Essays on the Theory of Incentives:
*Procurement, Franchise & Innovation Contracts*

Shane B. Evans

*A thesis submitted for the degree of*

Doctor of Philosophy

*of the*

Australian National University

Research School of Economics

College of Business & Economics

July 2010
Declaration

This thesis contains no material that has been accepted for the award of any other degree or diploma in any university. To the best of the author's knowledge and belief it contains no material previously published or written by another person, except where due reference is made in the text.

Signature

Date 24th September 2010.
Acknowledgements

I am greatly indebted to my supervisors, Martin Richardson and Ron Stauber for their excellent guidance over the course of my candidature. Every bit of feedback and advice they gave caused an improvement in the quality of this thesis, and I thank them very much for their support and encouragement. Many thanks also to Farshid Vahid for being part of my supervisory committee.

I am lucky enough to have had many other superb teachers, colleagues and friends while studying. They have all been a great positive influence on my work and experience at the ANU. In particular, I wish to thank Aki Asano, Phong Ngo, Paul Chen, Ben Smith, Rod Tyers, Chris Jones, Jose Rodrigues-Neto, Vipin Arora and Kris Funston.

I would also like to thank Lona Fowdur, Daniel Leonard and John Whitley who each had a part in helping me to get started; and Mardi Dungey and Graeme Wells who have been very encouraging during the final stages of writing.

Finally I wish to thank all my family and friends for their encouragement during my study. In particular, I have benefited greatly from Sally, who has proved to be a tough critic, a strong sounding board for ideas and without whose support I could not have undertaken, let alone have finished, this thesis.
Abstract

This thesis explores models of procurement, franchising and innovation through the lens of the theory of incentives. Chapter 2 examines the influence of type-dependent reservation utility on the optimality of linear contracts in a Principal-Agent model of procurement. Type-dependency of reservation utility, combined with the requirements of individual rationality and incentive compatibility in the principal's contracts induces a countervailing incentive effect, the strength of which depends on an index of quality or degree of competition that the agent would face in a private market. The results show how the curvature of the reservation utility dictates whether the optimal contracts can be implemented with a menu of linear contracts, and how the magnitude of the private market index influences the net-transfer rule.

Chapter 3 studies contracts between a manufacturer and a retailer when the retailer has ex ante private information, and is subject to limited liability. The contract takes place over two periods. In the first period, the retailer can take an action which influences the manufacturer's beliefs about the distribution of demand states for a final good in the second period. The retailer sells the manufacturer's intermediate good into a final output market according to a variable fee schedule. The interaction of the limited liability constraints with incentive compatibility gives rise to an expected surplus to the retailer, which the manufacturer can extract with a franchise fee. The franchise fee can also be used as a screening device or a means of eliciting the efficient first stage action from the retailer.

In Chapter 4 a developer contracts with a researcher for the production of a non-drastic innovation. Since effort is non-contractible, the developer offers an incentive contract dependent on the observed magnitude of the innovation. It is shown that the distribution of intellectual property rights (IPR) ownership does not affect the level of effort exerted for innovations where the developer would choose to license the innovation to its competitors. This is because the possibility of leakage of the innovation through licensing subsidies the developer's payment when IPR is delegated to the researcher, while at the same time eroding its profit.
## Contents

1 Introduction ......................................................... 1

2 Menus of Linear Contracts in Procurement .......... 9
   2.1 Introduction .................................................. 9
   2.2 Relation to the Literature ............................... 12
   2.3 Analytical Framework .................................. 15
   2.4 Implementation with a Linear Menu ............... 26
      2.4.1 Convexity of the Reservation Utility ........ 30
   2.5 The Extended Linearity Property .................... 35
      2.5.1 Concavity of the Reservation Utility .......... 38
   2.6 Conclusion ............................................... 39
   2.7 Appendix ................................................... 41
      2.7.1 Proof of Lemma 1 ................................... 41
      2.7.2 Proof of Lemma 2 ................................... 42
      2.7.3 Proof of Lemma 3 ................................... 43
      2.7.4 Proof of Corollary 1 ............................... 44
      2.7.5 Proof of Corollary 2 ............................... 44
      2.7.6 Proof of Lemma 4 ................................... 45
      2.7.7 Proof of Lemma 5 ................................... 45
      2.7.8 Proof of Corollary 3 ............................... 46

3 Franchise Contracts with Limited Liability .......... 47
   3.1 Introduction ............................................... 47
# CONTENTS

3.2 Relation to the Literature ................................. 50
3.3 Discrete Demand State Illustration ............ 53
  3.3.1 Double Marginalisation .......................... 53
  3.3.2 Vertical Restraints ............................... 55
  3.3.3 Franchise Contracts ......................... 56
3.4 Analytical Framework ....................... 61
  3.4.1 Stage 2: Fee Schedule ......................... 64
  3.4.2 Stage 1: Franchise Fee Component ......... 68
    Empty Action Space: \( a = \phi = A_1 \) ....... 70
3.5 Action Spaces .................................... 72
  3.5.1 Inducing Effort with Franchise Fee: \( a \in \{e_H, e_L\} = A_2 \) ....... 72
  3.5.2 Screening by Franchise Fees: \( a \in \{\sigma_L, \sigma_H\} = A_3 \) ....... 75
    Screening or Ignorance? ....................... 77
3.6 Conclusion .................................. 79
3.7 Appendix .................................... 80
  3.7.1 Proof of Lemma 1 ......................... 80
  3.7.2 Proof of Lemma 2 ......................... 80
  3.7.3 Proof of Corollary 1 ................. 82
  3.7.4 Proof of Proposition 1 ............... 82
  3.7.5 Proof of Lemma 3 ......................... 83
  3.7.6 Proof of Lemma 4 ......................... 83
  3.7.7 Proof of Lemma 5 ......................... 84
  3.7.8 Proof of Lemma 6 ......................... 85

4 Innovation Contracts with Leakage Through Licensing 87
  4.1 Introduction .................................. 87
  4.2 Relation to the Literature .................... 91
  4.3 Analytical Framework ....................... 94
  4.4 Weak Equivalence of Ownership ............ 102
  4.5 When IPR Ownership Matters .......... 107
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.6 Innovation in a Differentiated Product Market</td>
<td>110</td>
</tr>
<tr>
<td>4.7 Conclusion</td>
<td>114</td>
</tr>
<tr>
<td>4.8 Appendix</td>
<td>115</td>
</tr>
<tr>
<td>4.8.1 Proof of Lemma 1</td>
<td>115</td>
</tr>
<tr>
<td>4.8.2 Proof of Proposition 1</td>
<td>116</td>
</tr>
<tr>
<td>4.8.3 Proof of Proposition 2</td>
<td>116</td>
</tr>
<tr>
<td>4.8.4 Proof of Lemma 3</td>
<td>116</td>
</tr>
<tr>
<td>4.8.5 Proof of Proposition 3</td>
<td>117</td>
</tr>
<tr>
<td>4.8.6 Proof of Lemma 4</td>
<td>117</td>
</tr>
<tr>
<td>4.8.7 Proof of Proposition 5</td>
<td>118</td>
</tr>
<tr>
<td>4.8.8 Proof of Proposition 6</td>
<td>118</td>
</tr>
</tbody>
</table>

5 Concluding Remarks 121

Bibliography 126
# List of Figures

2.1 An Example Solution for $\nu(\theta, s)$ .................................................. 25  
2.2 Implementation with a Menu of Linear Contracts ............................... 28  
2.3 Costate Solutions for Various $s$-values .............................................. 32  
2.4 Utility, Costate & Effort Profiles – Convex $\pi(\theta, s)$ .................... 34  
2.5 Net-Transfer Schedules ....................................................................... 37  
2.6 Utility, Costate & Effort Profile – Concave $\pi(\theta, s)$ ...................... 39

3.1 Double Marginalisation ................................................................. 54  
3.2 Incentive to Mis-report ....................................................................... 57  
3.3 Incomplete Information Contracts ....................................................... 59  
3.4 General Timing .................................................................................. 63  
3.5 Timing - Stage 1 Moral Hazard ............................................................. 72  
3.6 Timing - Stage 1 Adverse Selection ....................................................... 75

4.1 Timing .................................................................................................. 98
LIST OF FIGURES
Chapter 1

Introduction

The essays comprising this thesis use techniques from the theory of incentives to explore the contracts governing the interaction between (i) a procurement agency and a private firm, (ii) a manufacturer and a retailer, and (iii) a researcher and a developer. This introduction briefly describes the background to the theory and outlines key assumptions employed throughout the thesis to highlight and contrast the main themes of each chapter.

According to the taxonomy of Laffont and Martimort (2002), the theory of incentives belongs to a broader literature known as contract theory. Contract theory evolved from an investigation of environments that the neoclassical theory of general equilibrium, as exposited by Arrow and Debreu (1954), could not deal with directly: namely, where an asymmetry of information between individuals led to opportunistic behaviour by the informed party. Contract theory also includes incomplete contract theory, and transactions cost theory. The former explores matters relating to the assignment of the residual rights of control in the contract (see for example Grossman and Hart (1988) and Hart and Moore (1985)), and the latter deals with the choice of alternative governance structures through merit of their relative transaction costs (which can be traced back to Coase (1937), and was developed later by Williamson (1979)).

The machinery of the theory of incentives is the Principal-Agent model, and it takes as inputs the institutions in which the principal and agent are embedded. The equilibrium of the model is a contract, a mutually agreeable coordination device, which is written to incorporate any informational asymmetries associated with their interaction and any
CHAPTER 1. INTRODUCTION

resulting opportunistic behaviour that is induced by the asymmetries. The theory of incentives most notably departs from the standard neoclassical assumptions by assuming a limited number of economic actors: at its simplest, a single principal and a single agent. However, other standard neoclassical precepts are retained:

- Economic agents are rational.
- A benevolent and rational court of law perfectly enforces the terms of the contract.
- There is common knowledge of the model.

The second of these assumptions implies that any disputes in the contract or breaches of it are dealt with by the benevolent and rational court of law, and without regard to the resource cost of arbitration. Of course, for this to be feasible, the terms of the contract have to be verifiable by a third party. If the contract is contingent on non-verifiable variables, then this assumption is null and the models of incomplete contract theory or transactions cost theory must be employed (see Brousseau and Glachant (2002)).

The third assumption uses the definition of common knowledge as defined by Aumann (1976). It implies that the structure of the contract and its payoffs are known to all participants, and that all participants know that all other participants know, and so on. In the many-agent context of neoclassical general equilibrium models, this means that the terms of trade between two parties are taken as exogenous: hence there is no need to consider the division of the surplus from trade. However, when there is only a principal and agent, the allocation of surplus between the two must be determined within the contract, possibly through some complex bargaining procedure (as in the alternating offer bargaining game of Rubinstein (1982)). To abstract from such bargaining games, the Principal-Agent model assumes that:

- The principal has all the bargaining power.

Hence, the principal is a Stackelberg leader, and offers take-it-or-leave-it contracts to the agent. The agent however, has the option to refuse the offer – if it accepts, it must do so voluntarily. Contracts are therefore constrained by individual rationality constraints: the principal must offer the agent at least their reservation utility or opportunity cost of
accepting the contract. In most models of contracting, the agent’s reservation utility is assumed to be normalised to zero.

To endow the principal with all the bargaining power seemingly reduces the scope for an agent to engage in opportunistic behaviour. The principal could simply specify precise terms of the contract and perfectly monitor the agent’s execution of those terms. However, monitoring the agent becomes problematic when the agent has an informational advantage over the principal. In such circumstances the principal must operate under uncertainty. This requires it to form beliefs about the probability of each possible outcome. The theory of incentives is defined in this type of system. Therefore, the following assumptions are maintained throughout this thesis:

- The agent has an informational advantage over the principal.
- There is a common knowledge probability distribution over the agent’s private information.

The second of these assumptions, combined with the assumption of rational economic agents from neoclassical theory, implies that uninformed players act as Bayesian expected utility maximisers when evaluating their payoffs from the contract. The agent’s informational advantage may arise in two ways. It may be endogenous, where the agent can take an action that the principal cannot observe. Or it may arise exogenously, for example due to the principal’s perceived unobservable heterogeneity in a population of agents. The case of endogenous private information is called hidden action or more commonly, moral hazard. Exogenous private information is called hidden knowledge, or adverse selection.

To deal with moral hazard, the principal typically ties the agent’s payment to the outcome of an observable variable that is correlated to its own objective, and which can be verified by a court of law. When the contracts are successfully designed to elicit a particular action from the agent they are said to be incentive compatible. In the case of adverse selection, the agent can learn its private information after or before signing the contract. In either case, the principal’s contracts are typically characterised as a menu of contracts, one contract for each possible type that the agent could take.

Restricting this menu of contracts to have the same cardinality as the agent’s type
space greatly simplifies the principal's problem. Moreover, this simplification does not come at the expense of generality in the setting of the Principal-Agent model as might be first thought. Rather, it is the result of a powerful theorem in contract theory known as the Revelation Principle. To see the importance of the Revelation Principle, first it is useful to establish that the menu of contracts itself may be more accurately defined as a mechanism: a rule which assigns a contract to each report that an agent can make to the principal about its type. In general, the report can be a complex message reflecting the environment in which the agent is embedded. So the principal's design problem is seemingly a difficult one, and the equilibrium contracts that it writes may need to embody sophisticated information and institutional constraints to induce the agent to reveal its private information.

Fortunately, the principal's task of constructing equilibrium contracts is made simpler by appealing to the Revelation Principle, a central tool of incentive theory developed by numerous authors, including Gibbard (1973), Green and Laffont (1977), Dasgupta, Hammond and Maskin (1979), and Myerson (1979, 1981). The Revelation Principle states that in designing its contracts, the principal can restrict its attention to the set of direct truthful mechanisms. That is, the equilibrium contracts can be chosen among a class of mechanisms that are based directly on reports of the agent's private information only, where the report is truthful in equilibrium. Thus, the Revelation Principle requires that equilibrium contracts satisfy an incentive compatibility constraint: the principal's contracts must embody the agent's opportunistic behaviour.

In summary, incentive theory describes the optimal contracts that a principal could write to alleviate the information advantage that the agent has over them. In doing so, the principal chooses contracts that are both incentive compatible and individually rational to incorporate the opportunistic behaviour of the agent. Each of the following three chapters evaluates the optimal contracts that a principal could write in three different informational and institutional environments. In each chapter, the standard Principal-Agent model is adapted to account for the unique features of the particular industry in which it is embedded. A brief overview of each chapter is provided below.

Chapter 2 deals with procurement contracts. In procurement contracting environ-
ments, the firm that is engaged by the government must be compensated for all its economic costs, including its opportunity costs - or reservation utility. Most models of procurement contracting normalise the reservation utility of the contracted firm to zero. However, in industries like pharmaceuticals and defense, normalisation of reservation utility is not desirable, as if the firm has an intrinsically low cost of production, its outside option of producing in a private market may be considerably higher than if it were an intrinsically high cost firm. It is exactly this ability of different types of firms to earn different profits in outside markets that makes it necessary to introduce type-dependent reservation utilities into the government's problem.

Chapter 2 incorporates type dependent reservation utility into the Laffont and Tirole (1986) procurement model between a government and a firm, where the government can only observe the total cost of the project. However, the firm has private information about its production cost and can exert an unobservable and non-verifiable level of cost-reducing effort. They show that under certain regularity conditions, the optimal procurement contract can be implemented with a menu of contracts that are linear in cost over-runs. The firm's type dependent reservation is modeled so that higher cost firms have lower outside reservation utility. This allows the firm to be opportunistic in two ways: it may over-report its cost so as to save on costly effort, or it may under-report its cost so as to be compensated with a higher reservation utility. This second effect is known as countervailing incentives. The analysis draws on the results for countervailing incentives in Maggi and Rodriguez-Clare (1995) to find the conditions that are required on the shape of the reservation utility to ensure the linearity property still holds.

The model of procurement in Chapter 2 assumes that adverse selection and moral hazard take place simultaneously. However, this does not have to be the case. In fact, in instances where vertically related firms interact, a more realistic timing assumption is one of moral hazard followed by adverse selection, or even sequential adverse selection.

Chapter 3 describes a franchise contracting arrangement between a manufacturer and a retailer with such a sequence of information revelation. Under the introduced timing of information revelation, the retailer does not observe the precise level of demand that it will face at the time of signing the contract. Hence, if demand turns out to be low, the terms of
CHAPTER 1. INTRODUCTION

the contract may penalise the retailer beyond its ex post reservation utility. However, in franchise environments, the retailers are typically small operators with limited assets, and the use of harsh penalties may be prohibited by explicit provisions by local competition authorities. Hence, a contract that prescribes severe punishments in low demand states is not feasible. The motivation for Chapter 3 is to analyse exactly this environment by introducing a limited liability constraint which rules out such penalties.

The impact of modeling limited liability in franchise arrangements is that the distribution of information rents to the retailer are skewed. So the manufacturer must give up greater information rents to higher demand retailers than it would absent limited liability, which results in a positive ex ante expected surplus accruing to the retailer. The model shows that the manufacturer can use a franchise contract consisting of a fixed fee payable upfront and a variable fee schedule specifying payments contingent on the realised demand state. The model also shows how use of the franchise fee can solve informational problems that arise before the demand state is realised.

In the models of Chapters 2 and 3, the agent benefits only from the principal’s payment. However, if the interaction of the principal and the agent through the contract creates some value to an external party, there is an additional source of benefit. Chapter 4 explores such an environment in a model of innovation contracts.

Innovation creates knowledge, which is durable, non-rivalrous in consumption and partially excludable. In short, these properties confer value to the holder of the property right to the innovation through either using the knowledge itself or licensing use of the knowledge to a third party. However, when two parties contract for an innovation, the entitlement to the stream of income from licensing to a third party may influence the strength of the incentives offered in the contract. Chapter 4 explores this environment.

In the model, a principal contracts for the production of an innovation which has value to the principal and also to its competitors. Here, ownership rights to the innovation are an important component of the contract because the innovation can be licensed to the external party. By exerting costly effort, the agent can increase the magnitude (or value) of the innovation; but effort is unobservable and non-verifiable by the principal, so moral hazard arises. Using the same arguments as above, one way for the principal to deal
with moral hazard is to tie the agent's reward to the magnitude of the innovation. This is undertaken in the context of the linear compensation wage model of Holmstrom and Milgrom (1987). However, if the agent owns the innovation, then the principal's provision of explicit incentives to exert effort through its contract is reduced since the agents is implicitly incentivised through its expected value from licensing. The model shows an equivalence result: if the principal would license the innovation to its competitors anyway, then it does not care who owns the intellectual property rights. This occurs because the amount that its profit is reduced by if the agent licenses the innovation to its competitors is exactly offset by the reduction in the expected payment it makes under the terms of the contract.

This brief overview shows that it will be the idiosyncrasies of each chapter's respective institutional environment that influences the nature of the contractual arrangements. In Chapter 2, it is the government's responsibility to compensate the agent for all economic costs of procurement in an environment of type-dependent reservation utility that affects its ability to use linear contracts; in Chapter 3, the timing of the revelation of information to the retailer and its subsequent limited liability will affect the form of the contract that the manufacturer may offer; in Chapter 4 the structure of the industry itself bears on the power of the incentive contract that the developer can offer. In each model, the common thread is the principal's need to resort to contracts to coordinate the agent's otherwise opportunistic behaviour in the context of the different institutions in which they are embedded.
Chapter 2

Menus of Linear Contracts in Procurement with Type-dependent Reservation Utility

2.1 Introduction

Procurement contracts are written in complex institutional environments, and where the procurer is often at an informational disadvantage to the firm with which it wishes to contract. As such, procurement contracts usually involve a sophisticated rule for transferring payments that are contingent on the observed project cost. Moreover, the transfer rule must embody the opportunistic behaviour of the contracted firm. The complexity of these optimal contracts presents not just a theoretical problem, but a practical one since complex contracts are not typically used by procurement agencies. To overcome some of the complexity, one line of research beginning with Laffont and Tirole (1986) has explored the feasibility of implementing procurement policy with a menu of contracts that are linear in cost over-runs.\(^1\) The generally accepted result is that implementation with a menu of linear contracts is feasible provided the transfer rule is decreasing and convex in the project cost. This is because the rule can then be

\(^1\)A larger body of research simply assumes this linearity property as a starting point for their work. See, for example, the recent work of Rogerson (2003) and Chu and Sappington (2007).
CHAPTER 2. PROCUREMENT CONTRACTS

replaced with a menu of its tangents, each tangent representing a linear contract. Hence, Laffont and Tirole's linearity property is important because it connects the theoretically complex contracts to the observation that procurement contracts are typically linear in cost over-runs.

The regularity conditions for linear contracts to be feasible are well established.\(^2\) However there is one aspect of procurement environments for which the linearity property has not been tested: when the contracted firm has type-dependent reservation utility. In most models of procurement, the reservation utility, or outside opportunity of the contracted firm is normalised to zero. However, relaxing this assumption to allow for type-dependent reservation utility affects the core of the agency problem faced by the procurer: the incentives facing the contracted firm allow for more complicated opportunistic behaviour. This chapter attempts to extend the linearity property to include such environments.

Normalisation of reservation utility across different agent types is not a desirable property to impose on procurement models in many cases. For example, in Australia, the Commonwealth government procures pharmaceuticals under a policy known as the Pharmaceutical Benefit Scheme. A pharmaceutical company can make an application to the government to be placed on a registered list, whereupon it can operate in a regulated prescription market. However, the pharmaceutical company could also choose to sell its drugs into a private prescription market, and record a profit that may depend on not just its cost of production, but also the quality of its drug, or the degree of competition it faces or other market variables. Another example is defense procurement: firms in the defense industry typically operate in many international markets, and when using their resources in one contract, have given up contracts in other markets. Again, the size of the profit given up by a firm depends on its production costs as well features of the market that it could have operated in. The problem with allowing for type-dependent reservation utilities is that the principal must compensate the agent for its opportunity cost. This introduces scope for opportunistic behaviour by the agent: by misreporting its type, the agent may receive more compensation for its reservation utility than is warranted.

\(^2\)Rogerson (1997) spells out the regularity conditions in detail.
To be more precise, in Laffont and Tirole's formulation the total observed cost of the project is \( C = \theta - e \) where \( \theta \) is the agent's privately known intrinsic cost of the project and \( e \) is the effort that it exerts to reduce cost. Hence, the principal cannot determine whether a high total cost is the result of a high cost agent or a low cost agent not exerting effort. The agent prefers to exert less effort, as it is personally costly, with disutility represented by \( \psi(e) \). Moreover, effort is unobservable and non-verifiable, and therefore cannot be written into a contract or enforced by a court of law. The principal must therefore design its contracts to deal with adverse selection from the agent's private information about its intrinsic cost \( \theta \), and moral hazard from the principal's inability to observe and verify the agent's effort. So the principal's contracts must incentivise the agent to both report the truth about its intrinsic cost, and to exert a certain level of effort.

To do this, the principal's contracts are based on a report from the agent about its privately observed intrinsic cost. After receiving the report \( \theta \), the principal reimburses the cost \( C(\theta) \) to the agent, and transfers an amount \( y(\theta) \). Hence, the agent's payoff is \( y(\theta) - \psi(e) \). Ideally, this transfer just covers the agent's opportunity cost and its disutility from effort. However, since effort is unobservable and non-verifiable, a lower cost agent has incentive to over-report, its intrinsic cost to receive the transfer designed for a higher cost agent. By doing so, the lower cost agent earns a rent since it can reduce its effort and save on disutility. This is the usual direct effort incentive effect. To reduce the expected rent that the principal has to pay out, it may design its contracts in a way that reduces the effort it requires of higher cost agents. This in turn reduces the reward for lower cost agents to over-report. In doing this, the principal therefore trades of rent extraction against a reduction in the efficient level of effort in cost reduction.

