidders?" *Review of Industrial Organization*, 40, 191–205.

Li, T., I. Perrigne, and Q. Vuong (2000): "Conditionally independent private information in OCS wildcat auctions," *Journal of Econometrics*, 98, 129–161.

Li, T. and X. Zheng (2009): "Entry and competition effects in first-price auctions: theory and evidence from procurement auctions," *The Review of Economic Studies*, 76, 1397–1429.

Malmendier, U. and J. Lerner (2010): "Contractibility and the Design of Research Agreements," *American Economic Review*, 100, 214–246.

Manelli, A. M. and D. R. Vincent (1995): "Optimal procurement mechanisms," *Econometrica: Journal of the Econometric Society*, 591–620.

Maskin, E. and J. Riley (2000a): "Asymmetric auctions," *Review of Economic studies*, 413–438.

——— (2000b): "Equilibrium in sealed high bid auctions," *The Review of Economic Studies*, 67, 439–454.

McAfee, R. P. and J. McMillan (1986): "Bidding for contracts: a principal-agent analysis," *The RAND Journal of Economics*, 326–338.

Milgrom, P. R. and R. J. Weber (1982): "A theory of auctions and competitive bidding," *Econometrica: Journal of the Econometric Society*, 1089–1122.

Monopolkommission (2013): "Sondergutachten 64: Bahn 2013: Reform zgig umsetzen!" Tech. rep.

——— (2015): "Sondergutachten 69: Bahn 2015: Wettbewerbspolitik aus der Spur?" Tech. rep.

Moroni, S. (2016): "Experimentation in Organizations," *Working Paper.*

Neher, D. V. (1999): "Staged Financing: An Agency Perspective," *The Review of Economic Studies*, 66, 255–274.

OECD (2015): "Size of public procurement," .

Paarsch, H. J. and H. Hong (2006): "An introduction to the structural econometrics of auction data," *MIT Press Books*, 1.

Parreiras, S. O. (2006): "Contributions to Theoretical Economics," .

Pindyck, R. S. (1993): "Investments of uncertain cost," *Journal of Financial Economics*, 34, 53 – 76.

Prat, J. and B. Jovanovic (2014): "Dynamic contracts when the agent's quality is unknown," *Theoretical Economics*, 9, 865–914.

Reny, P. J. and S. Zamir (2004): "On the Existence of Pure Strategy Monotone Equilibria in Asymmetric First-Price Auctions," *Econometrica*, 72, 1105–1125.

Rodriguez, G. E. (2000): "First price auctions: Monotonicity and uniqueness," *International Journal of Game Theory*, 29, 413–432.

Rubinstein, A. (1982): "Perfect equilibrium in a bargaining model," *Econometrica: Journal of the Econometric Society*, 97–109.

Sandler, T., F. P. Sterbenz, and J. Posnett (1987): "Free riding and uncertainty," *European Economic Review*, 31, 1605 – 1617.

Strausz, R. (1999): "Efficiency in Sequential Partnerships," *Journal of Economic Theory*, 85, 140 – 156.

Suzuki, A. (2010): "An Empirical Analysis of Entrant and Incumbent Bidding in Electric Power Procurement Auctions," .

Tadelis, S. (2012): "Public procurement design: Lessons from the private sector," *International Journal of Industrial Organization*, 30, 297–302.

Tadelis, S. and P. Bajari (2006): "Incentives and award procedures: competitive tendering vs. negotiations in procurement," *Handbook of Procurement, Dimitri, Pigo and Spagolo, Eds.*

Tas, B. K. O. (2017): "Effect of Entrant and Incumbent Bidding on Public Procurement Efficiency," .

Tirole, J. (2010): *The Theory of Corporate Finance*, Princeton University Press.

Wasserman, N. (2008): "The Founder's Dilemma," *Harvard Business Review*.

Curriculum Vitae

2011-2017	University of Mannheim, Germany *PhD Program in Economics,*
09/2016	Yale University *Visiting Scholar*
04-05/2016	Toulouse School of Economics *Visiting Scholar*
2012-2013	Yale University *Visiting Scholar*
2011-2013	University of Mannheim, Germany *Master of Science in Economics, field of study: Economic Research*
2008-2011	University of Mannheim, Germany *Bachelor of Science in Economics*