This chapter explores the additional incentive effect that arises when the agent's reservation utility depends on its cost type. The transfer \( y(\theta) \) must cover not only the disutility of the agent's effort, but must also account for its reservation utility as well. If the reservation utility decreases in the agent's type, then a higher cost agent has incentive to under-report its cost and receive the higher transfer designed for a lower cost agent. This effect works in the opposite direction to the direct effort incentive, hence it is known
as the countervailing incentive effect. This time, to reduce the expected rent that the principal has to pay out to induce truthful reporting, the contracts require agents to exert more effort which increases the disutility from under-reporting.

The balance of the opposing incentive effects depends on the magnitude of the agent's reservation utility, which in turn depends on the characteristics of the alternative market in which it could have operated. Private market features, like drug or weapon quality or fierceness of competition in the private market, affect the magnitude of the agent's reservation utility and hence the importance of the countervailing incentive in the principal's problem. Hence, these market features are summarised by a "private market parameter", indexed by the real-valued parameter $s \in \mathbb{R}_+$. This enables different strengths of the countervailing incentive effect on the principal's contracts to be explored.

The results show that feasibility of implementation with linear contracts rests on the curvature of the reservation utility in the agent's cost type. It will be shown that provided the reservation utility satisfies a convexity bound, Laffont and Tirole's linearity property is preserved. However, if the reservation utility is concave, an equilibrium menu of contracts may exist, but cannot be implemented with a menu of linear contracts. The chapter also demonstrates how different levels of the private market parameter change the strength of the countervailing incentive effect, but do not disturb the linearity property.

The next section reviews some of the relevant literature in procurement and type dependent reservation utility. Then the feasibility of implementation with linear contracts is studied by constructing the equilibrium contracts and determining the conditions on the curvature of the reservation utility that lead to convex net-transfer rules. The last section concludes.

2.2 Relation to the Literature

The optimality of menus of linear contracts has in the past been a subject of intensive study. Many models of procurement assume from the outset that contracts are linear, or use a linear menu of contracts as a benchmark for the relative performance of their incentive schemes, including some very recent research (see for example, Baron and Besanko (1988), Bower (1993), Rogerson (2003), Chu and Sappington (2007)). The Laffont and
Tirole (1986) procurement model first provided the sufficiency conditions for an optimal menu of contracts to be implemented with a menu of linear contracts. They showed that provided the net-transfer rule for procurement is decreasing and convex, then it can be replaced with a menu of its tangents, where each tangent represents a linear contract.

Laffont and Tirole's assertion was followed by a series of research papers directed at finding a generalisation of the linearity property of optimal contracts, as in McAfee and McMillan (1987), Melumad and Riechelstein (1989) and Rogerson (1987). The results from that line of research showed that the optimality of a menu of linear contracts does not hold in general. As qualified in Laffont and Tirole (1993, p. 107):

"We thus should not consider the linearity result as a general rule but rather as defining a class of environments in which one can conveniently work with linear schemes."

However, so far the environments studied in which the linearity property holds have been limited to the case where the reservation utility of each type of agent is normalised to zero. The task of this chapter is devoted to re-examining the linearity property in a procurement environment where the reservation utility of the agent is type dependent. It is well known that type dependent reservation utility introduces a new problem, where an agent may no longer have a systematic incentive to always under-report its cost. Hence, the monotonicity property required to properly implement separating contracts may become violated. Problems of this sort have come to be known as countervailing incentives.

In a paper closely related to this one, Lewis and Sappington (1989) study how countervailing incentives can arise in the seminal incentive regulation model of Baron and Myerson (1982). In their framework, adverse selection arises since the principal faces an agent with private information about its marginal production cost. In the equilibrium contracts, the principal trades off efficiency of production against limiting the information rents to low marginal cost types of agents. It does this by distorting downward the quantity it asks high marginal cost agents to produce, thus reducing the benefit to low marginal cost agents from over-reporting their cost. Lewis and Sappington's (1989)
extension analyses the case where countervailing incentives are created through the existence of a type-dependent fixed cost of production.

In their model, lower marginal costs are associated with higher fixed costs - so that reservation utility is decreasing in the agent’s type. As such, agents no longer have a systematic incentive to over-report their type as in Baron and Myerson. By under-reporting, the agent can be rewarded with a higher transfer to cover its perceived higher fixed cost. However, to ensure their problem satisfies the second order conditions for an optimum, Lewis and Sappington assume that the reservation utility is concave. This assumption leads to the result that their equilibrium contracts entail some pooling: for intermediate cost realisations, the principal induces the same level of production from those cost type agents.

Maggi and Rodriguez-Clare (1995) generalise the countervailing incentives problem in agency contracts. In doing so, they are able to explain the pooling equilibrium result of Lewis and Sappington. They demonstrate that whether an equilibrium contract exhibits pooling or is separating depends on the assumptions placed on the curvature of the reservation utility. Specifically, whenever the reservation utility is concave, pooling characterises the equilibrium contract - as in Lewis and Sappington. However, when the reservation utility is strictly convex, the equilibrium contracts are fully separating.

Maggi and Rodriguez’s optimal control technique is employed in this chapter in the context of Laffont and Tirole’s (1986) procurement model, which features both adverse selection and moral hazard. Their solution employs a rent extraction-efficiency trade off that is similar to the Baron-Myerson result, but involves an inefficiency in cost-reducing effort rather than production.³ This chapter introduces countervailing incentives to the Laffont-Tirole model in a similar manner to Lewis and Sappington: lower intrinsic cost

³Rogerson (1994) places the model more concretely: it is well known that first-best contracts can be achieved when there is (i) complete information and (ii) when all parties are risk-neutral and (iii) there is a deterministic relationship between cost and effort. Relaxing assumptions (ii) and (iii) lead to the literature of moral hazard models where the optimal incentive scheme trades off risk allocation for effort inducement (see Shavell (1979), Holmstrom (1979), Hart and Holmstrom (1987) and Mas-Colell et al., (1995 Ch. 14)). Relaxing only assumption (i) gives rise to the Laffont-Tirole analysis, where the fundamental trade-off between efficiency and rent extraction arises. The trade-off is driven by the agent’s private information about its exogenously given efficiency parameter (type) and an endogenous effort variable that has desirable cost-reducing consequences, but which brings unobservable disutility to the agent.
agents have higher reservation utility since they would be more profitable in the private market. However, unlike Lewis and Sappington, no restriction is made from the outset on the curvature of the reservation utility. Rather, this chapter derives the restrictions on the curvature that are required for the linearity property to hold. In addition, by controlling the private market parameter, this chapter is able to study the equilibrium solutions under varying relative strengths of the countervailing incentive effect to the direct effort incentive effect, including cases where one or other is dominant.

The results confirm the dependence of separating or pooling equilibria on the curvature of the reservation utility. By allowing the reservation utility to vary exogenously with a private market parameter, the solution technique of Maggi and Rodriguez-Clare (1995) enables this analysis to gain a clear insight into how the strength of the countervailing incentive affects the relationship between the information rents and the effort profiles that the principal chooses to minimise its expected procurement cost, and the ability of the principal to implement the optimal contract with a menu of linear contracts.

It must be noted that the countervailing incentive problem arises in a number of other contexts. It arises in Champsaur and Rochet's (1989) study of the interaction between two competing firms, in Laffont and Tirole's (1990) examination of bypass of a regulatory regime, in Biglaiser and Mezetti (1993) in the context of competing principals and in Acconcia et al (2008) in a model of vertical restraints. A more recent significant analysis of countervailing incentives is Jullien (2000), who provides a general treatment of the problem where the type-dependent individual rationality constraints give rise to countervailing incentives. While Jullien's paper is very general, this chapter follows the methodology of Maggi and Rodriguez-Clare (1995).

2.3 Analytical Framework

The agent is characterised by two parameters: its ex ante cost \( \theta \in [\theta_0, \theta_1] \) and a private market parameter, \( s \in \mathbb{R}_+ \), which indexes some aspect of the agent's performance in the outside private market for its services. For example, the parameter could index the degree of market power the agent exerts in the private market, or reflect the quality of the service that the agent offers. It is assumed that the private market parameter is observable.
by the principal, but that the cost \( \theta \) is not. More precisely, the principal has a prior belief of the distribution of the agent's cost (or type) given by the common knowledge cumulative distribution function \( F(\theta) \) with density \( f(\theta) \) defined over the interval \([\theta_0, \theta_1]\). The distribution \( F(\theta) \) is assumed to satisfy the monotone hazard rate property:

\[
\frac{d}{d\theta} F(\theta) \geq 0 \geq \frac{d}{d\theta} \frac{1-F(\theta)}{f(\theta)}
\]

The principal is a Stackelberg leader who contracts with the agent for one unit of a good or service, and compensates the agent with a monetary transfer, \( y \), over and above reimbursement of production costs. The agent's total cost of production is:

\[
C = \theta - e
\]  

(2.1)

where \( e \geq 0 \) is the agent's effort. If the agent exerts effort level \( e \), it decreases the monetary cost of production by one dollar and incurs a disutility of effort \( \psi(e) \) measured in units of utility. It is assumed that effort is positive or zero over the range of equilibrium efforts of interest. Note that disutility only occurs when effort is strictly positive and increases with effort at an increasing rate: \( \psi(0) = 0, \psi'(e) > 0, \) and \( \psi''(e) > 0. \) Also \( \lim_{e \to 0} \psi(e) = +\infty \), and it is assumed that \( \psi'' \approx 0. \)

Total cost is observed by the principal, and by accounting convention, cost is reimbursed to the agent by the principal. Public funds are used to finance the transfers made to the agent, so the cost of the project is as perceived by the taxpayer.

An alternative specification for the agent's total cost is to introduce additive noise:

\[
C = \theta - e + \epsilon
\]

where \( \epsilon \) is an accounting or forecast of production error with mean zero. Laffont and Tirole (1986) show that in such instances, if both the principal and agent are risk neutral then the optimal contract may still be implemented by a menu of linear contracts. This happens because the agent's incentives taken over the expectation of \( \epsilon \) remain identical as without noise. However, if the agent is risk averse, then the linearity property no longer holds in general (see section IV and Appendix D in Laffont and Tirole (1986)).

Hence, to maintain an environment in which linear contracts are robust
2.3. **ANALYTICAL FRAMEWORK**

to noise, it is necessary to confine the analysis to the case where both the principal and agent are risk neutral. Then the impact of type-dependent reservation utility on the optimal contracts can be isolated.

The reservation utility of an agent of type $\theta$ is represented by $\pi(\theta, s)$, where $s \in \mathbb{R}^+$ is the private market parameter. A number of assumptions are made on this function:

1. $\pi_\theta < 0$: the reservation utility is higher for lower cost agents, as they have greater value in the private market.

2. $\pi_s > 0$: the reservation utility increases in the private market parameter, which is some index like quality, or product differentiation, or degree of market power.

3. $\pi_{\theta s} \leq 0$: the marginal benefit to an agent from understating their cost is greater for higher values of the private market parameter.

No assumption is made on the curvature of the reservation utility in cost type yet: in finding feasible solutions, bounds for the curvature will be derived.

In order to induce the agent to enter the contract, the principal must offer at least as much utility to the agent as it would obtain in its outside opportunity, which depends on the magnitude of the private market parameter. Following Maggi and Rodriguez-Clare (1995), let $U$ denote the agent's "net-utility": the utility received in its relationship with the principal in excess of its reservation utility. The individual rationality constraint is then:

\[(IR) \quad U(y(\theta)) = y(\theta) - \psi(\epsilon) - \pi(\theta, s) \geq 0, \quad \forall \theta \in \Theta, \ & \text{& some } s \in \mathbb{R}^+ \quad (2.2)\]

Consumer benefit from production is captured by the real number $W \in \mathbb{R}^+$. Assuming that there is an excess burden of taxation, using public funds induces a distortion. Denote $\lambda > 0$ as the shadow cost of public funds. Then the net surplus to consumers is: $V = W - (1 + \lambda)(y + C)$. The principal is assumed to be utilitarian, so ex post social welfare is:

\[V + U = W - (1 + \lambda)(C + \psi + \pi) - \lambda U \quad (2.3)\]

powered incentives) than accounting forecast errors when the agent is risk averse.
where the individual rationality constraint (IR) has been used to substitute for the transfer $y$. The presence of the last term indicates that it is socially costly to leave rents to the agent. The principal’s objective is to maximise social surplus, which involves minimising the expected rent paid to the agent. In a situation of incomplete information about the agent’s cost type, the principal cannot avoid giving up rents to low cost agents as they can exploit the informational asymmetry to their advantage, as will be shown. However, before analysing that situation, it is useful to obtain the solution to the principal’s problem in the complete information case as a benchmark. This gives the first best solution.

The principal’s objective is to maximise social welfare, given by equation (2.3). Assuming that production is always worthwhile ($V + U > 0, \forall \theta$), then for each $\theta$, the principal chooses the smallest transfer $y(\theta)$ to induce that type of agent to participate. Under complete information about cost $\theta$, the first best condition is:

$$
\psi'(e^*(\theta)) = 1, \quad \Rightarrow \quad e = e^*(\theta)
$$

(2.4)

The interpretation of this condition is that the marginal disutility from exerting a unit of effort equals the marginal benefit - the monetary value of the reduction in cost due to that effort. Notice that the first best level of effort is invariant to the agent’s type - the effort profile is flat. Hence, after reimbursing the total production cost $C^*(\theta) = \theta - e^*$, the principal transfers $y^*(\theta)$ to compensate the agent for the disutility cost of its effort and to cover its reservation utility: $y^*(\theta) = \psi(e^*) + \pi(\theta, s)$. Hence, all types of agents earn zero information rents. However, the transfer schedule is decreasing in $\theta$. This is because a high cost agent has a smaller reservation utility, and needs to be compensated less over and above cost reimbursement for its disutility of effort than its low cost counterpart.

Under incomplete information about the agent’s type, the principal still observes the agent’s actual cost data, even if it does not know the agent’s cost or the level of effort exerted to reduce cost. Then a feasible mechanism for the principal is to offer a net-transfer $y$ for each level of observed cost $C$. However, such a mechanism is prone to perverse incentive effects. On one hand, a low cost agent may earn a rent by reporting its cost as high and completing the project at a higher (observable) total cost than in
2.3. ANALYTICAL FRAMEWORK

the first best case. In doing so, the agent reduces its effort to below the first-best level, thus reducing its disutility from effort. On the other hand, a high cost agent may earn a rent by reporting its cost as low and completing the project for a lower total cost than in the first best case. This time, the agent raises its effort above the first-best level, which increases its disutility from effort, but is rewarded through a higher transfer since the principal believes its opportunity cost is higher than it really is.

Hence, a feasible mechanism must now be based on the observed cost level and the private information. Fortunately, the Revelation Principle guarantees that there is no loss of generality in restricting the problem to truthful mechanisms. In such a mechanism, the principal asks that the agent make a report on its type, \( \theta_r \in [\theta_0, \theta_1] \), and based on that report the principal transfers \( y(\theta_r) \) to the agent, in addition to reimbursing its cost \( C(\theta_r) \). Equivalently, given the report \( \theta_r \), the principal reimburses production cost \( C(\theta_r) \) and leaves information rent \( U(\theta_r) \) for the agent. Truth-telling, or incentive compatibility, then requires:

\[
(\text{IC}) \quad \arg \max_{\theta_r} \{ y(\theta_r) - \psi(\theta - C(\theta_r)) - \pi(\theta, s) \}, \quad \forall \theta \in \Theta \text{ & some } s \in \mathbb{R}_+ \tag{2.5}
\]

The necessary and sufficient conditions that the principal is constrained by in choosing the amount of effort and rent to offer to the agent in its contract to maximise social welfare are specified in the following Lemma:

**Lemma 1** (Necessary and Sufficient Conditions) Necessary and sufficient conditions for incentive compatibility are: (i) \( U_\theta(\theta) = -\psi'(e(\theta)) - \pi_\theta(\theta, s) \) and (ii) \( e_\theta(\theta) \leq 1 \).

Note that part (ii) of Lemma 1 is really a monotonicity condition on the total cost: \( C_\theta(\theta) \geq 0 \). This means that the effort profile may be downward or upward sloping over subintervals of the type space, so long as it is never steeper than (positive) unit slope.

The principal's problem is to induce an effort profile \( e(\theta) \) and choose a utility profile \( U(\theta) \) to maximise expected social welfare, while respecting individual rationality (so as to keep each type of agent in the contract) and accounting for the incentive effects that occur due to the informational asymmetry about the agent's cost type. The principal's
problem can be stated formally:\(^5\)

\[
(P) \quad \max_{\{e,U\}} \int_{b_0}^{b_1} \left( W - (1 + \lambda) \left( \theta - e + \psi(e(\theta)) + \pi(\theta, s) \right) - \lambda U \right) f(\theta) d\theta \quad (2.6)
\]

subject to:

\[
(IR) \quad U(\theta) = y(\theta) - \psi(e) - \pi(\theta, s) \geq 0, \quad \forall \theta \in \Theta, \text{ and}
\]

\[
(IC) \quad U_\theta(\theta) = -\psi'(e(\theta)) - \pi_\theta(\theta, s), \text{ for some } s \in \mathbb{R}_+.
\]

Truth-telling, or incentive compatibility is achieved by the principal through allowing agents a distribution of information rents \(U(\theta)\), whose magnitude and rate of change in \(\theta\) is governed by the strengths of local marginal incentives for each type of agent. When deciding which cost type to report to the principal, the agent balances the effects of two sources for personal marginal gain: (i) the direct effect that arises from reducing its effort level when reporting a higher cost, and (ii) the indirect effect through receiving a lower reservation utility when reporting a higher cost. The necessary condition of Lemma 1 for incentive compatibility bears out these two effects:

\[
U_\theta(\theta) = -\psi'(e(\theta)) - \pi_\theta(\theta, s) \quad (2.7)
\]

The first term on the right hand side is the agent’s marginal disutility saving on effort from over-reporting its cost type. If an agent of type \(\theta\) purports to be an agent of type \(\theta + d\theta\), it can reduce its effort and still meet the higher cost target: \(dC = -de\). The personal disutility saving the agent makes on this marginal effort reduction is \(-\psi'(e)de\).

This acts to increase the rent \(U(\theta)\) accruing to the \(\theta\) cost agent from reporting to be a \(\theta + d\theta\) type. The second term on the right hand side is the agent’s marginal reduction in reservation utility from over-reporting its cost type. Since higher cost types receive lower reservation utilities, this acts to reduce the rent \(U(\theta)\) accruing to the \(\theta\) cost agent.

Hence, this second indirect incentive effect is called the “countervailing incentive effect”.

\(^5\)Note that this is the relaxed maximisation program, since the sufficient condition of Lemma 1 is omitted. The solution is checked ex post to ensure that it satisfies sufficiency.
2.3. ANALYTICAL FRAMEWORK

The balance of these two local incentive effects determines the slope of the utility profile at each particular value of \( \theta \).

Notice that the right hand side of the necessary condition for IC depends implicitly on two variables (holding constant the agent’s type): effort, and the private market index. The dependence of the utility profile on each of these variables will be examined in turn.

First, consider the private market effect. This variable impacts the utility profile exogenously. To see how, recall that \( \pi_{\theta s} \leq 0 \) by assumption. Therefore, when the private market parameter \( s \) is very small, so the agent has little or no market power for example, the countervailing incentive term \( \pi_{\theta}(\theta, s) \) is close to zero in absolute value (recall that \( \pi_{\theta}(\theta, s) < 0 \)). In this case, it is likely that the direct (effort) incentive effect dominates and the utility profile will be downward sloping over all cost types of the agent. This yields the standard monotonicity constraint (or constant sign condition \( CS^- \)) as in Guesnerie and Laffont (1984)). The \( CS^- \) condition holds because there is a systematic incentive for the agent to over-report its cost type, no matter what its actual cost type. When \( CS^- \) holds, the equilibrium contracts are fully separating - one contract is designed for each possible cost type of agent. Thus, the incentive constraint is locally downward binding. In fact, monotonicity means that the incentive and individual rationality constraints can be combined together to solve the initial value problem comprising the necessary condition in Lemma 1 and \( U(\theta_1) = 0 \) to yield an expression for the type \( \theta \) agent’s rent:

\[
U(\theta) = \int_{\theta}^{\theta_1} \left( \psi'(c(\tilde{\theta})) + \pi_{\theta}(\tilde{\theta}, s) \right) d\tilde{\theta} \tag{2.8}
\]

Now consider the case where the private market index is very large, so the agent has a lot of market power: then \( \pi_{\theta}(\theta, s) \) is relatively large and negative. In this case, it is likely that the countervailing incentive effect is dominant and the utility profile will be sloping upward over all cost types of the agent. In this case, the agent has a systematic incentive to under-report its cost type, and the incentive constraint is locally upward-binding. The monotonicity constraint would be of the \( CS^+ \) type, and the initial value
problem \( U_\theta = -\psi'(e(\theta)) - \pi_\theta(\theta, s) \) and \( U(\theta_0) = 0 \) solves to yield:

\[
U(\theta) = -\int_{\theta_0}^{\theta} \left( \psi'(e(\tilde{\theta})) + \pi_\theta(\tilde{\theta}, s) \right) d\tilde{\theta} \tag{2.9}
\]

For intermediate values of the private market index, the sign of the left hand side of the necessary condition for IC changes. When this happens, the monotonicity requirement does not hold: the constant sign condition is violated. In general, when this happens the optimal solution entails some bunching on the subinterval of types where the constant sign condition changes.\(^6\) However, here a special optimal control technique developed by Maggi and Rodriguez-Clare (1995) is used. This is the topic of the next section.

Before moving to the formal maximisation problem, consider the effort variable inside the necessary condition for IC. The magnitude of the direct incentive effect depends on the effort profile. Moreover, inspection of the rent equations above show that the effort profile has a cumulative effect on the rent for any given cost type in general. Hence, if the principal can control the amount of effort it induces the agent of type \( \theta \) to exert, it can distort the distribution of information rents to its advantage. In fact, as the next section will describe, the principal optimally chooses an induced effort profile to limit the information rents that it has to given up to lower cost agents. It does so by trading off rent reduction through an efficiency distortion away from the first best level.

Throughout the analysis, it is assumed that the gross project value, \( W \) is always large enough that shut down of some types is never optimal. That is, the principal finds that the project is viable for all types that the agent can take. It is conceivable that this would not always be true, in which case the principal would optimally shut down some types.\(^7\) In such cases, the net social surplus on the marginal type that the principal should shut down would be smaller than the expected rent the principal would have to pay out to achieve incentive compatibility. The principal would then shut down higher types if the direct effort incentive effect was dominant, or lower types if the countervailing incentive

\(^6\)Guesnerie and Laffont (1984) derive the full solution for cases where the constant sign condition is violated.

\(^7\)For example, a situation where positive reservation utilities lead to optimal shut down of types is discussed in Lu’s (2009) auction design problem where potential bidders have a known positive opportunity cost of bidding. He shows that the revenue-maximising auction may implement asymmetric entry across symmetric bidders.
2.3. ANALYTICAL FRAMEWORK

An optimal control technique developed in Maggi and Rodriguez-Clare (1995) is used here, and the derivation closely follows theirs. In step with the usual procedure for solving incentive problems, the global second order condition is checked ex-post. Hence, the principal’s problem is as before in (P) - equation (2.6). The control variable is the (induced) effort level \(e\), the state variable is net-utility \(U\) and the costate variable is \(\mu\). Hence, the Hamiltonian is:

\[
H = (W - (1 + \lambda)(\theta - e + \psi(e) + \pi) - \lambda U) f(\theta) - \mu(\theta)(\psi'(e) + \pi_\theta)
\]  

(2.10)

The point of departure here from usual incentive problems is that the issue of countervailing incentives as discussed in the previous section means that the IR constraint may bind on a subset of types in a non-trivial way. Hence, it is not possible to simply exploit the monotonicity of the utility profile to eliminate the IR constraint from the problem. To deal with this, the constraint is modeled explicitly by formulating the Lagrangian for the problem:

\[
\mathcal{L} = H + \tau U
\]

where \(\tau\) is the Lagrange multiplier for the IR constraint. Applying the Maximum Principle to this Lagrangian, as in Leonard and Long (1992), yields the following necessary conditions:

\[
\psi'(e(\theta)) = 1 - \frac{\mu(\theta)}{1 + \lambda f(\theta)} \psi''(e(\theta))
\]  

(E)

\[
\mu_\theta = \lambda f(\theta) - \tau(\theta)
\]  

(C)

\[
U_\theta(\theta) = -\psi'(e(\theta)) - \pi_\theta(\theta, \delta)
\]  

(S)

\[
\frac{\partial\mathcal{L}}{\partial \tau} = U(\theta) \geq 0, \quad \tau(\theta) \geq 0, \quad U(\theta) \tau(\theta) = 0
\]  

(CS)

\[
\mu(\theta_0) \leq 0, \quad \mu(\theta_0)U(\theta_0) = 0 \text{ and } \mu(\theta_1) \geq 0, \quad \mu(\theta_1)U(\theta_1) = 0
\]  

(TV)

---

\(^8\)Technically, it can be shown that if there exists a \(\theta^* \in (\theta_0, \theta_1)\) defined by \((W - (1 + \lambda)(\theta^* - e^* + \psi(e(\theta^*))) + \pi(\theta^*, \delta))f(\theta^*) = \lambda(\psi'(e(\theta^*)) + \pi_\theta(\theta^*, \delta))F(\theta^*)\), then cost types greater than \(\theta^*\) should be shut down whenever the direct incentive effort incentive dominates and cost types lower than \(\theta^*\) should be shut down whenever the countervailing incentive effect dominates.
CHAPTER 2. PROCUREMENT CONTRACTS

The transversality conditions nest the possibilities for the endpoints of the solution profiles. To see this, take the case where the direct (effort) incentive effect dominates in the IC. Then low cost agents will inevitably earn information rents by over-reporting their cost. To ensure incentive compatibility, the principal must leave rents to the low cost agent, but will reduce the utility profile over the type space to ensure that the highest cost type receives no rent, where the IR constraint will be binding. To minimise the expected information rent, the principal can distort the effort required for higher cost types lower than first best, which prevents the gain from high cost types saving on effort reduction. So at some $\hat{\theta}$, the costate variable $\mu^*(\hat{\theta})$ is the imputed value to the principal of retaining the rent for types greater than $\hat{\theta}$, and should therefore be positive. Hence, in such cases, the shadow price of rent to the principal at the highest cost type is positive, $\mu(\theta_1) > 0$, and the principal leaves that cost type with no rent: $U(\theta_1) = 0$. For the opposite case where the countervailing incentive effect dominates, the incentives work in the opposite direction. The value of the shadow price of rent to the principal is greatest at the lowest cost type because the principal receives no benefit from incentivising that cost type to report its true type. Hence $\mu(\theta_0) > 0$, and $U(\theta_0) = 0$ in that case.\(^9\)

Since $H(U, e, \mu, \theta)$ is concave in $U$, if the configuration $(U(\theta), e(\theta), \mu(\theta), \tau(\theta))$ satisfies equations (E)-(TV) it also satisfies the sufficient conditions for an optimum. Let $e^\mu(\theta)$ denote the value of $e$ that maximises $H$ given $\mu$ and $\theta$, defined by equation (E). Let $\nu(\theta, s)$ be the solution to the state differential equation (S): $U_{\theta} = -\psi'(e(\theta)) - \pi_\theta(\theta, s)$. Then $\nu(\theta, s)$ is the value of the costate variable such that $U_{\theta}(\theta) = 0$. It is clear that $\mu(\theta) = \nu(\theta, s)$ on any interval where the IR constraint is binding (and in that case, the principal should optimally pin down the net-utility to zero for that interval).

The $\nu(\theta, s)$ solution has a neat graphical interpretation: it is the locus of points in $(\mu, \theta)$ space where $U_{\theta}(\theta) = 0$. Figure 2.1 illustrates this for a particular solution of $\nu(\theta, s)$ indicated by the heavy trace. Evidently, in the same space the region above this locus corresponds to $U_{\theta}(\theta) < 0$, and everywhere below has $U_{\theta}(\theta) > 0$. Recall that $\mu(\theta) = \lambda F(\theta)$ is the optimal costate trajectory whenever $U_{\theta}(\theta) < 0$ and $\mu(\theta) = \lambda(F(\theta) - 1)$ is the

\(^{9}\)The economic interpretation of the costate variables are not discussed in Maggi and Rodriguez-Clare (1995), but are important to understanding the problem. This generalised "shadow price" interpretation of the costate variable is derived in Leonard (1987) and discussed intuitively in Leonard and Long (1992).
optimal costate solution whenever $U_\theta(\theta) > 0$. These are indicated in Figure 2.1 by the light traces. Hence, the optimal costate trajectory follows $\lambda F(\theta)$ whenever it lies below $\nu(\theta, s)$, and then becomes the $\nu(\theta, s)$ solution itself whenever $\lambda F(\theta)$ is above. Likewise, whenever $\lambda(F(\theta) - 1)$ lies above $\nu(\theta, s)$ it represents the optimal costate solution, and follows $\nu(\theta, s)$ when it is above. With this in mind, the conjectured solution is defined over a partition of the type space. In particular, let $\Theta = \Theta_I \cup \Theta_{II} \cup \Theta_{III}$ where the partitions are defined by $\Theta_I := \{\theta : \lambda F(\theta) \leq \nu(\theta, s)\}$, $\Theta_{II} := \{\theta : \lambda(F(\theta) - 1) \leq \nu(\theta, s) \leq \lambda F(\theta)\}$, $\Theta_{III} := \{\theta : \nu(\theta, s) \leq \lambda(F(\theta) - 1)\}$. Then the conjectured solution is:

$$
\mu^*(\theta) = \begin{cases} 
\lambda F(\theta) & \text{for all } \theta \in \Theta_I \\
\nu(\theta, s) & \text{for all } \theta \in \Theta_{II} \\
\lambda(F(\theta) - 1) & \text{for all } \theta \in \Theta_{III}
\end{cases}
$$

(2.11)

Exactly which profile the optimal costate variable takes depends on the position of the $\nu(\theta, s)$ solution, which in turn depends on the private market index.

Inspection of the (point-wise) first order condition for effort in equation (E) and the conjectured costate solution reveals an important insight into the principal’s equilibrium.
contracts. The distortion in the effort for incentive compatibility depends on the magnitude of the costate variable at each cost type. As such, if there is a change in the slope of the costate solution across partitions, then effort is not distorted in a uniform direction. This will be manifest as a non-convexity, or wrinkle in the net-transfer rule - which precludes the ability of the principal to use a menu of linear contracts. So in the next section, a careful study is made of the position and shape of the $\nu(\theta, s)$ solution for different values of the private market parameter. Note also the dependence of the slope of $\nu(\theta, s)$ on the curvature of the reservation utility through the state equation (S). Whether the reservation utility is concave or convex will dictate the feasibility of implementation with a menu of linear contracts.

In the next section, the conditions on the curvature of the reservation utility that preserve the linearity property of optimal procurement contracts are derived. Since the strength of countervailing incentives depends on the magnitude of the private market parameter, solutions corresponding to various levels of $s$ are examined. An equilibrium will be denoted $\left(U^*(\theta), e^*(\theta), \mu^*(\theta) | s \in \mathbb{R}_+ \right)$, as the solution to the principal's problem outlined in equations (E), (S), (C), (CS) and (TV) using the conjectured costate solution (2.11) for a particular value of the private market parameter, $s \in \mathbb{R}_+$.

2.4 Implementation with a Linear Menu

This section derives the net-transfer rule that implements the solution to the principal's problem, $\left(U^*(\theta), e^*(\theta), \mu^*(\theta) | s \in \mathbb{R}_+ \right)$ for various values of the private market parameter. Most importantly, the conditions on the convexity of the agent's reservation utility are obtained that allow the solution to be implemented with a menu of linear contracts. This linearity property of optimal procurement contracts holds whenever the net-transfer rule is decreasing, convex and continuously differentiable.\textsuperscript{10} This section establishes the conditions required for the net-transfer rule to be decreasing and convex. The next section discusses continuous differentiability.

From Lemma 1, in a solution to the principal's problem the total cost $C(.)$ is a strictly

\textsuperscript{10}A function that is continuously differentiable is one whose derivative is continuous. This simply rules out kinks in the net-transfer rule. If the net-transfer rule has a kink, then the left and right limits of the derivative at the kink are not equal, so the derivative of the rule is not continuous.
increasing function of $\theta$. Hence, it can be inverted so that $\theta = \theta^*(C)$. Rewriting equation (2.2) then yields:

$$Y^*(C) = \psi(e^*(\theta^*(C))) + \pi(\theta^*(C), s) + U^*(\theta^*(C)), \quad s \in \mathbb{R}_+ \quad (2.12)$$

To establish whether the net-transfer rule is decreasing and convex, differentiate equation (2.12):

$$\frac{dY^*}{dC} = -\psi'(e^*) \quad \text{and} \quad \frac{d^2Y^*}{dC^2} = -\psi''(e^*) \frac{e_\theta^*}{1 - e_\theta^*} \quad (2.13)$$

where the necessary condition for incentive compatibility has been used in the left hand equation. Since the disutility of effort function is assumed to be increasing and convex, the cost-reimbursement rule is always decreasing. However, global convexity requires the optimal effort profile to be decreasing: $e_\theta^*(\theta) \leq 0$. Laffont and Tirole's (1993) work shows that a sufficient condition for implementability with a menu of linear contracts is for the monotone hazard rate assumption to hold. This is equivalent to requiring that the effort profile be decreasing in cost type. Recall that this is a stronger condition than what is required for sufficiency for an equilibrium, as outlined out in part (ii) of Lemma 1: $e_\theta^* \leq 1$. The Laffont-Tirole linearity property is gathered in Proposition 1 below:

**Proposition 1 (Linearity Property)** If a solution to the principal's problem (P) given by $(U^*(\theta), e^*(\theta), \mu^*(\theta) \mid s \in \mathbb{R}_+)$ entails an effort profile that is downward sloping, $e_\theta^*(\theta) \leq 0$, then the optimal contract can be implemented with a menu of linear contracts.

The implication is that the requirements for implementation by a linear menu are stronger than for an equilibrium to exist: the net-transfer rule may in fact exhibit non-convexities, and yet still support a solution. Moreover, the slope of the equilibrium effort profile for a particular value of the private market parameter is inextricably tied to the curvature of the reservation utility, as will be discussed in detail below.

Notwithstanding, if the effort profile is strictly decreasing, then equations (2.13) imply that the net-transfer rule in equation (2.12) is decreasing and convex. Hence, it can be...
replaced with a menu of its tangents - a menu of contracts that are linear in cost over-runs or under-runs. This is illustrated in the left hand side of Figure 2.2. The dashed tangent lines to the net-transfer rule indicate linear contracts designed for two possible cost types, $\theta_L$ and $\theta_H$. If the agent is of type $\theta_L$, it will choose to take the steeper linear contract, exert effort level $e^*_L$ and land the project for $C_L$. The principal then reimburses the agent for total cost $C_L$ and transfers $Y(C_L)$. Similarly for the higher cost agent. Note that neither agent has incentive to take the other's contract. This is because the net-transfer rule embodies incentive compatibility. Now consider what happens if the rule has a non-

convexity, as in the right hand side of Figure 2.2.\textsuperscript{12} The higher cost agent clearly has incentive to exert more effort and report to be the lower cost agent that has experienced a cost over-run. The larger transfer from the principal more than compensates the agent for the extra effort it had to exert. Hence, incentive compatibility is violated. Importantly, the net-transfer rule (in bold) is a solution. However, the pertinent point here is simply that it cannot be implemented with a menu of linear contracts.

If the net-transfer rule is decreasing, convex and continuously differentiable, the menu

\textsuperscript{12}The right hand side of Figure 2.2 is drawn to be continuously differentiable. However, if there was a kink in the rule at $C_L$, a similar problem would exist.
of linear contracts can be represented as:

\[ y(\hat{e}, C) = y^*(\hat{e}) - \psi'(e^*(\hat{e}))(C - C^*(\hat{e})), \quad \forall \hat{e} \in \Theta \]  \tag{2.14} 

so a linear contract is offered to each cost type of agent. Note that \( y^*(\hat{e}) \) is simply the net-transfer that the agent of type \( \hat{e} \) would receive if facing the optimal menu. Also, since \( \psi'(.) > 0 \), the coefficient of cost over-runs represents the power of the incentive contract, which is higher for lower cost types. For the lowest cost type of agent, \( \psi'(e^*(\theta_0)) = 1 \), and the menu prescribes a fixed price contract (the highest powered incentives).\(^{13}\)

To check that an agent of cost type \( \hat{e} \) is induced to report its true cost type, note that the agent's problem is to simultaneously choose a cost type to report, \( \hat{e} \), and a level of effort to exert to ensure the project is landed for its report cost: \( \hat{C} = \hat{e} - e \):

\[
\max_{\{\hat{e}, e\}} \left[ \psi(e^*(\hat{e})) + \pi(\hat{e}, s) + U^*(\hat{e}) - \psi'(e^*(\hat{e}))(\hat{e} - e) - (\hat{e} - e^*(\hat{e})) - \psi(e) \right] \tag{2.15}
\]

where the first three terms in the maximisation comprise the transfer \( y^*(\hat{e}) \). The first order conditions for the agent's problem are:

\[
\psi'e^*_\hat{e} + \pi_\theta - \psi'(e^*_\theta - 1) - \psi''(\theta - e - (\hat{e} - e^*(\hat{e})))e^*_\theta = 0
\]

\[
\psi'(e^*(\hat{e})) - \psi'(e) = 0
\]

The second of these first order conditions shows that the menu of linear contracts induces the same level of effort from the agent as the optimal menu induces: \( e^*(\hat{e}) = e \). The first condition then shows that the agent reports the truth: \( \hat{e} = e \).

Now that the conditions on the effort profile for the linearity property to hold have been established in Proposition 1, and it has been shown that a linear menu can induce truth-telling, the next step is to establish the conditions on the convexity of the agent's reservation utility that ensure the solution entails an effort profile that is downward sloping. This is undertaken in the next section.

\(^{13}\)It is also worth pointing out that the agent experiences no uncertainty in this model: in equilibrium there is no cost under-run or over-run.
2.4.1 Convexity of the Reservation Utility

This section constructs the regularity conditions on the convexity of the reservation utility that are required for the linearity property to be extended to this generalised procurement environment. Recall that the linearity property holds when the agent’s reservation utility is normalised to zero, provided the effort profile is always downward sloping. However, now the equilibrium effort profile in general depends on the magnitude of the private market parameter, and the convexity of the reservation utility. This is because the private market parameter influences the position of the \( \nu(\theta, s) \) solution in \([\mu^*, \theta]\) space, and the convexity of the reservation utility affects its slope, as seen below in equation (2.16).

Moreover, through equation (E), the size and direction of the distortion of effort away from the first best level, and the slope of the effort profile both depend on the position and slope of \( \nu(\theta, s) \) whenever countervailing incentives bind in the solution. Then to proceed, the slope of the \( \nu(\theta, s) \) solution and the impact of \( s \) on its position are analysed.

To obtain the shape of the \( \nu(\theta, s) \) costate solution, differentiate equation (S) at the equilibrium solution (where \( \mathcal{U}_o(\theta) = 0 \)):

\[
\nu(\theta, s) = \left(-\theta^\nu - \frac{\pi^\theta \theta}{\psi^\theta}\right) / \psi^\theta
\]  

(2.16)

Returning to the first order conditions from the Maximum Principle, condition (CS) implies that \( \mathcal{U}(\theta) = 0 \) whenever \( \tau(\theta) \geq 0 \). Hence equation (C) gives \( \tau(\theta) = \lambda f(\theta) - \nu(\theta) \geq 0 \).\(^{14}\) Thus, for an interval of types over which the IR constraint binds to be an optimum, the slope of the costate profile must not be too steep: \( \nu(\theta) \leq \lambda f(\theta) \), which is the analogous condition to Maggi and Rodriguez-Clare (1995).

Turning now to the linearity property requirement that the slope of the effort profile be always downward sloping, differentiating equation (E) yields:\(^{15}\)

\[
\frac{d e^*}{d \theta} = -\frac{1}{1 + \lambda d \theta} \left( \frac{\mu^*}{f} \right)
\]  

(2.17)

\(^{14}\)When the IR constraint is not binding, \( \mu_\theta(\theta) = \lambda f(\theta) \) and then the transversality conditions (TV) give the usual cases of always downward binding or always upward binding incentive compatibility.

\(^{15}\)Formally, \( \psi'' \frac{d e^*}{d \theta} = -\frac{\psi''}{f^2} \left( \frac{\mu^*}{f} \right) + O(\psi''', \theta), \) where \( O(\psi''') \approx 0 \).
2.4. IMPLEMENTATION WITH A LINEAR MENU

From the definition of the conjectured costate solution in equation (2.11), it follows from (2.17) that the effort profile is downward sloping on partitions $\Theta_I$ and $\Theta_{III}$ since $\frac{d}{ds} \left( \frac{r}{f} \right) > 0$ by assumption. On partition $\Theta_{II}$, equation (2.17) implies that the requirement is $\frac{d}{ds} \left( \frac{r}{f} \right) > 0$. Lemma 2 states the convexity requirements on the reservation utility that meet these conditions:

Lemma 2 Provided that $\pi_{\theta_0}(\theta, s) \in [0, \tilde{\pi}_{\theta_0}]$ then: (i) $\nu_0(\theta, s) \leq \lambda f(\theta)$, and (ii) $e^*_\theta(\theta) \leq 0$, where $\tilde{\pi}_{\theta_0} := \psi'' \frac{\lambda}{1 + \lambda} \left( \frac{F(\theta)}{f(\theta)} \right)$.

Under the conditions specified in the Lemma, the $\nu(\theta, s)$ solution never has a slope greater than the slopes of the costate variable on the other two partitions. This implies that to maximise expected welfare, the principal need not distort the effort profile away from the first best as much over the interval on which the individual rationality constraint is binding. This is because the principal optimally reduces the agent’s induced effort to offset the countervailing incentive at the exact rate required to keep each type of agent on its reservation utility.

Having established the convexity requirements for the shape of the $\nu(\theta, s)$ costate solution, now the position of the solution with respect to the private market parameter will be established. Note that for very low values of $s$, say very low quality, the reservation utility of the agent is essentially flat: the standard direct effort incentive should dominate. Conversely, when quality is very high, the agent’s outside opportunity is high, so the countervailing incentive effect should be dominating. This suggests that for higher values of the private market parameter, the $\nu(\theta, s)$ solution should shift downward in $[\mu^*, \theta]$ space. The following Lemma formalises this intuition:

Lemma 3 The $\nu(\theta, s)$ solution shifts down as the value of the private market parameter increases: $\nu_0 = -\frac{\pi_{\theta_0}}{e^*_\theta \psi''} < 0$.

Inspection of Figure 2.3 reveals that where the $\nu(\theta, s)$ solution intersects either $\mu^* = \lambda F(\theta)$ or $\mu^* = \lambda (F(\theta) - 1)$ defines the boundary of partitions I and II or II and III respectively. For values of the private market parameter like $s_L$, $s_3$ or $s_H$, there is no intersection of partitions. Then the critical cost type $\tilde{\theta}$ is defined as the cost type on
CHAPTER 2. PROCUREMENT CONTRACTS

Figure 2.3: Costate Solutions for Various $s$-values

the boundary of these partitions: $\hat{\theta} \in \{\Theta_I \cap \Theta_{II}, \Theta_{II} \cap \Theta_{III}, \phi\}$, where $\phi$ represents no partition.

The equilibrium effort and information rent solutions to the principal’s problem (P) for the private market parameter values from Figure 2.3 when the curvature of the reservation utility is $\pi_{\theta\theta}(\theta, s) \in [0, \pi_{\theta\theta}]$ are listed in Table 1.

Table 1. Solution Profiles for: $\pi_{\theta\theta} \in [0, \pi_{\theta\theta}]$

<table>
<thead>
<tr>
<th>Solution Profile</th>
<th>Effort Profile</th>
<th>Rent Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type L: $\Theta = \Theta_I$</td>
<td>$e^L(\theta) \leq e^{fb}$</td>
<td>$U^*(\theta) = \int_0^{\phi}(\psi'(e) + \pi_{\theta\theta})d\theta$</td>
</tr>
<tr>
<td>Critical $\hat{\theta} = \phi$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type $s_2$: $\Theta = \Theta_I \cup \Theta_{II}$</td>
<td>$e^{fb} \geq e^{2}(\theta) \geq e^L(\theta)$</td>
<td>$U^*(\theta) = \int_0^{\phi}(\psi'(e) + \pi_{\theta\theta})d\theta$ for $\theta \leq \hat{\theta}$ and 0 otherwise</td>
</tr>
<tr>
<td>Critical $\hat{\theta} = \Theta_I \cap \Theta_{II}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Theta = \Theta_{II}$</td>
<td>$e^2(\theta) \geq e^{fb}$ for $\theta \leq \hat{\theta}$, otherwise $e^{fb} \geq e^2(\theta)$</td>
<td>$U^*(\theta) = 0$ for all $\theta$</td>
</tr>
<tr>
<td>$\hat{\theta} = {\theta</td>
<td>\nu(\theta, s) = 0}$</td>
<td></td>
</tr>
<tr>
<td>Type $s_4$: $\Theta = \Theta_{II} \cup \Theta_{III}$</td>
<td>$e^{H}(\theta) \geq e^{fb}$ for $\theta \geq \hat{\theta}$, otherwise $e^{fb} \geq e^3(\theta)$</td>
<td>$U^*(\theta) = -\int_0^{\phi}(\psi'(e) + \pi_{\theta\theta})d\theta$ for $\theta \geq \hat{\theta}$ and 0 otherwise</td>
</tr>
<tr>
<td>Critical $\hat{\theta} = \Theta_{II} \cap \Theta_{III}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type H: $\Theta = \Theta_{III}$</td>
<td>$e^H(\theta) \geq e^{fb}$</td>
<td>$U^*(\theta) = -\int_0^{\phi}(\psi'(e) + \pi_{\theta\theta})d\theta$</td>
</tr>
<tr>
<td>Critical $\hat{\theta} = \phi$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Solution types $L$ and $H$ correspond to the scenarios where either the direct effort incentive is dominant (type $L$), or the countervailing incentive is dominant (type $H$). As depicted in the light traces of the left hand panel in Figure 2.4, for the type $L$ solution the principal optimally distorts the effort levels downward from first best in order to
2.4. **IMPLEMENTATION WITH A LINEAR MENU**

limit the information rent that it must give up to low cost types of the agent to prevent them from over-reporting their cost. On the other hand, the right hand panel of Figure 2.4 shows that for the Type $H$ solution, the principal optimally distorts the effort levels upward from first best in order to limit the reservation utility it must pay out to higher cost agents to prevent them from under-reporting their cost. The interesting cases occur for intermediate levels of the private market parameter, where the balance of the direct effort and countervailing incentives changes over the type space.

The heavy traces in the left hand side of Figure 2.4 depicts the case where countervailing incentives becomes dominant on a subset of the higher cost type agents. Then the conjectured costate solution of equation (2.11) forces the state variable, net-utility, to be governed along a path of zero net-utility corresponding to $v(\theta, s)$ for those higher cost types. Since the values of $\nu(\theta, s)$ over the subset of higher cost types is still positive, but smaller than $\lambda F(\theta)$ then equation (E), effort is distorted to a lessor degree than under the fully dominant direct effort case. A similar effect occurs for higher values of the private market parameter like $s_4$ in the far right panel of Figure 2.4. This statement is made precise in the following Corollary:

**Corollary 1** Suppose $\pi_{00}(\theta, s) \in [0, \pi_{00}]$. Then $\forall \theta \in \Theta_{II}: e^L(\theta) \geq e^\nu(\theta) \geq e^H(\theta)$

The corresponding utility profiles for each case, as illustrated in Figure 2.4 suggest two things. First, that the critical value of $\bar{\theta}$ (locally) decreases as the private market parameter increases. Second, that the value of the utility profile decreases for all cost types less than $\bar{\theta}$ on partition I, and increases for all cost types greater than $\bar{\theta}$ on partition III. The details are in the proof of the following Corollary:

**Corollary 2** Suppose $\pi_{00}(\theta, s) \in [0, \pi_{00}]$. Then: $\frac{\partial U(\bar{\theta})}{\partial s} \leq 0 \text{ for } \bar{\theta} \in \Theta_I$, $\frac{\partial U(\bar{\theta})}{\partial s} = 0 \text{ for } \bar{\theta} \in \Theta_{II}$ and $\frac{\partial U(\bar{\theta})}{\partial s} \geq 0 \text{ for } \bar{\theta} \in \Theta_{III}$.

Figure 2.4 sheds light on the reasoning for the conjectured costate variable. In the top left hand side panel, the $v(\theta, s)$ solution intersects $\mu(\theta) = \lambda F(\theta)$ at $\bar{\theta}$. Remembering that in the region below $v(\theta, s)$ it must be that $U_\theta(\theta) < 0$, then $\mu(\theta) = \lambda F(\theta)$ is optimal and the induced effort profile on $\Theta_I = [\theta_0, \bar{\theta}]$ is given implicitly by $\psi'(e^L) = 1 - \frac{\lambda}{1+\lambda} \frac{E}{\psi''(e^L)}$.

16This is not a global relationship, since there is a discontinuity defined for values of the private market parameter like $s_3$.  

CHAPTER 2. PROCUREMENT CONTRACTS

Moving along the $\lambda F(\theta)$ curve from the left, $\mu(\theta)$ is increasing - the value of relaxing the IC and thereby reducing the information rent left to higher-cost types impacts positively on the principal's objective function (since information rents are socially costly), so that all higher types should receive lower rent. Moving through the intersection of $\lambda F(\theta)$ with $\nu(\theta, s)$ leads to a region where $U_{\theta}(\theta) > 0$ so that $\lambda F(\theta)$ is no longer optimal. Pursuing that trajectory would lower the expected value of the principal's objective function, because in that region $\tau(\theta) > 0$ and the IR constraint binds, so $U(\theta) = 0$ in the solution to (P). Instead, the costate solution optimally tracks $\nu(\theta, s)$, so that $U_{\theta}(\theta) = 0$. Also, by requiring $U(\hat{\theta}) = 0$, it follows that the optimal costate variable satisfies $U_{\theta}(\theta) \leq 0$ ($\mu_{\theta}(\theta) = \lambda f(\theta)$) and $U(\hat{\theta}) = 0$ from the left, and $U_{\theta}(\theta) = 0$ with $U(\hat{\theta}) = 0$ from the right. The induced effort profile to the right of $\hat{\theta}$ is given implicitly by $\psi'(e^{\nu}) = 1 - \frac{1}{1+\lambda} \psi''(e^{\nu})$. Since $\nu(\theta, s) < \lambda F(\theta)$, the effort distortion is less than it would be without the countervailing incentive effect, and there is a kink in the induced effort profile at $\hat{\theta}$, as shown in the middle left hand panel.

If the value of the private market parameter is consistent with Partition III, the middle panel of Figure 2.4 shows that there is no part of the optimal solution for $U_{\theta}(\theta) < 0$.
(\mu(\theta) = \lambda F(\theta)) that intersects with the region below \( \nu(\theta, s) \), where \( U_\theta(\theta) < 0 \), and therefore it cannot form part of the solution. Similarly, no part of \( \mu(\theta) = \lambda (F(\theta) - 1) \) intersects with the region above \( \nu(\theta, s) \) where \( U_\theta(\theta) > 0 \), hence it cannot form part of the solution. In fact it follows from the conjectured solution that the costate variable in this cases is \( \mu^*(\theta) = \nu(\theta, s) \) over the entire type space. This means that the IR constraint binds for all types, and no agent type receives information rent, as depicted at the top of the middle panel. All types receive compensation for their reported cost and their disutility of effort exerted, plus, they receive a transfer equal to the profit they would have made in the private market.

Since the IC binds on each type, for all cases discussed so far the equilibrium is fully separating. This is borne out in the middle of Figure 2.4, where it is seen that the induced effort profile is still downward-sloping - a result due to the fact that \( \nu(\theta, s) \) is non-decreasing. Importantly, whenever \( \nu(\theta) > 0 \), the induced effort profile is distorted downward from first-best to relieve the principal from leaving costly information rent to low-cost types. On the other hand, whenever \( \nu(\theta) < 0 \), the effort profile is distorted upward so as to mitigate the costly information rents left to high-cost types. Of course when \( \nu(\theta) = 0 \), the first best is effort is induced for that particular type where it occurs.

### 2.5 The Extended Linearity Property

Having established the convexity requirements on the agent’s reservation utility to ensure that the equilibrium effort profiles are downward sloping for all values of the private market parameter, this section examines the properties of the corresponding net-transfer rules. In particular, it remains to be seen whether the net-transfer rules are continuously differentiable: if they are not, then implementation with a menu of linear contracts is not permissible. Recall that the net-transfer rule is:

\[
Y^*(C) = \psi(\varepsilon^*(\theta^*(C))) + \pi(\theta^*(C), s) + U^*(\theta^*(C)), \quad s \in \mathbb{R}_+ \tag{2.18}
\]

Where the countervailing incentive is strong enough that the solution is defined over two partitions (where, for example, quality is greater enough that higher cost agents begin
CHAPTER 2. PROCUREMENT CONTRACTS

to have incentive to under-report their cost), the solution \( (U^*(\theta), c^*(\theta), \mu^*(\theta) \mid \theta \in \mathbb{R}_+ ) \) is “patched” across partitions. At the borders of the partition, there is a kink in the costate profile and therefore in the effort profile, as seen for example in Figure 2.4 in the leftmost and rightmost cases. At first glance, this non-differentiability in the effort profile could translate into a non-convexity in the net-transfer rule. If it does, then the linearity property is void. However, Lemma 4 shows that even patched solutions give rise to decreasing and convex net-transfer functions.

**Lemma 4** If \( \pi_{\theta}(\theta, s) \in [0, \pi_{\theta}] \), the net-transfer rule is continuously differentiable, decreasing and convex in \( \theta \) for all \( s \).

Intuitively, for any value of the private market parameter, \( s \), at the border of a patched solution the slope of the utility profile approaches zero. This reflects the rate at which is optimal to trade-off observed total cost levels for transfers when the type crosses from (say) region I into II. This may seem surprising in view of the non-differentiability in the effort and costate profiles on the border of partitions. However, the non-differentiability only translates into a discontinuity in the convexity of the net-transfer rule: while the rule itself is smooth and continuous, its curvature undergoes a change on the border of a patched solution. The next Lemma formalises this result.

**Lemma 5** When the IR constraint is binding for a subset of agent cost types, the net-transfer rule is less convex than for subsets where it is relaxed.

The implication of Lemmata 4 and 5 are shown in Figure 2.5. As the private market parameter increases so that the countervailing effect begins to bind on higher cost types, the new net-transfer function “peels off” from the original function and lies below the hypothetical net-transfer rule. Intuitively, the reason is that since the principal does not distort the effort profile downward as much for higher cost types when the IR binds, those higher cost types must exert more effort, closer to the first-best level. The higher effort requirement for higher types increases their marginal disutility from effort, making the slope of the net-transfer function steeper. Any types along the patched portion of the curve earn zero net-utility, however the effort schedule is still downward sloping.
For high enough values of the private market parameter where countervailing incentive binds on lower types, a similar effect occurs. Now the optimal effort profile entails a smaller distortion above the first best effort than when countervailing incentives would bind on all cost types. Hence lower cost types do not exert as much effort, and the marginal disutility from effort is therefore smaller. Consequently, the new net-transfer function peels off from the original function and again lies below the hypothetical net-transfer rule.

For values of the private market parameter like $s_3$, where the balance of the direct effort and countervailing incentives result in a flat utility profile, the net-transfer rule is simply:

$$Y^*(C) = \psi(e'(\theta(C))) + \pi(\theta''(C), s_3)$$

In this case, the slope of the net-transfer rule is determined by the shape of the reservation utility, since from the necessary condition for IC, $-\psi'(e'(\theta)) = \pi_\theta(\theta, s_3)$, since the utility profile is flat: $U_\theta(\theta) = 0$ over the entire cost type interval.

Lemmata 4 and 5 and the descriptive reasoning above lead to the conclusion that the net-transfer rule is decreasing and convex for all values of the private market parameter, provided that the reservation utility meets Lemma 2's convexity requirement. In such cases, it can always be replaced by the family of its tangents. This is the key result of this
analysis, which generalises Laffont and Tirole's (1986) linearity property, as captured in Proposition 2:

**Proposition 2 (Extended Linearity Property)** If the curvature of the agent's reservation utility is in the interval \( \pi_{\theta}(\theta, s) \in [0, \hat{\pi}_{\theta}] \), then a solution to the principal's problem (P) given by \( (U^*(\theta), e^*(\theta), \mu^*(\theta) \mid s \in \mathbb{R}_+) \) can be implemented with a menu of linear contracts.

The result implies that the class of environments in which it is convenient to work with linear incentive schemes is necessarily restricted to situations where the agent's outside opportunity is convex in cost type.

Laffont and Tirole's precise hypothesis for the linearity property to hold specified that the monotone hazard rate assumption should hold. Earlier it was argued that this is equivalent to requiring that the effort profile be downward sloping. That equivalence lead to the focus on the optimal costate, effort and rent profiles for different values of the private market parameter of the previous section to derive Proposition 2. Interestingly however, a counterpart to their hypothesis can be obtained in this generalised procurement environment that has the same substantive meaning:

**Corollary 3** A solution to the principal's problem (P) given by \( (U^*(\theta), e^*(\theta), \mu^*(\theta) \mid s \in \mathbb{R}_+) \) can be implemented with a menu of linear contracts whenever: \( \frac{d}{d\theta} \left( \frac{\mu^*(\theta)}{\hat{\pi}_{\theta}} \right) \geq 0 \).

So the extended linearity property amounts to a monotone restriction on the generalised costate variable, \( \mu^*(\theta) \) of equation (2.11). The conjectured costate variable was in turn a response to the generalised procurement environment of type dependent reservation utilities.

To complete the analysis, attention is finally turned to the case where the reservation utility is concave. While the extended linearity property of Proposition 2 show that implementation with a menu of linear contracts will not be feasible if the reservation utility is concave, it is still possible to obtain an optimal menu, as the next section briefly illustrates.

### 2.5.1 Concavity of the Reservation Utility

When the reservation utility is concave, it may still be possible to support a solution \( (U^*(\theta), e^*(\theta), \mu^*(\theta) \mid s \in \mathbb{R}_+) \) with a net-transfer rule, however the rule necessarily has a
non-convexity. As a result, it cannot be implemented with a menu of linear contracts. Figure 2.6 shows the solution profiles for such a case. As can be seen, the downward sloping $\nu(\theta, s)$ costate solution induces a switch in the trajectory of the effort profile on partition $\Theta_H$. The effort goes from below first best to above in this partition. However, provided that over the entire interval the slope of the effort profile never exceeds 1: $e^*_0 \leq 1$, then part (ii) of Lemma 1 still holds - the second order condition for the profile to form part of a solution is satisfied. Moreover, since $C_\theta(\theta) \geq 0$, the solution can still be inverted and a net-transfer rule can be derived. However, the rule is decreasing but not convex. The right hand side of Figure 2.2 shows the non-convexity that is induced in the net-transfer rule. Hence, the linearity property is violated when the reservation utility is concave in cost type. This is analogous to the result in Lewis and Sappington (1989).

2.6 Conclusion

The objective of this chapter was to extend the linearity property of optimal procurement contracts to the more general environment of type-dependent reservation utilities. This situation is likely to arise in the pharmaceutical or defence industries, where firms may have significant market power in a private market or a valuable outside opportunity.
which they must be compensated for in the procurement contract. In the pharmaceutical industry for instance, this means that lower cost types have higher reservation utility since they forgo greater profits in the private prescription market. Moreover, the higher the quality of the drug under procurement, the greater those foregone profits are likely to be.

The results of this chapter showed that in such industries, a countervailing incentive effect must be accommodated in the optimal procurement contract. Furthermore, it was shown that such optimal contracts can only be implemented with a menu of linear contracts if the reservation utility satisfies a certain convexity requirement, in addition to the standard regularity assumptions (as in Laffont and Tirole (1986)). The convexity requirement is robust to various strengths of the countervailing incentive effect. The strength was controlled through the degree of quality, or market power or some other private market parameter, and lead to various sub-cases which induced different rent extraction-efficiency trade-offs.

Concave reservation utility alone does not rule out an optimal menu of contracts, but it does rule out the linearity property. The reason is that concave reservation utility results in the principal switching the trajectory of the effort profile on a subset of cost types, which induces a non-convexity in the net-transfer rule, which if used would result in pooling of cost types. This finding parallels Lewis and Sappington's (1989) result that pooling characterises the equilibrium solution when the reservation utility is concave. However, here a separating contract still exists for concave reservation utility provided that the effort profile is not steeper than unit slope.

The implication for procurement in industries like pharmaceuticals and defence that may exhibit type-dependent reservation utilities is that the linearity property cannot be assumed from the outset without due reference to the curvature of the reservation utility. If the return from under-reporting cost is weakly greater for lower cost firms in those industries, then the results of this chapter suggest that the linearity property holds; otherwise optimal contracting does not entail the linearity property.
2.7 Appendix

2.7.1 Proof of Lemma 1

The proof is almost identical to Laffont and Tirole (1993, Chapter 2). From the Revelation Principle, attention is restricted to the class of direct truthful mechanisms without loss of generality: \( \{y(\theta'), C(\theta')\}_{\theta' \in [\theta_l, \theta_u]} \), where \( \theta' \) is the agent’s report of their cost parameter. Then the agent’s payoff as a function of its report is:

\[
U(\theta, \theta') = y(\theta') - \psi(\theta - C(\theta')) - \pi(\theta, s)
\]

For the necessary and sufficient conditions for local incentive compatibility, every agent \( \theta \) chooses to report \( \theta' \) to maximise its transfer from the principal. The local first-order necessary condition is:

\[
y_\theta(\theta') - \psi'(\theta - C(\theta'))C_\theta(\theta') = 0 \quad (2.19)
\]

Hence, for truth-telling to be optimal for every type, the necessary condition follows by evaluating the previous condition at \( \theta = \theta' \):

\[
y_\theta(\theta) - \psi'(\theta(\theta))C_\theta(\theta) \equiv 0 \quad \forall \theta \in \Theta \quad (2.20)
\]

As this holds for every type, the condition is an identity. Now differentiating equation (2.19) again with respect to the agent’s true type, and evaluating under the condition that the agent reports the truth yields the expression in part (i) of the Lemma:

\[
U_\theta(\theta) = -\psi'(\theta(\theta)) - \pi_\theta(\theta, s)
\]

The local second-order condition for truth-telling is:

\[
y_{\theta \theta}(\theta) - \psi'(\theta - C(\theta))C_\theta^2 \leq 0
\]
Differentiating the local first-order necessary condition for truth-telling (which is identically zero for all \( \theta \)), the second-order condition is equivalent to:

\[
C_\theta(\theta) \geq 0
\]

That is, the cost function must be non-decreasing in type.

The requirement for global incentive compatibility follows from the weak-reversal property: for any pair \( \theta, \theta' \in \Theta \):

\[
U(\theta, \theta) \geq U(\theta, \theta') \quad \& \quad U(\theta', \theta) \geq U(\theta', \theta')
\]

Adding these two and integrating gives:

\[
\int_{\theta}^{\theta'} \int_{C(\theta')}^{C(\theta)} \psi''(x - y) \, dx \, dy \geq 0
\]

Since \( \psi'' \geq 0 \), it follows that provided \( \theta \geq \theta' \), then \( C(\theta') \geq C(\theta) \). Hence, non-decreasing \( C(\theta) \), or \( e_\theta \leq 1 \) is a necessary condition for incentive compatibility. ■

### 2.7.2 Proof of Lemma 2

For partitions \( \Theta_I \) and \( \Theta_{III} \), this is straightforward. Consider partition \( \Theta_{II} \). Recall that:

\[
\nu_\theta(\theta) = \left( -e_\theta^\nu - \frac{\pi_{\theta \theta}}{\psi''} \right) / e_\mu^\nu
\]

It is convenient to make the following definition: \( V := (1 + \lambda)(1 - \psi'(e)) \), then equation (E) is employed to recover \( e_\mu^\nu \): First, rearranging gives:

\[
\mu(\theta) = V f / \psi''(e)
\]

Differentiating this expression with respect to \( \theta \) yields:

\[
\psi''(e_\theta^\nu + e_\mu^\nu \nu_\theta) = \frac{1}{1 + \lambda} \left( \frac{\mu f' - \mu_\theta f}{f^2} \right) \psi'' + O(\psi''')
\]
2.7. APPENDIX

Taking \( \psi'' \approx 0 \) and rearranging this expression to solve for \( e''_\mu \) yields:

\[
e''_\mu = \frac{\mu f' - \mu_\theta f - \psi''_\mu (1 + \lambda) f^2}{(1 + \lambda) f^2 \mu_\theta}
\]  

(2.23)

Where the last equality follows from the definition of \( V \). Now replacing \( e''_\mu \) in equation (2.21) yields:

\[
\nu_0 = (f'V + f(1 + \lambda)\pi_{\theta\theta}) / \psi''
\]

(2.24)

To establish that \( \nu_0 \leq \lambda f \), first note that:

\[
\nu_0 - \lambda f = \frac{V f' - \pi_{\theta\theta}(1 + \lambda) f}{\psi''} - \lambda f
\]

\[
= \frac{f'\mu}{f} + \frac{f(1 + \lambda)\pi_{\theta\theta}}{\psi''} - \lambda f \quad \text{(using definition of } V \text{)}
\]

\[
= -f \left( \lambda - \frac{f'\mu}{f^2} \right) + \frac{f(1 + \lambda)\pi_{\theta\theta}}{\psi''}
\]

Now recall \( \mu(\theta) \in \{\lambda(F(\theta) - 1), \lambda F(\theta)\} \) and that \( \frac{d}{d\theta} \left( \frac{\psi''}{f'} \right) = \lambda - \frac{f'}{f} f'' > 0 \) by the monotone hazard rate assumption on the distribution of types. So the last line can be written as:

\[
\nu_0 - \lambda f = -f \frac{d}{d\theta} \left( \frac{\mu}{f} \right) + \frac{f(1 + \lambda)\pi_{\theta\theta}}{\psi''}
\]

(2.25)

Now it can be seen from equation (2.25) that if \( \pi_{\theta\theta} \leq \psi'' \frac{\lambda}{1 + \lambda} \frac{d}{d\theta} \left( \frac{\psi''}{f} \right) \), then \( \nu_0 - \lambda f \leq 0 \).

To establish that \( e''_\theta(\theta) \leq 0 \), recall that the effort profile on partition \( \Theta_{III} \) depends on the value of the \( \nu(\theta, s) \) costate solution, therefore: \( \frac{de''}{d\theta} = \frac{\partial e''}{\partial \theta} + \frac{\partial \psi''}{\partial \nu} \nu_\theta \). Using (2.21), this yields:

\[
\frac{de''}{d\theta} = -\frac{\pi_{\theta\theta}}{\psi''}
\]

(2.26)

So for the effort profile to be downward sloping on partition \( \Theta_{III} \) requires \( \pi_{\theta\theta} \geq 0 \) since \( \psi'' > 0 \) by assumption.

2.7.3 Proof of Lemma 3

Differentiating equation (S) with respect to \( s \) yields: \( \nu_s = -\frac{\pi_{\theta\theta}}{\psi'' e_\nu} \). Differentiating equation (E) with respect to \( \nu \) yields: \( \psi'' e_\nu = -\frac{1}{(1 + \lambda) f} \psi'' + O(\psi''') \). Hence \( e_\nu < 0 \). Since \( \psi'' > 0 \)
and $\pi_{\theta} < 0$ by assumption, then $\nu_s < 0$. That is, the $\nu(\theta, s)$ solution shifts down for greater values of the private market parameter, $s$. ■

2.7.4 Proof of Corollary 1

The result follows directly from consideration of the conjectured solution of the costate variable (equation (2.11)), and Figure 2.4. Since $\nu \in \{\lambda F, \lambda(F - 1)\}$, then take (say) $\tilde{\theta} \in \text{int}(\Theta_{II})$ in case II. Then it follows that $0 < \nu(\tilde{\theta}) < \lambda F(\tilde{\theta})$. Hence, the distortion implicit in the first order condition for optimal effort is smaller, so $e^v$ solutions are closer to the first best level. The result is analogous for case IV. ■

2.7.5 Proof of Corollary 2

First, consider a value of $s$ like $s^*$ such that on an interval of cost types, $U_0(\theta) \leq 0$,

$$U(\theta) = \begin{cases} \int_0^{\tilde{\theta}(s^*)}(\psi' + \pi_\theta)d\tilde{\theta} & \text{for } \theta \leq \tilde{\theta}(s^*) \\ 0 & \text{for } \theta \geq \tilde{\theta}(s^*) \end{cases}$$

So differentiating this expression with respect to the private market parameter yields:

$$\frac{\partial U(\theta)}{\partial s} = \left[ \psi'(c(\tilde{\theta}(s^*))) + \pi_\theta(\tilde{\theta}(s^*), s^*) \right] \frac{\partial \tilde{\theta}(s^*)}{\partial s} + \int_\theta^{\tilde{\theta}(s^*)} \pi_{\theta\theta}(\theta, s^*)d\tilde{\theta}$$

The expression labeled (a) is positive by assumption that $U_0(\theta) \leq 0$ - the direct effort incentive dominates the countervailing incentive. To sign the term (b), recall that the critical value $\tilde{\theta}$ is the cost type at the border of partition I and II (in this case): $\nu(\tilde{\theta}, s^*) = \lambda F(\tilde{\theta})$. Differentiating this equation gives the sensitivity of the critical cost type to the private market parameter:

$$\frac{d\tilde{\theta}(s)}{ds} = \frac{-\nu_s}{\nu_\theta - \lambda f} < 0$$

since Lemma 2 says that $\nu_\theta \leq \lambda f$ and Lemma 3 says that $\nu_s < 0$. Hence (b) is negative.

Since $\pi_{\theta\theta}(\theta, s) \leq 0$ by assumption, the integral over this function is negative. Hence (c) is negative. Consequently, an increase in the private market parameter decreases information rents: $U_s(\theta) \leq 0$ when $\pi_{\theta\theta} \in [0, \hat{\pi}_{\theta\theta}]$. The proof for the critical value on the
border of partitions II is straightforward. For III, term (a) is negative, (b) is negative, and the endpoints of the integral in (c) are switched, so that (c) is positive. ■

2.7.6 Proof of Lemma 4

Differentiating equation (2.18) with respect to \( \theta \) yields:

\[
\frac{dY^*}{dc} = y_0^*(\theta) \frac{1}{c_0} = -\psi'(e^k) \leq 0 \quad \text{and} \quad \frac{d^2Y^*}{dc^2} = -\frac{\psi''(e^k)e^k}{c_0} \geq 0, \quad k = \{L, H, \nu\}
\]

where the first order condition for truth-telling outlined in equation (2.20) has been used in the left hand equation, and the result of Lemma 2 has been used in the right hand side equation. Hence, as stated, the net-transfer function is decreasing and convex for each of the separate solution profiles. To show continuity for a patched solution, take type \( \theta \) on the border of region I and II. Since \( \lim_{\theta \to \hat{\theta}^-} e^L(\hat{\theta}) = e^\nu(\hat{\theta}) \) and \( \lim_{\theta \to \hat{\theta}^+} e^\nu(\theta) = e^L(\hat{\theta}) \) then:

\[
\lim_{\theta \to \hat{\theta}^-} y^L(\hat{\theta}) = \lim_{\theta \to \hat{\theta}^-} (\psi(e^L(\hat{\theta})) + \pi(\hat{\theta})) = \psi(e^\nu(\hat{\theta})) + \pi(\hat{\theta}) = \lim_{\theta \to \hat{\theta}^+} y^\nu(\hat{\theta})
\]

and \( \psi(e^L(\hat{\theta})) = \psi(e^\nu(\hat{\theta})) \) exist. Now for differentiability, note that:

\[
\lim_{\theta \to \hat{\theta}^-} \frac{dY^*}{dc} = \lim_{\theta \to \hat{\theta}^-} -\psi'(e^L(\hat{\theta})) = -\psi'(e^\nu(\hat{\theta})) = \lim_{\theta \to \hat{\theta}^+} \frac{dY^*}{dc}
\]

Also, \( \psi'(e^L(\hat{\theta})) = \psi'(e^\nu(\hat{\theta})) \) exist. The same method can be applied to the border of region II and III. ■

2.7.7 Proof of Lemma 5

Differentiating equation (E) yields:

\[
e^k_\theta = -\frac{1}{1 + \lambda \frac{d}{d\theta} \left( \frac{\mu}{f} \right)}, \quad k = \{L, H, \nu\}
\]

Take \( j = L, H \). At \( \theta = \hat{\theta} \), from equation (E): \( e^j(\hat{\theta}) = e^\nu(\hat{\theta}) \) since \( \mu^*(\hat{\theta}) = \nu(\hat{\theta}) \), but \( e^j_\theta \neq e^\nu_\theta \). To see this, note that:

\[
\frac{d}{d\theta} \left( \frac{\mu}{f} \right) = \lambda - \frac{\mu}{f^2} f' \quad \text{whereas} \quad \frac{d}{d\theta} \left( \frac{\nu}{f} \right) = \frac{\nu_\theta}{f} - \frac{\nu}{f^2} f'
\]
For an interval of types to be binding at the optimum, the required condition is \( \nu_\theta \leq \lambda f \).

Hence,

\[
\frac{d}{d\theta} \left( \frac{\nu(\theta)}{f} \right) = \frac{\nu(\theta)}{f} \frac{d}{df} f' \leq \lambda - \frac{\nu(\theta)}{f^2} f' = \lambda - \frac{\mu(\theta)}{f^2} f' \iff \frac{d}{d\theta} \left( \frac{\mu(\theta)}{f} \right) = \frac{d}{d\theta} \left( \frac{\nu(\theta)}{f} \right)
\]

Where the first inequality comes from Lemma 2 which says that \( \nu_\theta(\theta, s) \leq \lambda f(\theta) \), and the second equality follows from the definition of the costate variable at the border of the partitions: either \( \nu(\theta) = \lambda F(\theta) \) or \( \nu(\theta) = \lambda (F(\theta) - 1) \). It follows that \( e_\nu^L(\theta) \leq e_\nu^V(\theta) \).

Now recall that:

\[
\frac{d^2 Y^*}{dc^2} = -\psi''(e^k) \frac{e_\theta^k}{c_\theta} = -\psi''(e^k) \frac{e_\theta^k}{1 - e_\theta^k} = \psi''(e^k) \left| \frac{e_\theta^k}{1 - e_\theta^k} \right|, \quad k = \{L, H, V\}
\]

Since \( e_\theta^L < e_\theta^V < 0 \) for \( j = L, H \), then it follows that \( \frac{e_j^L}{1 - e_j^L} < \frac{e_j^V}{1 - e_j^V} \). Again, since the slope of the effort profiles are negative regardless of mode,

\[
\left| \frac{e_\theta^L}{1 - e_\theta^L} \right| > \left| \frac{e_\theta^V}{1 - e_\theta^V} \right|
\]

Using this inequality and the expression for the second cost derivative of the net-transfer function from above, it is straightforward to see that:

\[
\frac{d^2 Y_j}{dc^2} > \frac{d^2 Y_V}{dc^2}, \quad j = L, H
\]

Hence, the net-transfer function is less convex on subsets of cost types where the costate solution is \( \nu(\theta, s) \) - which is where the IR binds. □

### 2.7.8 Proof of Corollary 3

From equation (2.17) and (2.26),

\[
\frac{d e^\nu}{d\theta} = -\frac{1}{1 + \lambda} \frac{d}{df} \left( \frac{\nu}{f} \right) = -\frac{\pi_\theta}{\psi''}
\]

The result follows from requiring \( 0 \leq \pi_\theta \leq \frac{\psi''}{1 + \lambda} \frac{d}{df} \left( \frac{f}{f} \right) \). □
Chapter 3

Franchise Contracts with Limited Liability

3.1 Introduction

The interaction of vertically related firms is typically complicated by the presence of externalities and informational asymmetries. For example, the contract written between a manufacturer and retailer may need to account for the possibility that the retailer may not internalise the impact of its pricing decision on the manufacturer, or deal with opportunism that can arise when the retailer has an informational advantage over the manufacturer. It is well known that the inefficiencies that arise from externalities can be resolved by the use of contractual provisions, known as vertical restraints. Moreover, the performance of those vertical restraints under information asymmetries has been studied in a number of recent examinations using Principal-Agent models, where contracts involve a system of incentives and penalties.\(^1\) However, there are two feature of franchise arrangements that have so far received little attention. First, retailers may often be bound by limited liability constraints. Hence, the manufacturer may be restricted in its enforcement of harsh penalties. Second, the retailer may be able to take an action that favourably influences the level of demand it ultimately faces, or may have private information about demand that is imprecise at the time of signing the contract.

\(^1\)See, for example, Rey and Tirole (1986), Blair and Lewis (1994) and Martimort and Piccolo (2007).
CHAPTER 3. FRANCHISE CONTRACTS

This chapter explores a franchise relationship under uncertainty when the retailer (i) has limited liability to comply with the terms of the contract once it realises its precise level of demand, and (ii) only learns private information relating to the precise demand state subsequently to signing the contract. At the time of signing the contract, the retailer and manufacturer share a common belief about the probability distribution of demand states, and both know that only the retailer will eventually learn the precise demand state. This informational asymmetry confers a strategic advantage to the retailer, which the manufacturer can resolve by incorporating self-selection, or incentive constraints into its contracts. These constraints ensure that the retailer has an incentive to reveal its true private information to the manufacturer. The resulting contracts are administered through a fee schedule: the retailer selects a quantity from the schedule to sell and pays a corresponding fee. However, in this chapter, the interaction of the incentive constraints on the manufacturer's contracts, and the impact of the retailer's limited liability on the capacity for those contracts to be implemented give rise to an expected surplus to the retailer that stems from its informational advantage. This chapter demonstrates how the manufacturer can extract this expected surplus with an additional vertical restraint in its contracts: an ex ante franchise fee, that together with a fee schedule, forms a contract that is referred to in this chapter as a franchise contract.

There are a number of reasons to expect that retailers are bound by limited liability. First, limited liability may be the result of legal implications that are embedded in local competition policy legislation. It may also follow from consideration of the retailer's operating decision, or the decision to shutdown. If a franchisee finds that its average operating revenue from executing the contract is less than the amount to be paid under

---

2There are a variety of different vertical relationships that can be classed as franchises. Broadly speaking, franchises can be separated into three categories: manufacturing, product and business franchises. Manufacturing franchises typically involve a technology licensing arrangement, where manufacturing firm holds intellectual property rights to a technology that it may license another manufacturer to use. A product franchise may be a distributorship, or a licensed dealer or reseller of a manufacturer's product. In a business franchise, the upstream firm typically issues the right for a downstream firm to use its business plan to retail to consumers.

3In Australia there are a number of provisions that amount to limited liability in the Trade Practices Act (1974). For instance, Australian franchise agreements must meet a Franchise Code of Conduct, which precludes the use of harsh penalties. Another reason for limited liability that is perhaps more applicable to business franchises is the size of the franchisees assets: typically such firms are small operators and do not have large assets.
3.1. *INTRODUCTION*

the terms of the contract, it may opt out. Limited liability only binds when the manufacturer would like to incorporate penalties into the contract — that is, when there is uncertainty.

Uncertainty arises here over the precise level of demand in the final market. Specifically, the precise state of final demand for the industry's output is uncertain to both the manufacturer and the retailer at the time of signing the contract, but the retailer privately observes the precise demand state in the process of serving its market. Furthermore, this chapter incorporates a realistic aspect of franchise arrangements: the retailer may either have some coarse private information before signing the contract (for example, a survey of the composition of consumers in the market), or can take an action that influences the likelihood of the demand state faced (for example, a targeted marketing campaign).

The results will demonstrate how the franchise fee component of the conjectured franchise contract can be used to extract the retailer's perceived expected gross surplus from entering the contract. Moreover, the model explores the contracting possibilities when the retailer has the ability to engage in opportunistic behaviour before the precise demand state that it will face has been realised. It is shown how the manufacturer can use the franchise fee as a device to elicit the efficient action in the case of moral hazard, or as a screening instrument in the case of ex ante hidden knowledge. In this latter case, the results also offer an interesting insight into the value to the manufacturer of obtaining the retailer's information. The manufacturer can weigh up the expected benefit of paying the retailer for its information against the expected efficiency cost of implementing more finely-tailored contracts.

The next section reviews some of the related literature in vertical contracting. Then, to establish the intuition for the existence of the expected surplus that arises from the interaction of the retailer's limited liability constraints and incentive compatibility, a simple discrete demand state illustration is examined. Two hypotheses are drawn from this simple setting that are scrutinised in the general model of the following section. The general model then allows the issue of ex ante private information to be rigorously explored. The last section concludes.
3.2 Relation to the Literature

It is well known following Spengler (1950) that when vertically related firms, each with some degree of market power, make independent linear pricing decisions, a vertical externality emerges: each firm fails to take into account the effect of its own price on the other’s profit. The result is an inefficiency in the form of a loss of producer surplus relative to an integrated structure. The price in the market for the industry’s output embodies two mark-ups over marginal cost: double marginalisation. Any arrangement that can eliminate one of the mark-ups necessarily improves allocative efficiency. Contractual arrangements known as vertical restraints are one way for vertically related firms to restore allocative efficiency.

Contractual provisions allow for the manufacturer to correct the vertical externality. Tirole (2000) details how vertical restraints broadly include the use of resale price maintenance, quantity fixing, and use of franchise fee arrangements. These restraints respectively allow the manufacturer to control the retailer’s price or quantity directly, or decentralise the pricing decision in a way that implements the integrated profit solution which the manufacturer and retailer can bargain over with the use of a fixed fee transfer.

With the extreme bargaining power assumption of the Principal-Agent framework, the manufacturer will appropriate all the rent under complete information with a fixed fee. However, when the relationship is exposed to uncertainty, there is scope for the retailer to behave opportunistically. Subsequent analysis must appeal to the theory of incentives.

Matthewson and Winter (1985) were among the first to study incentives in a business franchise arrangement. Two particular features of their model stand out as relevant for the study of this chapter. First, they conclude that the presence of a vertical externality is essential for the use of franchise contracts to resolve agency issues. In contrast they show that horizontal externalities, which give rise to effects such as under-provision of services between competing downstream firms, do not motivate the use of franchise con-

---

4There is a vast literature related to the theory and policy of vertical restraints. Much of the literature is devoted to a welfare analysis of the various vertical restraints that are used in practice in many industries, and the implication for competition policy. The results are not clear cut, as evidenced by the many changes that competition agencies have made to their legislation over the years. This chapter does not attempt to evaluate the welfare properties of vertical restraints, but rather conjectures a variety of contract that could be employed by a manufacturer in a realistic contracting environment with a retailer.
tracts. Only a vertical externality is at the heart of the results of this chapter, since the manufacturer exploits the inefficiency from double marginalisation to reduce the expected information rent it inevitably must pay out.

Second, Matthewson and Winter describe a positive relationship between the variance of demand and the variance of the nonlinear tariffs that emerge optimally in their model. This result partly reflects their assumption that the retailer is risk averse. However, it also arises because they compare demand distributions through the lens of second order stochastic dominance: where a more favourable distribution is one with smaller variance. In this chapter, demand state distributions are compared by first order stochastic dominance: distributions are ranked by their mean. Moreover, here both the manufacturer and retailer are risk neutral. Notwithstanding, this analysis arrives at a similar intuition for the fixed franchise fee. The more favourable the distribution of demand states, the lower the franchise fee. However, in contrast to Matthewson and Winter, this chapter finds that the variable fees are higher for the favourable distribution, since the manufacturer is more intent on extracting information rents in that case.

Other papers that are close to this analysis are Crocker (1983), Rey and Tirole (1986), Blair and Lewis (1994), and Martimort and Piccolo (2007). Crocker (1983) describes a vertical contracting environment where the retailer already has private information about its production cost at the time of signing the contract. The thrust of his analysis is to weigh the transaction cost of the agency relationship, measured in terms of the rent extraction - efficiency trade-off, against the benefit of eliminating it through vertical integration. His results are similar to the second stage contracts of this chapter, however here the retailer signs the contract before realising the precise demand state, so the timing is different.

The timing of the revelation of the uncertainty in Rey and Tirole (1986) is similar to that employed here: manufacturers and retailers are symmetrically informed about the precise demand state at the time of contracting, and retailers become privately informed about the precise demand state after the contract is signed. Hence, the retailer’s attitude to risk in their paper plays a role in the efficacy of contractual provisions, and varies under different types of uncertainty. Their analysis chiefly describes the efficacy of
contractual provisions in achieving the integrated profit within a vertical structure under uncertainty and retailer risk aversion. However, Rey and Tirole's model does not capture the intertemporal effect of the retailer's action on the distribution of demand states as the model presented here does. Nor does it explicitly analyse the manufacturers trade-off between extracting rents from high demand retailers on the one hand, and deliberately employing double marginalisation on the other to induce an allocation inefficiency to penalise misreporting.

In fact, the timing of the model presented here more closely resembles that of the information acquisition process in the optimal contest design literature. For example, in Fu and Lu (2010) an entry fee or subsidy induces the effort-maximising contest while extracting all the surplus from the contestants in the subgame perfect equilibrium of the game. This will prove to be similar to the role of the franchise fee in the contracts of the model presented later. In contrast however, in this model the second stage involves a revelation game where the retailer privately observes its demand, hence a rent-extraction efficiency trade-off is required for incentive compatibility.

The rent extraction-efficiency trade-off is dealt with in Blair and Lewis (1994), and Martimort and Piccolo (2007). Their papers analyse vertical delegation in a model of simultaneous adverse selection and moral hazard. In their models, the retailer is already privately informed at the time of contracting, and moral hazard is defined over the degree of promotional effort the retailer exerts to increase the demand for their product. Blair and Lewis focus on establishing resale price maintenance as a provision that emerges endogenously to the informational environment of their model. Their results are similar to the second stage results of the model here, in the respect that the manufacturer optimally employs the vertical externality to their overall advantage. However, both models differ from the one presented here in the respect that moral hazard in their models takes place after the retailer has learnt their private information, whereas in this chapter it takes place before. Moreover, it is this difference in the timing combined with

---

5Since there is no uncertainty on the part of the retailer, attitude to risk does not play any role in their results, unlike Rey and Tirole (1986).

6Martimort and Piccolo (2007) is technically similar to Blair and Lewis (1994), however their focus is on the difference between resale price maintenance and quantity fixing and the implications for welfare from both private and social perspectives.
limited liability constraints on payoffs for realised demand states that creates an expected surplus for the retailer that is central to the results of this chapter.

The impact of limited liability constraints in agency contracts was described in Sappington (1983). His analysis showed that in a model where the retailer signs the contract before realising its private information, limited liability constraints restrict the manufacturer’s ability to impose penalties that would otherwise induce the socially efficient outcome. Hence, limited liability constraints result in a pattern of information rents that is similar to the case of interim contracting. The consequences for limited liability are very similar in this chapter: ex post constraints on the payoff to the retailer in the lowest demand state skew the information rents to higher demand state types. It is exactly this skewed pattern of information rents that creates the context for a fixed franchise fee to be incorporated into the contract.

3.3 Discrete Demand State Illustration

This section sketches the nature of the conjectured franchise contracts for discrete demand states. First, the double marginalisation problem is set out for a single demand state in a non-integrated manufacturer-retailer hierarchy and compared to a vertically integrated structure. This is followed by a brief discussion of how vertical restraints can improve efficiency. Then a two demand state problem under complete information is constructed to obtain the first-best contracts. Moving to a situation of incomplete information shows how the retailer can strategically use its private information to its advantage, and how the manufacturer employs the inefficiency from double marginalisation to ensure incentive compatibility, thus improving its ex ante position. Finally, the retailer’s limited liability constraint is shown to interact with incentive compatibility to give rise to an expected surplus accruing to the retailer, which the manufacturer can appropriate with a franchise fee.

3.3.1 Double Marginalisation

To illustrate the inefficiency that occurs in this framework, first consider a non-integrated industry that faces a single known demand state where both the manufacturer and retailer
are restricted to use linear pricing. Panel (a) on the left hand side of Figure 3.1 depicts the consumer demand curve, \( D \), which the retailer faces. Suppose the retailer’s marginal cost of retailing is zero. Then the manufacturer faces a derived demand, \( D_M \) which comes from the retailer’s profit maximisation problem. Hence, the \( D_M \) curve in Figure 3.1(a) is both the manufacturer’s derived demand curve and the retailer’s marginal revenue curve, \( MR_R \). Since the manufacturer is a monopolist in the upstream market, it maximises its profit by selling to the retailer a quantity \( x^M \) where its marginal revenue \( MR_M \) from the sale of that quantity is equal to its constant marginal production cost, \( c > 0 \). The manufacturer charges the wholesale price \( p_w \) per unit to the retailer, and earns wholesale profits equal to the area \( \pi^M \), depicted in panel (b) for clarity.

Since the manufacturer’s problem came from a derived demand curve, it internalised the retailer’s problem in its quantity decision. Since the retailer is a monopolist in the downstream market, it resells the quantity \( x^R \) which it would have optimally chosen if its marginal cost was \( p_w \), and earns a retail profit equal to the area between its marginal revenue and marginal cost curves, denoted as \( \pi^R \) in Figure 3.1(b). The price each unit is sold for can be read off the final demand curve at \( x^R \) units of the quantity, which will be denoted \( p^R \). Hence, the final price faced by consumers includes two markups over marginal cost: the manufacturer’s markup \( p_w > c \) and the retailer’s markup: \( p^R > p_w \).
3.3. DISCRETE DEMAND STATE ILLUSTRATION

This is Spengler's (1950) double marginalisation phenomenon.

In an integrated structure, a monopolist would choose the quantity $x^i$ where the marginal (retail) revenue of the last unit sold is equal to the marginal (production) cost. In doing so, it would earn profit equal to the area $\pi^R + \pi^M + el$ in Figure 3.1(b) by charging the per unit integrated price $p^i$, as in panel (a). By comparing to the non-integrated case, it is straightforward to see that by making sequential markups over their respective marginal costs, the manufacturer-retailer hierarchy sells for a higher price: $p^{ni} > p^i$, and sells too few units into the final market: $x^{ni} < x^i$ relative to an integrated structure. The area $el$ is a measure of the relative allocative efficiency loss on the producer side to a manufacturer-retailer hierarchy from the vertical externality.7

3.3.2 Vertical Restraints

If the manufacturer is no longer restricted to use linear prices, it can employ various contractual provisions known as vertical restraints to overcome the efficiency loss on the producer side from double marginalisation. Vertical restraints include instruments like resale price maintenance, quantity fixing, and non-linear pricing techniques which are generally referred to as franchise fees. In an environment of complete information, all these types of vertical restraints achieve the same outcome.8 For example, if the manufacturer stipulated that the retailer must fix the consumer's price to the integrated monopoly price $p^i$ - resale price maintenance - then the wholesale price can be used simply to distribute the integrated monopoly profit between the manufacturer and retailer. A quantity fixing contract, which stipulates that $x^i$ units must be sold, has the same effect. Both these methods require the manufacturer to exert direct control over the behavior of the retailer.

The manufacturer can also decentralise the retail decision making and still implement the integrated outcome by selling using a two-part tariff. It could sell as many units to the retailer as it wanted to purchase at marginal cost - in which case the retailer optimally

---

7Note that there is an associated loss of consumer surplus from the double markup, and that consumer's prefer the integrated case since more output is then sold.

8With incomplete information, different vertical restraints give different outcomes according to the nature of the informational asymmetry and the existence of intra-brand or inter-brand competition. See Rey and Tirole (1986).
purchases \(x^i\) units - and charge a fixed fee to extract all or part of the integrated profit that the retailer would otherwise earn. In such a set up, the fixed fee which the retailer and manufacturer bargain over is called the franchise fee.

In what follows the manufacturer is assumed to offer the retailer a type of franchise fee contract, however it will be slightly more restrictive: the manufacturer will infer the integrated monopoly quantity contingent on the demand state, and ask the retailer to sell that quantity into the final market in return for a fixed fee equal to the entire revenue obtained.\(^9\) Note that this implicitly confers full bargaining power to the manufacturer: the retailer will make zero economic profit, and the manufacturer will receive the entire integrated monopoly profit. Hence, the manufacturer will offer a contract for each demand state consisting of a fee to be paid and a quantity to be sold: \(\{t, x\}\). The set of these contracts constitutes the fee schedule.

### 3.3.3 Franchise Contracts

Now suppose that the level of demand can take one of two possible values: high or low. Figure 3.2 depicts the two demand state case, where the demand curves have been omitted for clarity. If there is complete information about which demand state the retailer faces, then the manufacturer will offer one of two contracts. If the demand is low, then the contract is: \(\{CAc, x_L^{fb}\}\). If the demand is high, then the manufacturer offers contract: \(\{DBc, x_H^{fb}\}\). Notice that, because the marginal revenue curve is further out from the origin for the high demand type than the low demand type, \(x_H^{fb} > x_L^{fb}\), and \(DBc > CAc\).

If it is only the retailer that knows the state of demand, then there is incomplete information. In this case, the manufacturer knows that there are two possible states of demand, but not which is the prevailing state. The manufacturer has the common knowledge belief that demand is high with probability \(g\), and low with probability \(1 - g\), where \(0 \leq g \leq 1\). Moving to a situation of incomplete information introduces an information revelation problem: if the manufacturer continues to offer the first best contracts, a retailer facing the high demand state actually prefers to take the low demand

\(^9\)An equivalent resale price maintenance or quantity fixing contract could be found, however it would involve setting an appropriate wholesale price in each case. It will be notationally simpler to use fixed fees in the sequel.
3.3. DISCRETE DEMAND STATE ILLUSTRATION

Figure 3.2: Incentive to Mis-report

contract. Thus, the manufacturer faces an adverse selection problem. The reason is illustrated in Figure 3.2. By reporting that the state of demand is low, a high demand retailer earns an information rent equal to the difference in the total revenue they can achieve, $DEx^b_L$, and the fixed fee it must pay the manufacturer, $CAc$. Hence, the high demand retailer earns information rent equal to $DEAC$. Note that the source of the information rent is from the high demand retailer selling the low level of output for a higher price than the integrated monopoly price: the vertical externality returns, albeit in a slightly different guise.

In this situation, the manufacturer's payoff is always $CAc$, since the high demand contract would never be taken. However, if the manufacturer offered to give up at least the rent $DEAC$ in the case of high demand, then the retailer in that instance would weakly prefer to accept the high demand contract. Hence, the manufacturer would prefer to offer the separating contracts $\{CAc, x^b_L\}, \{DBc - DEAC, x^b_H\}$ only if:

$$E\pi = (1 - g)CAc + g(DBc - DEAC) \geq gDBc$$

That is, it pays the manufacturer to use separating contracts if the expected payoff from inducing the high demand retailer to report the true state of demand exceeds the payoff from only ever selling the low demand quantity.

Inspection of Figure 3.2 and the manufacturer's expected payoff from eliciting a truth-
ful report from the high demand retailer suggests that the manufacturer can improve their ex ante position. It can do this by imposing an efficiency loss from double marginalisation on the low demand retailer. By reducing the quantity its asks the retailer to sell in the low demand state to the second-best quantity $x_L^b$, the size of the information rent that the manufacturer has to give up to the high demand retailer reduces to $DFGC$ in Figure 3.3. The cost to the manufacturer of exploiting the vertical externality is a reduction in the revenue it can ask the low demand retailer to pay by area $GAx_L^b x_L^a$ in Figure 3.3. Hence, the manufacturer trades off extracting the information rent from the high demand retailer against an allocative efficiency loss from the low demand retailer. The manufacturer prefers second-best contracts whenever:

$$E\pi^b = (1 - g)CAc + gDBC - ((1 - g)GAH + gDFGC) \geq gDBC$$  \hspace{1cm} (3.1)

In the expression for expected profit in equation (3.1), the first term (a) corresponds to the first-best expected efficiency. Out of this, the manufacturer expects to lose efficiency from optimally exploiting the vertical externality for low demand types - term (b), and expects to payout an information rent to the high demand type - term (c). Importantly, the manufacturer controls the low demand retailer's quantity to improve their position. Note also that the magnitudes of the distortion in the low demand quantity and information rent depends on the probability distribution over the demand states.

Now suppose that before information is revealed to either party, the retailer can take a private action to influence the probability of the high demand state occurring. Specifically, suppose that the retailer has an action space containing two actions: $a, a' \in A$. Furthermore, suppose that action $a$ induces a greater probability of the high demand state occurring than action $a'$: $g(a) > g(a')$. Then action $a$ is said to induce a discrete probability distribution that first-order stochastically dominates the probability distribution induced by action $a'$. This is generalised in the next section to a probability distribution over a continuum of demand states.

Finally, consider the case of incomplete information where neither the manufacturer nor the retailer are informed about the precise state of demand at the time of signing the
3.3. DISCRETE DEMAND STATE ILLUSTRATION

Figure 3.3: Incomplete Information Contracts

contract, but they both share the same objective belief about the likelihood of the demand states. Moreover, suppose that just after the contract is signed the retailer privately realises the state of demand. To evaluate its payoff, the retailer also must form an expectation over their payoffs in each state of demand. Then once again utilising Figure 3.3, the retailer's expected payoff is:

\[ u = (1 - g(a))0 + g(a)DFGC. \]

Note that the retailer's payoff in the low demand state is bound at zero. This is due to the assumption of limited liability. Limited liability hence interacts with incentive compatibility to create a positive gross expected surplus to the retailer. However, if the retailer's outside opportunity is normalised to zero, then the manufacturer can extract the full expected surplus with a franchise fee payable at the time of signing the contract: \( T = g(a)DFGC. \)

Two hypotheses can be drawn from this simple illustration of ex ante contracting and limited liability. First, there is a reciprocal relationship between the size of the inefficiency induced by the vertical externality and the expected surplus \( T = g(a)DFGC. \) This is a straightforward consequence of the timing of information revelation and the basic rent-extraction efficiency result that frames the equilibrium contracts. Second, the more weight the manufacturer assigns to the high demand state occurring, the greater is the inefficiency from the vertical externality that the manufacturer imposes on the low demand type. This hypothesis is the key observation of this chapter. To see why it is important, consider Figure 3.3 again. The marginal expected benefit to the manufacturer
CHAPTER 3. FRANCHISE CONTRACTS

from distorting the low demand quantity is the revenue $FEAG$ from rent extraction, and the marginal expected cost is the loss of revenue $GAH$ from the vertical externality on the low demand retailer. Then increasing the probability of the high demand state occurring increases the manufacturer's relative benefit from distorting the low demand quantity.

From these two hypotheses, it follows that the size of the franchise fee $T$ depends on the probability of the high demand state occurring, which in turn depends on the action taken by the retailer prior to realising the state of the world. Since the action becomes relevant before the retailer privately realises the precise demand state, there is potential for either a moral hazard problem or an adverse selection problem over the action. For instance, if the action space of the retailer consists of a set of costly efforts that the retailer can exert which are unobservable and non-verifiable, say in promoting the manufacturer’s good or service, then the manufacturer’s problem is one of moral hazard followed by adverse selection. If the action space of the retailer consists of a set of reports of signals that the retailer receives which are perfectly correlated with the demand distribution it will be facing, then the manufacturer’s problem is one of sequential adverse selection. In either case, this chapter conjectures that the manufacturer can offer a set of franchise contracts of the form $\{T, (t, x)\}$ that can either induce the efficient effort level, or induce truthful signal report from the retailer. Both these situations are studied later in a more general model.

The timing of the action that the retailer takes and the revelation of the precise demand state gives rise to a commitment game that will be described here for clarity. The game is slightly different depending on whether there is moral hazard or adverse selection from the outset. In the moral hazard case the manufacturer will offer a single franchise contract $\{T, (t, x)\}$. By offering this contract, both the manufacturer and retailer commit to the variable fee schedule $(t, x)$ which is incentive compatible for reports of the precise demand state once it is realised by the retailer, contingent on the efficient action being taken. That is, the retailer must commit to the contract before it knows the precise state of demand. Since the fee schedule also must abide by ex post limited liability, the retailer’s expectation of the rent it will receive from this part of the contract is positive. To solve the commitment problem, the manufacturer must design the franchise fee $T$
3.4 ANALYTICAL FRAMEWORK

in a way that simultaneously elicits the efficient action with respect to which \((t, x)\) is designed, and that extracts the retailer’s positive expected surplus from the variable fee schedule.

For the adverse selection case, the commitment game is essentially the same, except the manufacturer and retailer commit to a menu of variable fee schedules, and the manufacturer must design a corresponding menu of franchise fees that elicit truthful reports about the signal that the retailer receives. In this case, the manufacturer must also consider whether it is worthwhile to extract the retailer’s signal or to remain ignorant. The manufacturer’s beliefs that support it’s commitment to its contracts are explored later.

The next section establishes the two hypotheses in a general model with a continuum of demand states. This then allows a characterisation of the optimal franchise contracts. The section after that discusses the use of the franchise fee component of the franchise contracts to solve ex ante informational asymmetries that might emerge.

3.4 Analytical Framework

The Principal-Agent problem is between a manufacturer and a retailer. The risk neutral manufacturer produces an intermediate product, \(x\), at constant unit cost \(c > 0\). The risk neutral retailer sells each unit supplied to it on a final output market, generating total revenues \(R(x, \theta)\) where \(\theta\) is the retailer’s private information about the precise state of demand, which belongs to the compact support \(\Theta := [\underline{\theta}, \bar{\theta}]\). The retailer’s marginal retailing cost is assumed to be zero, without loss of generality. Also, the retailer is assumed to be a single price monopolist in the final output market.

The revenue function, \(R(x, \theta)\) is concave in sales, \(x\), and increasing in the level of demand, \(\theta\). Moreover, since the retailer is a monopolist in the final market, the relevant part of the total revenue curve is where it is increasing in output. Putting these restrictions more concisely: over the relevant domain of \(x\), \(R_x > 0\), \(R_{xx} < 0\) and \(R_\theta > 0\), where subscripts denote partial derivatives. Importantly, the retailer’s marginal revenue is assumed to be increasing in the realised demand state: \(R_{x\theta} > 0\): this is the Spence-Mirrlees single crossing property. The following technical assumptions are also made: \(R_{x\theta\theta} \geq 0\) and \(R_{\theta xx} \geq 0\) to ensure the contracts are well behaved.
The model has two time periods, indexed by $s = 1, 2$. In the first period, the retailer takes an action, $a \in A$, where $A$ is the set of possible actions that the retailer can take. The action $a$ affects the probability distribution of demand states that the retailer faces in the second period. In the second period, the retailer privately realises the level of demand, $\theta \in \Theta$. Thus, the timing of the revelation of precise demand state information satisfies an ex-ante constraint: the retailer must sign to a contract with the manufacturer before the uncertainty about demand for the product is resolved. Moreover, the contract is signed before the retailer takes any action $a \in A$. Given this timing, the manufacturer must construct its contracts in a way that elicits the action it finds most appropriate, and induces the retailer to truthfully report its information about the level of demand in the second period.

The Revelation Principle permits attention to be restricted to the class of direct revelation mechanisms.\(^\text{10}\) The conjecture of this chapter is that the manufacturer may offer to the retailer a franchise contract, which is defined to be the triple: $\{T, (t(\theta), x(\theta))_{\theta \in \Theta}\}$. The first component, $T$, is a franchise fee that the retailer is liable for when the contract is signed in the first period. The second component is a menu of contracts, $(t(\theta), x(\theta))$, consisting of the variable fee $t(\theta)$ payable by the retailer to the manufacturer for a quantity $x(\theta)$ of the intermediate product that the retailer of type $\theta$ purchases from the manufacturer in the second period. Then the retailer's second period surplus from the fee schedule $(x(\theta), t(\theta))$ is given by:

$$U(\theta) := R(x(\theta), \theta) - t(\theta) \quad (3.2)$$

The fee and quantity profile represent a fee schedule that the manufacturer and retailer commit to using when the contract is signed. The exact timing of the contracts are detailed in Figure 3.4.

Stage 1 encompasses the first three steps, where along the equilibrium path, the contract is offered and accepted and the retailer chooses some action $a \in A$. Stage 2 encompasses the last steps in the contract, where the demand state $\theta$ is privately realised.

\(^\text{10}\)Laffont and Martimort (2002) show that the Revelation Principle can be extended to cover contracts with sequential information revelation (p. 274).
by the retailer and the retailing decision is made according to the fee schedule.

The manufacturer chooses its contracts to maximise its expected payoff given by:

\[ V = T + \int_{\Theta} (t(\theta) - cx(\theta))g_a(\theta)d\theta \]

The manufacturer designs the contract using backward induction. Hence, the fee schedule \((t(\theta), x(\theta))\) is chosen first to maximise the second term in its expected payoff above. Then, given this second period behavior, the manufacturer optimally chooses the franchise fee that elicits its most preferred, or efficient, action from the retailer through use of the franchise fee component. The retailer’s first stage action influences the probability distribution over demand states in the sense of first order stochastic dominance.

More precisely, when the retailer exerts action \(a \in A\), the demand state \(\theta \in \Theta\) is distributed according to the continuous, conditional cumulative distribution function \(G_a(\theta) := G(\theta|a)\) on the compact interval \(\Theta := [\underline{\theta}, \bar{\theta}]\) with conditional density function \(g_a(\theta) := g(\theta|a) > 0\). For technical convenience, \(g(\theta|a) = g(\theta|a')\) and \(g(\bar{\theta}|a) = g(\bar{\theta}|a')\) for \(a, a' \in A\). These distributions are assumed to satisfy the monotone hazard rate condition:

\[ (MHR) \quad \frac{d}{d\theta} \frac{g(\theta)}{1 - G(\theta)} \geq 0, \quad \forall \theta \in \Theta \]

Some actions are assumed to induce favorable demand state distributions, so a means of ranking probability distributions is needed. For this purpose, the cumulative distribution
property is required to satisfy hazard rate dominance. If the distribution induced by action \( a \) is more favourable distribution than the distribution induced by action \( a' \), then:

\[
(HRD) \quad \frac{g_a(\theta)}{1 - G_a(\theta)} \leq \frac{g_{a'}(\theta)}{1 - G_{a'}(\theta)}, \quad \forall \theta \in \Theta
\]

HRD means that the hazard rate of the favourable distribution is point-wise lower than the hazard rate of the unfavourable distribution. Intuitively, this means the conditional probability of eliminating a demand type in the interval \([\theta, \theta + d\theta]\) is smaller for the more favorable distribution.

A convenient implication of assuming HRD is that it also implies that demand realisations are correlated with the action of the retailer in the sense of first-order stochastic dominance. That is, an action \( a \) induces a more favourable distribution than action \( a' \) if:

\[
(FOSD) \quad G_a(\theta) \leq G_{a'}(\theta), \quad \forall \theta \in \Theta
\]

Lemma 1 proves this stochastic ordering:

**Lemma 1** *(Stochastic Orders)* HRD \( \Rightarrow \) FOSD.

The next section establishes the manufacturer’s second period fee schedule as the solution to a second period maximisation problem that takes as given an action \( a \in A \) that the retailer takes in the first period. Hence, the optimal second period fee schedule is the \( a\)-contingent solution \((t^a(\theta), x^a(\theta))\) for some \( a \in A \). After that, the franchise fee is derived for various action spaces \( A \).

### 3.4.1 Stage 2: Fee Schedule

The limited liability constraint requires retailers to *realise* a level of surplus at least as good as their outside opportunity, which is normalised to zero. Note that limited liability is different from an individual rationality constraint. Individual rationality applies to the entire contract, so will constrain the franchise fee, as will be shown in the next section. Individual rationality will therefore apply before the demand state \( \theta \) is realised by the retailer, whereas the limited liability constraint must be satisfied after the demand state
3.4. ANALYTICAL FRAMEWORK

θ is realised. The second period limited liability constraint is:

\[(LL2) \quad U(\theta) = R(x(\theta), \theta) - t(\theta) \geq 0, \quad \forall \theta \in \Theta\] (3.3)

To ensure that a retailer with a high realisation of demand does not take a contract designed for a low demand state retailer, self-selection or second period incentive compatibility constraints apply. By the Revelation Principle, attention is restricted to the class of direct mechanisms where the retailer is asked to make a truthful report on their demand type, \(\theta\). Using the notation that a retailer of demand type \(\theta\) reports their demand state as type \(\hat{\theta}\), then their second period utility is \(U(\theta, \hat{\theta})\). Incentive compatibility requires:

\[U(\theta, \theta) \geq U(\theta, \theta') \quad \forall \text{ pairs } (\theta, \theta') \in \Theta \times \Theta\]

That is, the contracts are constructed so that truthfully reporting is optimal for the retailer. Formally, locally incentive compatibility requires that the truthful report satisfies:

\[(IC2) \quad \theta \in \arg \max_{\hat{\theta} \in \Theta} \left\{ R(x(\hat{\theta}), \theta) - t(\hat{\theta}) \right\}, \quad \forall \theta \in \Theta\] (3.4)

Intuitively, this means that the manufacturer must give up an information rent equal to the marginal revenue that a retailer of type \(\theta\) could obtain by reporting its realised demand state to be \(\theta - d\theta\). Lemma 2 formalises this intuition:

Lemma 2 (Necessary & Sufficient Conditions for IC2) Necessary and sufficient conditions for incentive compatibility are: (i) \(U_\theta(\theta) = R_\theta(x(\theta), \theta)\) and (ii) \(x_\theta(\theta) > 0\).

Part (ii) of Lemma 2 simply implies that for truth-telling to be locally (and globally), optimal, the contract must specify that higher demand state retailers sell greater quantities of the manufacturer's good or service. The proof of the Lemma shows that part (ii) is a global condition.

A corollary of Lemma 2 that is immediately apparent is the existence of limited liability-induced information rents that the manufacturer has to give up to the retailer in order to prevent high demand states from being mis-reported. The rents arise due to the interaction of the ex-post limited liability constraints and the implementability
condition for separating second stage contracts. This is because limited liability implies that the manufacturer cannot penalise the retailer if a low demand state occurs. In fact, if the retailer faces the lowest demand state, then its ex post payoff must be non-negative: \( U(\theta) = 0 \). This skews the information rents that must be given up to higher demand types to ensure incentive compatibility.

**Corollary 1 (Demand State Information Rents)** The second period information rent accruing to a retailer that realises demand state \( \theta \) is:

\[
U(\theta) = \int_\theta^0 R_\theta(x(\theta), \theta) d\theta
\]  

The analogous information rents in the discrete demand state example were \( U_H = DFGC \) and \( U_L = 0 \) in Figure 3.3.11

Note that the expression for the retailer’s information rents depends cumulatively on the quantity profile for any given demand state. The manufacturer must inevitably give up the information rents if it wishes to achieve incentive compatibility, but by choosing an appropriate profile of quantities, the manufacturer can distort the distribution of information rents to its advantage. Recall that the manufacturer wishes to maximise its expected profit, conditional on an action that the retailer takes. Recall that the manufacturers expected profits are \( V = T + \int_\theta (t(\theta) - cb(\theta)) g_\theta(\theta) d\theta \). Hence, the manufacturer’s problem is to maximise its \( a \)-contingent expected profits subject to part (i) of Lemma 2 and (LL2):

\[
\begin{align*}
(P) \quad & \max_{\{x, u\}} V = \int_\theta (R(x(\theta), \theta) - U(\theta) - cb(\theta)) g_\theta(\theta) d\theta \\
\text{subject to:} & \quad (LL2) \ U(\theta) \geq 0 \quad \text{and} \quad (IC2) \ U_\theta(\theta) = R_\theta(x(\theta), \theta)
\end{align*}
\]  

In the optimal control formulation, \( U(\theta) \) is the state variable, \( x(\theta) \) is the control variable and \( \mu(\theta) \) is the costate variable. The transversality conditions are: (TV) \( U(\bar{\theta}) = 0 \) and \( \mu(\bar{\theta}) = 0 \). Using the individual rationality constraint to eliminate \( t(\cdot) \) from the manufacturer’s profit function, the Hamiltonian is:

\[
H = (R(x(\theta), \theta) - U(\theta) - cb(\theta)) g_\theta(\theta) + \mu(\theta) R_\theta(x(\theta), \theta)
\]  

\(^{11}\)Note that \( DFGC \) may also be interpreted as the marginal revenue in the discrete state as well.
Application of the Maximum Principle, as described in Leonard & Long (1992), yields the following first order conditions:

\[
\frac{\partial H}{\partial x} = 0 \iff (R_x - x^a)g_a(\theta) + \mu(\theta)R_\theta(x^a(\theta), \theta) = 0 \tag{3.8}
\]

\[
\mu_\theta(\theta) = \frac{\partial H}{\partial U^a} \iff \mu_\theta(\theta) = g_a(\theta) \tag{3.9}
\]

\[
U^a_\theta(\theta) = \frac{\partial H}{\partial \mu} = R_\theta(x^a(\theta), \theta) \tag{3.10}
\]

The optimally controlled solution to the manufacturer’s second stage problem is an a-contingent menu of contracts, or fee schedule \((t^a(\theta), x^a(\theta))_{\theta \in \Theta}\). The retailer’s rents when evaluated at the optimally controlled solution yields the realised a-contingent surplus:

\[
U^a(\theta) := U(\theta, t^a(\theta), x^a(\theta)).
\]

The precise nature of the fee schedule is given in the following Proposition:

**Proposition 1 (Fee Schedule)** The a-contingent contracts in the second period with incomplete information and limited liability, contingent on the retailer choosing action \(a \in A\) in the first period, are given by the fee schedule \((t^a(\theta), x^a(\theta))\) where:

\[
t^a(\theta) = R(x^a(\theta), \theta) - \int_\theta^\theta R_\theta(x^a(\bar{\theta}), \bar{\theta})d\bar{\theta}, \quad \text{and} \quad R_x - c = \frac{1 - G_a(\theta)}{g_a(\theta)} R_\theta x
\]

where \(x^a(\theta)\) is the implicit solution to the first-order condition on the right.

Inspection of the right hand condition shows the implicit solution to the optimal quantity profile. For the highest demand state, the manufacturer requires the retailer to sell into the market the (first-best) integrated monopoly quantity, where \(MR(x^a(\bar{\theta})) = c\). This is the largest profit that can be made in the output market. The fee that the retailer must pay to the manufacturer, \(t^a(\bar{\theta})\), is equal to the entire monopoly revenue that it makes from the output market, \(R(x^a(\bar{\theta}), \bar{\theta})\), less the information rent that the manufacturer must give up to it to elicit a truthful report of the demand state, \(U^a(\theta) = \int_\theta^\theta R_\theta(x^a(\bar{\theta}), \bar{\theta})d\bar{\theta}\). The analogous contract in the discrete demand case was \(\left\{DBC - DFGC, x_{IH}^D\right\}\).

Compare this to the lowest demand state. Since the manufacturer cares about leaving information rents to higher demand state retailers, it sacrifices allocative efficiency by
distorting the quantity it asks the lowest demand retailer (thereby inducing a form of the vertical externality). To see this note that for the lowest demand state $\theta$, the right hand side of the implicit quantity solution is the largest. As such, there is wedge between marginal revenue and marginal cost: $MR(x^a(\theta)) > c$. However, by sacrificing allocative efficiency on low demand states, the manufacturer is able to save on information rents that it must give up in the event of higher demand states occurring. In addition, since the manufacturer obtains no benefit to leaving information rents to the lowest demand state retailer, it extracts all the profit from the output market in that state of the world: $t^a(\theta) = R(x^a(\theta), \theta)$. The analogous contract in the discrete demand case was $\{CGHc, x^e_b\}$.

Part (ii) of Lemma 2 required that the control path for the quantity schedule is monotonically increasing for the variable component of the franchise contract to be implementable. This is shown to be true in Lemma 3:

**Lemma 3 (Implementability)** The optimal $a$-contingent quantity schedule $x^a(\theta)$ is monotone increasing in $\theta$.

The equilibrium franchise contract can now be defined as the triple $\{T^a, (t^a(\theta), x^a(\theta))_{\theta \in \Theta}\}$ for some $a \in A$. This section has just characterised the $a$-contingent fee schedule, $(t^a(\theta), x^a(\theta))_{\theta \in \Theta}$ as the solution to the manufacturer’s optimal control problem. The next section takes a step back in the timing to where the retailer has not yet realised the demand state. This will enable a characterisation of the $a$-contingent franchise fee component of the contract: $T^a$.

### 3.4.2 Stage 1: Franchise Fee Component

In the first period the manufacturer wishes to charge a franchise fee to extract all the expected surplus from the second period retailing. The expected second period surplus is non-zero since the manufacturer must abide by ex-post limited liability constraints on retailing. Hence, while it will give up information rents in the second period, it can extract a first period value equal to the retailer's expected value of engaging in the retailing activity. To simplify the expressions, the following definition is employed:

$$
E_{\theta,a} U^a(\theta) := \int_{\Theta} U^a(\theta^*) g_a(\theta^*) d\theta^* = \int_{\Theta} \int_{\mathbb{R}} R_\theta(x^a(\hat{\theta}), \tilde{\theta}) g_a(\theta^*) d\tilde{\theta} d\theta^*, \quad a, \hat{a} \in A
$$
This represents the retailer's expected rent from signing to an \( \hat{a} \)-contingent fee schedule if it chooses action \( a \in \mathcal{A} \). Hence, while the second period information rents are generated from the \( \hat{a} \)-contingent fee schedule that the manufacturer optimally controls for that action, the expectation over realisations of \( \theta \) is evaluated according to the distribution induced by the action \( a \in \mathcal{A} \) that the retailer takes. This reflects the scope for the retailer to act contrary to the designs of the manufacturer, the subject of the next section. Note that in the discrete demand case, the analogous expression was: \( E_{g,a}U = (1 - g(a))0 + g(a)DFGC \), where the size of \( DFGC \) depended on the action that the manufacturer believes the retailer to have taken. In that case, the manufacturer charged a fixed fee to extract this expected surplus and leave the retailer indifferent to entering the contract. The same task is undertaken here.

It will be convenient in what follows to define the efficient action \( a^* \) from the manufacturer's perspective:

\[
a^* := \{ a \in \mathcal{A} | V^a \geq V^{a'} \ \forall a' \in \mathcal{A} \}
\]

Where \( V^a \) is the manufacturer's surplus evaluated at the equilibrium solution. That is, \( a^* \) is the action that is most favoured by the manufacturer.

For the continuum of demand states case, first period payoffs to the retailer are:

\[
u^a := -T^a + E_{g,a}U^\hat{a}(\theta), \quad a, \hat{a} \in \mathcal{A}
\]

where \( T^a \) is the franchise fee charged by the manufacturer to the retailer. The manufacturer must respect a first period individual rationality constraint. To induce the retailer to sign the contract, the manufacturer must at least provide the retailer with their ex ante reservation value, which is normalised to zero when the retailer takes the efficient action. So individual rationality means that the retailer's expected value of entering the contract with the manufacturer when the contract induces the efficient action \( a^* \) must

\[12\]It is assumed that the set \( a^* \) is single-valued. That is, there is a unique efficient action.
be greater than zero:

\[(IR1) \quad u^* = -T^* + E_{\theta, a} U^*(\theta) \geq 0 \quad (3.11)\]

Before turning to examine the consequences of specific action spaces in the next section, first the case where no action can be taken by the retailer is examined as a benchmark.

**Empty Action Space: \( a = \phi = A_1 \)**

In this case, the retailer cannot take an action in the first period that influences the distribution of the demand states. Hence, both the manufacturer and retailer believe the probability distribution of demand states to be given by the unconditional cumulative distribution function \( G(\theta) \). The manufacturer then simply offers a contract that consists of a single franchise fee, and a fee schedule as in Proposition 1. The franchise contract in this case corresponds to a generalised two-part tariff case, where the franchise fee serves to extract expected surplus. The contract in this case must abide by \((IR1)\). Then the franchise fee is set to extract the expected surplus that accrues to the retailer from its operation in the final output market due to the ex post limited liability constraint.

**Proposition 2 (Empty Action Space)** The optimal franchise contract when the retailer cannot take an action prior to the realisation of the demand state is given by \( \{T^\phi, (t^\phi(\theta), x^\phi(\theta))\} \), where:

\[
T^\phi = E_{\theta, a} U^\phi(\theta), \quad t^\phi(\theta) = R(x^\phi(\theta), \theta) - \int_\theta R_\theta(x^\phi(\bar{\theta}), \bar{\theta}) d\bar{\theta}, \quad R_x - c = \frac{1 - G(\theta)}{g(\theta)} R_{dx}
\]

where \( x^\phi(\theta) \) is the implicit solution to the first order condition on the right.

The interpretation of the franchise contract is straightforward. The fee schedule \( (t^\phi(\theta), x^\phi(\theta)) \) still embodies the incentive compatibility and limited liability constraints, as in Proposition 1. However, note that when committing to choosing a second period contract from that menu, the retailer imputes its ex ante payoff to be the expectation over all the rents it could possibly receive, according to the unconditional probability distribution in this case. Since the \((LL2)\) constraint results in a positive expected payoff from the fee schedule, the manufacturer can charge a franchise fee \( T^\phi \) equal to the size of the retailers
expected payoff, and just induce the retailer to sign the ex ante contract. In the discrete demand formulation, the franchise fee was found to be $T = g.DFGC = E_g U$, which is exactly analogous to the franchise fee in Proposition 2 for the continuum of demand states.

Recall the second hypothesis from the discrete demand state illustration: "the more weight the manufacturer assigns to the high demand state occurring, the greater is the inefficiency from the vertical externality that the manufacturer imposes on the low demand type". Proposition 2 can be used to formalise this hypothesis by examining the impact of different actions on the franchise contracts. First, recall that favourable actions induce HRD. Hence, if the retailer takes an action $a$ that is favourable instead of an action $a'$, then the inverse hazard rates respect the inequality: $1 - \frac{G_a(\theta)}{s_a(\theta)} \leq 1 - \frac{G_{a'}(\theta)}{s_{a'}(\theta)}$. Using this and the concavity of the revenue function in $x$, $R_{xx} < 0$, in the right hand condition of Proposition 2 shows that the distortion in the quantity profile for the favourable distribution is more severe: $x^a(\theta) \leq x^{a'}(\theta)$. From Corollary 1, the information rent depends accumulatively on the marginal revenues in output. Hence, a direct consequence of the difference in the output distortions is that the variable fee $t(\theta)$ is pointwise higher for the favourable distribution because the manufacturer extracts greater rents for each type due to the relatively lower expected cost of allocative inefficiency.

Broadly speaking, the manufacturer cares more about extracting rents when it believes it is facing a favourable demand state distribution. However, when it believes it is facing an unfavourable demand state distribution, it cares more about allocative efficiency: it prefers not to penalise the retailer as much by exploiting the vertical externality for incentive compatibility.

Now all components of the franchise contract have been accounted for. Using this characterisation, the next step is to examine two different non-empty action spaces that could feasibly arise in franchise contracting arrangements. This is undertaken in the next section.
CHAPTER 3. FRANCHISE CONTRACTS

3.5 Action Spaces

This section analyses the retailer’s choice of action \( a \in \mathcal{A} \), when the action space consists of (i) a set of efforts: \( \mathcal{A}_2 = \{e_L, e_H\} \) that influence the conditional probability distributions over demand states, and (ii) a set of reports of signals: \( \mathcal{A}_3 = \{\sigma_L, \sigma_H\} \) that are perfectly correlated with the likelihood of the distribution of demand states. The first case corresponds to a situation of moral hazard followed by adverse selection. It describes a situation where the retailer could engage in, for example, a marketing campaign in its territory, or where the retailer’s managers could stake their reputation on the manufacturers product. In both cases, the level of effort exerted by the retailer is costly, and is not something that can be written into a contract: the reward through the contract must induce the efficient action. The second case corresponds to a situation of sequential adverse selection. It describes a situation where the retailer already possesses some private information about the likelihood of the demand state distribution before the contract is signed. For example, the retailer may have better knowledge of the composition of residential and business consumers for the manufacturers good or service in its territory. In this case the manufacturer may find it useful to offer two types of franchise contract to screen retailers that have favourable information.

3.5.1 Inducing Effort with Franchise Fee: \( a \in \{e_H, e_L\} = \mathcal{A}_2 \)

Here it is demonstrated how the manufacturer can devise a franchise contract that is guaranteed to elicit the efficient level of effort from the retailer when the first stage is characterised by moral hazard. Figure 3.5 illustrates the timing of the contract. The

\[
\begin{align*}
t = 0 & \quad t = 1 & \quad t = 2 & \quad t = 3 & \quad t = 4 & \quad t = 5 \\
\downarrow & \quad \downarrow & \quad \downarrow & \quad \downarrow & \quad \downarrow & \quad \downarrow \\
\text{M offers contract} & \quad \text{R accepts or refuses} & \quad \text{R pays } T^* & \quad \text{Demand state } \theta \text{ is } x^*(\theta) \& \text{are executed} & \quad \text{R sells transfers} & \quad \text{realised pays } t^*(\theta) \\
\{T^*, (t^*(\theta), x^*(\theta))\} & \quad a \in \{e_L, e_H\} & \quad & \quad & \quad \text{and } \theta \text{ is } x^*(\theta) \quad \text{are executed} & \quad (T^*, (t^*(\theta), x^*(\theta))) \end{align*}
\]

Figure 3.5: Timing - Stage 1 Moral Hazard
3.5. ACTION SPACES

action space for the retailer in this case consists of two efforts levels, \( a \in \{e_H, e_L\} \) with \( e_H > e_L \). The effort level \( e_H \) is considered more favourable than \( e_L \), since when the retailer exerts more effort, say in marketing activities, it induces a more favourable demand state distribution. However, effort is costly to the retailer, with respective costs equal to:

\[
\psi^a = \begin{cases} 
\psi^H = \psi & \text{if } e = e_H, \\
\psi^L = 0 & \text{if } e = e_L.
\end{cases}
\]

with \( \psi > 0 \). The manufacturer wishes to induce one of the actions, \( e_H \) or \( e_L \). Hence, it offers one optimal contract: \( \{T^*, (t^*(\theta), x^*(\theta))_{\theta \in \Theta}\} \), which can take on two values, corresponding to \( a^* = e_H \) and \( a^* = e_L \). Then in equilibrium, the retailer’s first-period utility from exerting effort level \( a \) is:

\[
u^a = -T + E_{\theta, a} U^a(\theta) - \psi^a, \quad a \in \{e_H, e_L\}
\]

Notice that by assumption, there is not a stochastic relationship between effort and the outcome of the effort in this setup. For instance, if the retailer exerts high effort, then the outcome is the favorable distribution over demand states. For incentive compatibility, the retailer’s first period optimal effort choice is the action \( a^* \in A \) that the manufacturer targets with franchise contract \( \{T^*, (t^*(\theta), x^*(\theta))_{\theta \in \Theta}\} \):

\[
a^* \in \text{argmax}_{a \in \{e_H, e_L\}} \{-T^* + E_{\theta, a} U^{a^*}(\theta) - \psi^a\}, \quad a^*, a \in \{e_H, e_L\}
\]

It is assumed that the exertion of effort is unobservable by the manufacturer and non-verifiable by a court of law. This means that the manufacturer cannot simply write a contract on the amount of effort. Even so, it need only write one franchise contract, and the retailer will be induced to exert the efficient amount of effort under the circumstances. In doing this, the manufacturer must therefore respect the individual rationality constraint on the action it wishes to induce and the incentive compatibility constraint from above.

From Proposition 1 and the subsequent discussion, for each demand state, the infor-
CHAPTER 3. FRANCHISE CONTRACTS

mation rent accruing to the retailer is smaller when it exerts effort since the manufacturer
distorts the quantity schedule away from first best more rapidly for lower demand states.
However, ex-ante more weight is placed on realising a higher demand state when effort
is exerted. As a consequence, there is ambiguity in the ex ante expected surplus for each
effort level: \( E_{\theta,H} U^H(\theta) \leq E_{\theta,L} U^L(\theta) \). Therefore, there are two cases that emerge in this
situation, as captured in Lemma 4 below:

**Lemma 4**  The optimal franchise fee elicits the high level of effort only when \( E_{\theta,H} U^H(\theta) - E_{\theta,L} U^L(\theta) \geq \psi \), otherwise the manufacturer optimally induces the retailer to exert low effort.

Without imposing more structure on the conditional probability distributions, no finer
prediction about the level of effort the retailer should be induced to exert can be made.
However, Lemma 4 does help in describing the two contracts that are feasible for the
manufacturer to offer. These are stated in the following Proposition.

**Proposition 3** (Moral Hazard Franchise Contract) The optimal franchise contract when
the retailer’s action space is \( A = \{e_H, e_L\} \) is given by the single contract:

\[
\{T^j, (t^j(\theta), x^j(\theta))\}, \quad \text{where: } j = \begin{cases} H & \text{if } E_{\theta,H} U^H(\theta) - E_{\theta,L} U^L(\theta) \geq \psi \\ L & \text{if } E_{\theta,H} U^H(\theta) - E_{\theta,L} U^L(\theta) < \psi \end{cases}
\]

and where the components of the franchise contract are determined by:

\[
T^j = E_{\theta,e_i} U^j(\theta) - \psi^j, \quad t^j(\theta) = R(x^j(\theta), \theta) - \int_0^\theta R_\theta(x^j(\theta), \theta) d\theta, \quad R_\theta - c = \frac{1-G_\theta(\theta)}{g_\theta(\theta)} R_{\theta|x}
\]

where \( x^j(\theta) \) is the implicit solution to the first order condition on the right.

A nice feature of the optimal moral hazard franchise contract is that there is only one
franchise fee: the manufacturer can elicit whichever level of effort it finds in its interests
with a single upfront fee and a commitment to sell its good or service from the corre­
sponding fee schedule. However, as will be shown in the next section, when the first stage
informational environment is best described by adverse selection, the manufacturer may
need to offer a menu of franchise contracts to ensure efficient screening of information.
3.5. ACTION SPACES

3.5.2 Screening by Franchise Fees: \( a \in \{\sigma_L, \sigma_H\} = \mathcal{A}_3 \)

In this section, the action space of the retailer is a report \( a \in \{\sigma_H, \sigma_L\} \), where \( \sigma_H > \sigma_L \) are privately observed signals that the retailer receives prior to accepting the manufacturer's contract. Figure 3.6 illustrates the timing of the contract. The manufacturer's beliefs about which signal the retailer received are: \( Pr(\sigma = \sigma_H) = \lambda \) and \( Pr(\sigma = \sigma_L) = 1 - \lambda \) where \( 0 \leq \lambda \leq 1 \). Since the signal is private information, the manufacturer designs the franchise contracts in a way that induces truthful revelation of the signal. This means that the manufacturer offers contracts that are contingent on the outcome of the retailer's action. Suppose that \( u(\sigma, \delta) \) represents a retailer of type \( \sigma \) making a report that it is type \( \delta \). Then first-period individual rationality conditions (\( IR_1 \)) are:

\[
u(\sigma_i, \sigma_i) \geq 0, \quad i, j = \{L, H\}\]

First-period incentive compatibility conditions (\( IC_1 \)) are:

\[
u(\sigma_i, \sigma_i) \geq u(\sigma_i, \sigma_j) \quad i, j = \{L, H\}, \quad i \neq j\]

If the manufacturer did not charge franchise fees, first-period high-signal types of retailers have an incentive to mimic low-signal types. This result is proved in Lemma 5:

**Lemma 5** A high-signal retailer has more to gain from misreporting its signal than does a low-signal retailer.
The intuition for Lemma 5 as to why high-signal retailers have more to gain from misreporting their signal follows from considering the rent extraction efficiency trade-off under the manufacturer’s belief about the distribution of demand states the retailer is facing. When the manufacturer believes the retailer is facing the unfavorable distribution of demand states, it places more weight on allocative efficiency, since it believes the probability of the retailer realising a high demand state is relatively low. Hence, it does not distort the quantity profile as much as when the distribution of demand states is favorable. This means that a retailer in any given demand state deciding whether to mimic a lower demand state retailer could sell more into the market if the quantity profile is not as distorted, and pay a smaller fee on that contract. Hence, high-signal type retailers will earn an expected rent premium on their first-period information.

In view of Lemma 5, the first-period low-signal retailer’s (IR1) constraint is binding, and the first-period high-signal retailer’s (IC1) constraint is binding. The following Lemma then characterises the incentive feasible franchise fees:

**Lemma 6** The incentive-feasible franchise fees for the low and high-signal retailers are:

\[
\begin{align*}
T_L &= E_{\theta,L}U(\sigma_L) \\
T_H &= E_{\theta,H}U(\sigma_H) - E_{\theta}DU(\sigma_L)
\end{align*}
\]

with \( T_H \leq T_L \) (3.15)

where \( E_{\theta}DU(\sigma_L) := E_{\theta,H}U(\sigma_H, \sigma_L) - E_{\theta,L}U(\sigma_L, \sigma_L) \) is the first-period information rent.

The expression \( E_{\theta}DU(\sigma_L) := E_{\theta,H}U(\sigma_H, \sigma_L) - E_{\theta,L}U(\sigma_L, \sigma_L) \) is the expression for the expected rent premium. The first term is the expected payoff to a high-signal retailer from selling the low-signal contract quantities weighted with the favorable probability distribution. The second term is simply the franchise fee, \( T_L \), payable by the low-signal retailer. Hence, the expected rent premium is just the excess over the low-signal franchise fee that the high-signal retailer could earn for itself by mimicking the low-signal retailer.

Note that there is a priori no reason why \( T_H \) cannot be negative: in that case, the manufacturer subsidises the high-signal retailer to sign a contract, and recoups the loss with a higher variable fee. The optimal adverse selection franchise contracts are stated in Proposition 4:
Proposition 4 (Adverse Selection Franchise Contracts) The optimal menu of franchise contracts when the retailer's action space is \( A = \{\sigma_H, \sigma_L\} \) is given by:

\[
\{T^H, (t^H(\theta), x^H(\theta))\}, \{T^L, (t^L(\theta), x^L(\theta))\}
\]

where the components of the franchise contract are determined by Lemma 6 and:

\[
t^j(\theta) = R(x^j(\theta), \theta) - \int_0^\theta R_\theta(x^j(\theta), \theta) d\theta, \quad \text{and} \quad R_x - c = \frac{1-G_j(\theta)}{g_j(\theta)} R_{\sigma_x}
\]

for \( j = L, H \) and where \( x^j(\theta) \) is the implicit solution to the first order condition on the right.

From the manufacturer's perspective, screening signal types may not always be attractive. The price of limiting information rents under a favorable demand state distribution is a relatively severe distortion in allocative efficiency. So committing to a fee schedule that embodies the severe rent extraction efficiency trade-off may only be worthwhile if the probability that a retailer receives a high-signal is great enough. Otherwise the manufacturer may prefer to remain ignorant of the retailer's first period information. This idea is explored in the next section.

Screening or Ignorance?

Lemma 6 shows that separating franchise fees exist, and will allow for franchise contracts that are characterised by first stage adverse selection to be written, as in Proposition 4. However, if the manufacturer's beliefs about the signal received by the retailer place a large weight on the low signal, then separating the types may not be optimal. To see this, remember that when beliefs about the distribution of the demand state are conditional on the low signal being received, the manufacturer considers achieving allocative efficiency to be relatively more important than rent extraction.

However, if the manufacturer did not revise its beliefs, it would prefer to distort the quantity profile more steeply than it otherwise would if it had received the low signal: it would be more concerned with giving up information rents. Hence, using separating franchise contracts when the probability of receiving a high-demand signal is small may
CHAPTER 3. FRANCHISE CONTRACTS

leave the manufacturer with a smaller expected surplus than if they did not try to separate out the signal types.

It is possible to establish the critical probability $\lambda^*$ where the manufacturer is indifferent between offering separating franchise contracts and being ignorant of the retailer’s first stage signal. To make the exposition simpler, assume without further loss of generality that the manufacturer’s constant marginal production cost, $c$, is zero. Then the risk neutral manufacturer’s ex ante profit is:

$$E_{\theta,k}V = T_k + \int_{\Theta} t(\theta_k)g_k(\theta)d\theta, \quad k = \phi, L, H$$

Making the substitutions for $k = \phi, L, H$ yields:

$$E_{\theta,\phi}V^\phi = \int_{\Theta} R(x^\phi(\theta))g_\phi(\theta)d\theta = E_{\theta,\phi}R^\phi$$
$$E_{\theta,L}V^L = \int_{\Theta} R(x^L(\theta))g_L(\theta)d\theta = E_{\theta,L}R^L$$
$$E_{\theta,H}V^H = -\Delta E_{\theta,L}(\theta) + \int_{\Theta} R(x^H(\theta))g_H(\theta)d\theta = -\Delta E_{\theta,L}(\theta) + E_{\theta,H}R^H$$

Then total manufacturer payoffs are:

$$E_{\theta,H}V^H + T^H = E_{\theta,H}R^H - E_{\theta,L}U^L(\theta), \quad E_{\theta,L}V^L + T^L = E_{\theta,L}R^L, \quad E_{\theta,\phi}V^\phi + T^\phi = E_{\theta,\phi}R^\phi$$

A separation of signal types is only worthwhile for the manufacturer if the expected payoff from the high-signal retailer is greater than from the low-signal retailer, which requires: $E_{\theta,H}R^H - E_{\theta,L}R^L \geq E_{\theta,L}U^L(\theta)$. Moreover, the manufacturer prefers offering a screening franchise contract to remaining ignorant whenever: $\lambda E_{\theta,H}R^H + (1 - \lambda)E_{\theta,L}R^L - \lambda E_{\theta,L}U^L(\theta) - E_{\theta,\phi}V^\phi \geq 0$. These two inequalities taken together imply a lower bound on $\lambda$ that is required for existence of screening franchise contracts:

$$\lambda \geq \frac{E_{\theta,L}R^L - E_{\theta,L}R^L}{E_{\theta,H}R^H - E_{\theta,L}R^L - E_{\theta,L}U^L(\theta)} \equiv \lambda^*$$

From the definition of $\lambda^*$, existence of screening contracts requires the following inequalities to be satisfied: $E_{\theta,H}R^H \geq E_{\theta,\phi}R^\phi \geq E_{\theta,L}R^L$, and $E_{\theta,H}R^H - E_{\theta,L}R^\phi > E_{\theta,L}U^L(\theta)$. In
3.6. CONCLUSION

this case, $\lambda^* \in (0, 1)$. The single “ignorant” franchise contract is exactly as in Proposition 2, where the retailer cannot take an action. The following Proposition gathers the results on screening franchise contracts.

**Proposition 5** If $\lambda \in [\lambda^*, 1]$, the manufacturer offers the menu of screening franchise contracts in Proposition 4. If $\lambda \in [0, \lambda^*)$ then the manufacturer offers the pooling franchise contract in Proposition 2.

There is no guarantee of the existence of $\lambda^* \in (0, 1)$. Depending on the structure imposed on the parameters of the model, it may be that the manufacturer always prefers to screen, or always prefers to remain ignorant.

3.6 Conclusion

This chapter has demonstrated how a manufacturer can use franchise contracts in a vertical relationship with a retailer when the retailer has ex ante private information, and the manufacturer’s contracts are bound by the retailer’s limited liability. It was shown how the interaction of second stage incentive compatibility constraints with limited liability constraints yield an expected surplus to the retailer, and how the manufacturer may employ a franchise fee to appropriate it. This franchise fee, combined with the fee schedule was defined as a franchise contract.

The franchise fee can be used as an instrument for the manufacturer to either elicit an efficient action to influence the distribution of demand states, or screen ex ante private information from the retailer, according to the type of ex ante information that the retailer possessed. In the context of first stage moral hazard, it was shown that the manufacturer can always offer a single franchise contract that elicits the most efficient level of effort in the first stage and induces a truthful demand state report in the second stage. In the context of first stage adverse selection, it was shown that a pair of screening franchise contracts performs better than when the manufacturer is ignorant of the retailer’s first stage information only where the manufacturer’s beliefs about the probability that the retailer has received the high-demand signal is larger than some critical threshold, $\lambda^*$.

The model in this chapter is static: it does not account for situations in which the manufacturer and retailer contract repeatedly. In such situations, the retailer’s concern
for its reputation may solve the moral hazard issue. Furthermore, if the demand state that the retailer faces is correlated over time, then the simple franchise contract outlined in this chapter may no longer allow the manufacturer to screen privately informed retailers. Issues of renegotiation and commitment arise, so that the mechanics of dynamic contracting must be employed.\textsuperscript{13} These matters are left for future research.

The conclusion that the manufacturer may prefer to remain ignorant rather than screen retailers with different signals about the likelihood of the demand state distribution raises an issue for further research. As there is expected surplus at stake, it may be possible to construct a game where the strategies of the manufacturer are to attempt to "muddy the waters" or signal or screen information. The mechanics of the timing of the model and the use of limited liability constraints could easily be applied to other incentives theory contexts, for example, in non-linear pricing or in credit rationing contracts.

3.7 Appendix

3.7.1 Proof of Lemma 1

First note that: \( \frac{d}{d\theta} \ln(1 - G(\theta)) = -\frac{g(\theta)}{1 - G(\theta)} \).

Integrating this and rearranging for \( G(\theta) \) yields: \( G(\theta) = 1 - e^{-\int_{\theta}^{\bar{\theta}} \frac{g(\theta)}{1 - G(\theta)} d\theta} \).

By hypothesis, the HRD condition holds point-wise, so integrating both sides of the inequality results in:

\[
\int_{\theta}^{\bar{\theta}} \frac{g(\theta)}{1 - G(\theta)} d\theta \leq \int_{\theta}^{\bar{\theta}} \frac{g(\theta)}{1 - G(\theta)} d\theta \Rightarrow 1 - e^{-\int_{\theta}^{\bar{\theta}} \frac{g(\theta)}{1 - G(\theta)} d\theta} \leq 1 - e^{-\int_{\theta}^{\bar{\theta}} \frac{g(\theta)}{1 - G(\theta)} d\theta} \Rightarrow G(\theta) \leq G(\theta)
\]

which holds \( \forall \theta \in \Theta \). \( \blacksquare \)

3.7.2 Proof of Lemma 2

For part (i), differentiating \( R(x(\hat{\theta}), \hat{\theta}) - t(\hat{\theta}) \) with respect to the retailer's report \( \hat{\theta} \) yields:

\[
R_{x}(x(\hat{\theta}), \hat{\theta})x_{\theta} - t_{\hat{\theta}}(\hat{\theta}) = 0 \tag{3.16}
\]

\textsuperscript{13}See for example Hart and Tirole (1988) and Laffont and Tirole (1988).
For truth-telling to be optimal, (3.16) becomes the identity \( R_x(x(\theta), \theta)x_\theta - t_\theta(\theta) = 0, \forall \theta \in \Theta \). Now differentiating the argument of (3.4) with respect to \( \theta \) and evaluating at \( \hat{\theta} = \theta \) yields:

\[
U_\theta = R_x x_\theta + R_\theta - t_\theta
\]

Using the first order condition for truth-telling in this expression yields the result.

For part (ii) of Proposition 3.4, the second derivative of (3.16) again with respect to the retailers report \( \hat{\theta} \) must be less than or equal to zero:

\[
R_{xx}(x(\hat{\theta}), \theta)x_\theta^2 + R_x(x(\hat{\theta}), \theta)x_{\theta\theta} - t_{\theta}(\hat{\theta}) \leq 0
\]

(3.17)

Now since the truth-telling identity is always zero, differentiating again with respect to \( \theta \) yields: \( R_{xx}(x(\theta), \theta)x_\theta^2 + R_x(x(\theta), \theta)x_{\theta\theta} + R_{x\theta}(x(\theta), \theta) - t_{\theta}(\hat{\theta}) = 0 \). Plugging this into (3.17) yields:

\[-R_{x\theta}(x(\theta), \theta)x_\theta \leq 0\]

Recalling the assumption that \( R_{x\theta} > 0 \) yields the local incentive result.

For global incentive compatibility, note that for any pair \( \theta, \theta' \in \Theta \), with \( \theta > \theta' \), that:

\[
R(x(\theta), \theta) - t(\theta) \geq R(x(\theta'), \theta) - t(\theta') \quad \& \quad R(x(\theta'), \theta') - t(\theta') \geq R(x(\theta), \theta') - t(\theta)
\]

Adding these together and rearranging yields:

\[
R(x(\theta), \theta) - R(x(\theta'), \theta) \geq R(x(\theta), \theta') - R(x(\theta'), \theta')
\]

Rewriting yields:

\[
\int_{x(\theta')}^{x(\theta)} R_x(z, \theta)dz \geq \int_{x(\theta')}^{x(\theta')} R_x(z, \theta')dz
\]

Then finally:

\[
\int_\theta^{\theta'} \int_{x(\theta')}^{x(\theta)} R_{x\theta}(z, y)dzdy \geq 0
\]

Since \( \theta > \theta' \) and \( R_{x\theta} > 0 \) by assumption, then this inequality holds only if \( x(\theta) > x(\theta') \).
3.7.3 Proof of Corollary 1

First note that Part (i) of Lemma 2 is a continuum of constraints, given by the solution to the differential equation: \( U^a(\theta) = R_\theta(x(\theta), \theta) \). A particular solution to this differential equation requires an initial condition. Since \( R_\theta(.) > 0 \) by assumption, then the utility profile is always upward sloping, hence the (LL) constraint will only bind at the lowest type in the interval, so \( U(\theta) = 0 \). Integrating Lemma 2(i) and using the initial condition \( U(\theta) = 0 \) yields:

\[
U(\theta) - U(\theta) = U(\theta) = \int_\theta^\theta R_\theta(x(\theta), \hat{\theta}) d\hat{\theta}
\]

which is the required result. ■

3.7.4 Proof of Proposition 1

First note that due to the concavity of the manufacturer's problem and Lemma 2, these necessary conditions are also sufficient. Equation (3.9) and the TV, \( \mu(\bar{\theta}) = 0 \), defines a differential IVP whose solution is:

\[
\mu(\bar{\theta}) - \mu(\theta) = \int_\theta^{\bar{\theta}} g_a(\bar{\theta}) d\bar{\theta} = (1 - G_a(\theta)) \Rightarrow \mu(\theta) = -(1 - G_a(\theta)) \quad (3.18)
\]

Putting this expression into equation (3.8) yields:

\[
(R_x - c)g_a(\theta) = \frac{1 - G_a(\theta)}{g_a(\theta)} R_{\theta x}(x^a(\theta), \theta)
\]

which is the left hand side equation in the Proposition. For the right hand side equation, note that for each \( \theta \in \Theta \) the expression in (3.3) must hold. Rearranging and replacing the surplus \( U^a(\theta) := U(\theta, t^a(\theta), x^a(\theta)) \) with the corresponding information rent from Corollary 1 yields the result. ■
3.7. APPENDIX

3.7.5 Proof of Lemma 3

Totally differentiating equation (3.8) yields: 
\[ R_{\theta \bar{z}} + R_{xx}x_{\theta}^z = \frac{d}{d\theta}(\frac{1-G_\theta(\theta)}{g_\theta(\theta)}) + \frac{1-G_\theta(\theta)}{g_\theta(\theta)}(R_{\theta \theta \bar{z}} + R_{\theta \theta \bar{x} \bar{x}}). \]

Solving this expression for \( x_{\theta}^z \) gives:

\[ x_{\theta}^z = \frac{\left( \frac{d}{d\theta}(\frac{1-G_\theta(\theta)}{g_\theta(\theta)}) - 1 \right) R_{\theta \bar{z}} + \frac{1-G_\theta(\theta)}{g_\theta(\theta)}R_{\theta \theta \bar{z}}}{R_{xx} - \frac{1-G_\theta(\theta)}{g_\theta(\theta)}R_{\theta \theta \bar{x} \bar{x}}} \]  

(3.19)

Since the HR is monotone increasing, the numerator of this expression is monotone decreasing. By assumption, \( R_{\theta \bar{z}} > 0, R_{xx} < 0, R_{\theta \theta \bar{z}} \leq 0 \) and \( R_{\theta \theta \bar{x} \bar{x}} \geq 0 \). Then it follows that \( x_{\theta}^z(\theta) \) is monotone increasing in \( \theta \). Note that under complete information, \( x_{\theta}^z = -\frac{R_{\theta \theta \bar{x} \bar{x}}}{R_{xx}} > 0 \), and that due to incomplete information, \( x_{\theta}^z < x_{\theta}^z^* \).

3.7.6 Proof of Lemma 4

The proof simply requires checking the incentive constraint for effort exertion for the retailer. When the manufacturer sets the franchise fee to \( T \), the payoffs under each level of effort are:

\[ u^H = -T + E_{\theta,H}U^H - \psi \]
\[ u^L = -T + E_{\theta,L}U^L \]

So first, if \( E_{\theta,H}U^H < E_{\theta,L}U^L \) then for the franchise fee \( u^H < u^L \). The manufacturer chooses the fee to satisfy the IR1 constraint on the low-effort level, \( T = E_{\theta,L}U^L(\theta) \). In this case, the retailer exerts no effort.

If \( E_{\theta,H}U^H > E_{\theta,L}U^L \) then there are two cases. Case (i): \( 0 \leq E_{\theta,H}U^H(\theta) - E_{\theta,L}U^L(\theta) \leq \psi \). In this case the retailer has no incentive to exert effort and the best the manufacturer can do is to set the franchise fee to extract the expected rent at the low-effort level: \( T = E_{\theta,L}U^L(\theta) \). Case (ii): \( E_{\theta,H}U^H(\theta) - E_{\theta,L}U^L(\theta) \geq \psi \). Hence, \( u^H \geq u^L \). In this case the manufacturer sets the franchise fee so that IR1 binds for the high-effort level: \( T = E_{\theta,H}U(\theta) - \psi \) and the retailer always exerts effort.
3.7.7 Proof of Lemma 5

The expected gain to the high-signal type of retailer from reporting a low-signal is the gain from having the manufacturer believe that its consumption valuations in period 2 comes from the distribution \( G_L(\theta) \) rather than \( G_H(\theta) \):

\[
\Delta_H U := \mathbb{E}_{\sigma_H, \hat{\sigma}_L}(\sigma_H, \hat{\sigma}_L) - \mathbb{E}_{\sigma_H, \hat{\sigma}_H}(\sigma_H, \hat{\sigma}_H) = \int_{\Theta} U_L(\theta) g_H(\theta) d\theta - \int_{\Theta} U_H(\theta) g_H(\theta) d\theta
\]

Similarly, the gain to the low-signal retailer from reporting a low-signal is the gain from having the manufacturer believe that their consumption value in period 2 (correctly) comes from the distribution \( G_L(\theta) \) rather than \( G_H(\theta) \):

\[
\Delta_L U := \mathbb{E}_{\sigma_L, \hat{\sigma}_L}(\sigma_L, \hat{\sigma}_L) - \mathbb{E}_{\sigma_L, \hat{\sigma}_H}(\sigma_L, \hat{\sigma}_H) = \int_{\Theta} U_L(\theta) g_L(\theta) d\theta - \int_{\Theta} U_H(\theta) g_L(\theta) d\theta
\]

Then the difference in the expected difference in surplus is:

\[
\Delta_H U - \Delta_L U = \int_{\Theta} (U_L(\theta) - U_H(\theta))(g_H(\theta) - g_L(\theta)) d\theta
\]

Integration by parts gives:

\[
\Delta_H U - \Delta_L U = (U_L(\theta) - U_H(\theta))(G_H(\theta) - G_L(\theta))|_\Theta - \int_{\Theta} (\frac{\partial U_L(\theta)}{\partial \theta} - \frac{\partial U_H(\theta)}{\partial \theta})(G_H(\theta) - G_L(\theta)) d\theta
\]

Using Leibniz' rule that \( \frac{\partial U_k(\theta^*)}{\partial \theta^*} = R_\theta(x_k(\theta^*)) > 0 \) for \( k = L, H \):

\[
\int_{\Theta} (R_\theta(x_L(\theta)) - R_\theta(x_H(\theta)))(G_L(\theta) - G_H(\theta)) d\theta > 0
\]

where the last inequality follows from (i) FOSD and (ii) \( x^L(\theta) \geq x^H(\theta) \) pointwise for \( \theta \in \Theta \).
3.7.8 Proof of Lemma 6

The low-signal retailer franchise fee follows directly since (IR1) binds for the low-signal retailer. For the high-signal retailer franchise fee, since (IC1) binds:

\[
\begin{align*}
  u(\sigma_H, \sigma_H) &= -T^L + \mathbb{E}_{\theta, L} U(\sigma_L, \sigma_L) - \mathbb{E}_{\theta, L} U(\sigma_L, \sigma_L) + \mathbb{E}_{\theta, H} U(\sigma_H, \sigma_L) \\
  &= u(\sigma_L, \sigma_L) + \mathbb{E}_{\theta} \Delta U(\sigma_L)
\end{align*}
\]

The high-signal retailer franchise fee can be derived using the definition of \( u(\sigma_H, \sigma_H) \) and again, since (IR1) binds for the low-signal retailer.

To show that \( T^H \leq T^L \), first write out \( T^H \) in full:

\[
T^H = \int_{\Theta} U_H(\theta^*) g_H(\theta^*) d\theta^* - \left( \int_{\Theta} U_L(\theta^*) g_H(\theta^*) d\theta^* - \int_{\Theta} U_L(\theta^*) g_L(\theta^*) d\theta^* \right)
\]

where the last two terms comprise \( \mathbb{E}_\theta \Delta U(\sigma_L) \), which itself can be more usefully expressed using integrations by parts:

\[
\begin{align*}
  \mathbb{E}_\theta \Delta U(\sigma_L) &= \int_{\Theta} U_L(\theta^*) [g_H(\theta^*) - g_L(\theta^*)] d\theta^* \\
  &= U_L(\theta^*) [G_H(\theta^*) - G_L(\theta^*)] \bigg|_{\Theta} - \int_{\Theta} \left( \frac{\partial U_L(\theta^*)}{\partial \theta^*} \right) [G_H(\theta^*) - G_L(\theta^*)] d\theta^*
\end{align*}
\]

Using Leibniz' rule: \( \frac{\partial U_L(\theta^*)}{\partial \theta^*} = R_\theta (x_L(\theta^*)) > 0 \). Hence, by the property of FOSD:

\[
\mathbb{E}_\theta \Delta U(\sigma_L) = \int_{\Theta} R_\theta (x_L(\theta^*)) [G_L(\theta^*) - G_H(\theta^*)] d\theta^* > 0 \quad (3.20)
\]

Returning to the expression for \( T^H \) and focusing on the first term in the expression, using integration by parts yields:

\[
\begin{align*}
  \int_{\Theta} U_H(\theta^*) g_H(\theta^*) d\theta^* &= U_H(\theta^*) G_H(\theta^*) \bigg|_{\Theta} - \int_{\Theta} \left( \frac{\partial U_H(\theta^*)}{\partial \theta^*} \right) G_H(\theta^*) d\theta^* \\
  &= U_H(\tilde{\theta}) - \int_{\Theta} R_\theta (x_H(\theta^*)) G_H(\theta^*) d\theta^*
\end{align*}
\]

Recalling that \( U_H(\tilde{\theta}) = \int_{\Theta} R_\theta (x_H(\theta)) d\theta \) and combining equations (3.20) and (3.21) into
CHAPTER 3. FRANCHISE CONTRACTS

\[ T^H \text{ gives:} \]

\[ T^H = \int_\Theta R_\Theta(x_H(\theta))d\theta - \int_\Theta R_\Theta(x_L(\theta))G_L(\theta)d\theta - \int_\Theta [R_\Theta(x_H(\theta)) - R_\Theta(x_L(\theta))]G_H(\theta)d\theta \]

Now, recall that:

\[ T^L = U(\sigma_L) = \int_\Theta U(\sigma_L)G_L(\theta)d\theta = U(\sigma_L) - \int_\Theta R_\Theta(x_L(\theta))G_L(\theta)d\theta \]

where integration by parts has been used again. Substituting and rearranging yields:

\[ T^H = \int_\Theta R_\Theta(x_H(\theta))d\theta + T^L - U(\sigma_L) - \int_\Theta (R_\Theta(x_H(\theta)) - R_\Theta(x_L(\theta)))G_H(\theta)d\theta \]

Rearranging yields:

\[ T^H - T^L = \int_\Theta R_\Theta(x_H(\theta))d\theta - \int_\Theta R_\Theta(x_L(\theta))d\theta - \int_\Theta (R_\Theta(x_H(\theta)) - R_\Theta(x_L(\theta)))G_H(\theta)d\theta \]

\[ = \int_{\Theta_2} (R_\Theta(x_H(\theta)) - R_\Theta(x_L(\theta))) (1 - G_H(\theta))d\theta \]

Rearranging and noting from Proposition 1 that \( R_\Theta(x_L(\theta)) - R_\Theta(x_H(\theta)) > 0 \) for all \( \theta \), gives the required result:

\[ T^H = T^L - \int_\Theta (R_\Theta(x_L(\theta)) - R_\Theta(x_H(\theta))) (1 - G_H(\theta))d\theta \]

From Proposition 1(i), \( x_L(\theta) \geq x_H(\theta) \). Hence, \( T^H \leq T^L \).
Chapter 4

Innovation Contracts with Leakage Through Licensing

4.1 Introduction

Innovation and technology are matters of ongoing importance in developed countries. There has been a great deal of research on incentives to undertake research and development, protection of the resultant intellectual property rights (IPR) by patenting or trade secrets, and licensing or knowledge sharing of innovations.\(^1\) However, there has recently been greater focus on the way that innovators are compensated, since there is an increasing reliance of R&D intensive industries on outside research.\(^2\) Issues of bargaining strength, ownership of IPR and the propensity for innovators to leak knowledge to parties outside a contract are at the focus of recent work. This chapter contributes by examining the impact of the ownership of the IPR to an innovation on the incentive contract between developers and researchers.

Innovation creates knowledge, which has the peculiar nature of being durable, non-rivalrous in consumption and partially excludable. Durability means that once a party obtains knowledge, it does not wear out or depreciate making it difficult to commit to not

---


CHAPTER 4. INNOVATION CONTRACTS

using wherever it has value. Non-rivalrous in consumption means that different parties may use the knowledge simultaneously. Partial excludability implies that appropriation of the benefits of an innovation is problematic since the strength of property rights may vary considerably. These properties ensure that IPR plays a central role in the contracts written between a developer and researcher, since innovation knowledge can be leaked to parties outside of the contract.

An innovation contract in this chapter involves a compensation wage, and an assignment of ownership of the IPR that results. Despite the capital intensity of most research programs, wages paid to researchers represent the largest component of R&D expenditures, averaging around 50% of total R&D expenditures in OECD countries (Public Support for Science and Innovation, PC Research Report 2007, pg. 588). Moreover, there is ample evidence of incentive contracting between large research intensive firms and innovators. However, there is little theoretical attention on the wage contracting arrangements in innovation markets. The second part of an innovation contract, ownership, is crucial because it bestows the right to license new technologies, an important source of revenue for the holders of the IPR. One estimate of the market for innovation licensing in the software industry puts the figure at $US100 billion in 2003 (see Bhattacharya et al (2006)).

For a developer, the process of obtaining innovation knowledge is plagued by at least two agency problems. The first is a moral hazard problem of unobservable and non-verifiable effort. The second is a leakage problem that results because innovation knowledge is durable, partially excludable and non-rivalrous. In this chapter, the first agency problem is tackled with the usual incentive contract approach. The second agency problem is avoided by assuming perfectly enforceable property rights, and that ownership rights can be assigned in a contract before the innovation is created.

The model consists of a Development Unit (DU, for example, a pharmaceutical company, or a Silicon Valley firm) who enters into a contract with a Research Unit (RU,  

\footnote{Lerner and Wulf (2006) empirically investigate the link between compensation of research personnel to the objectives of large US corporations.}  

\footnote{The alternative would be to model the imperfect protection that patents infer as in Lemley and Shapiro (2005), who examine the implication of uncertain property rights under patenting, or "probabilistic patents".}
for example, a biotechnology laboratory, or a software designer) for the creation of an innovation. The DU is assumed not to have expertise in the RU’s field, or the ongoing need for its skills to justify integration as an ownership configuration. Alternatively, the RU owns a bundle of complementary skills or assets that warrant a vertically separated structure.\(^5\)

While each party may hold IPR for existing innovations, at the time of contracting the exact specification of the innovation to be created in the contract is unknown. As such, the RU has no bargaining power, since there are many RUs that could perform the research task. Moreover, the DU provides the framework for the innovation to be generated in: without this framework, the RU could not independently create the innovation. The question of this chapter is then: under what conditions would the DU prefer to retain control over the IPR and residual control rights to licensing?

One reason for asking this question is concerned with the power of the incentive contract that the DU can offer. Since it retains all the bargaining power, any trade gains can be transferred to it through ex post lump sums. However, more total producer surplus could be generated if the RU is given high powered incentives to work. Such incentives could come through entitlement to the stream of revenue from licensing the innovation to other DUs in the industry.

It is the uncertain nature of R&D and the inability of the DU to observe and verify the RU’s action, or research effort, that necessitates the use of incentive contracting. In this respect, the model is not different from many other Principal-Agent relationships. Usually the interaction between the principal and the agent results in surplus that can be divided in some way between the two, according to a predetermined rule or contract. For example, in Holmstrom’s (1979) example of a machine repairman whose unobservable action influences the expected time before the machine breaks down, the surplus generated is the expected net value of the production when the machine is put in use. The demand for the repairman’s services is a derived demand extending from the principal alone. Unlike Holmstrom’s example however, the agent’s effort in this model confers two benefits: an internal and an external demand for the RU’s services. The internal demand

CHAPTER 4. INNOVATION CONTRACTS

stems from a direct payoff to the Principal from utilising the resulting innovation. This is analogous to the result of the efforts of the repairman. Note that the incentives for the repairman depend only on the *explicit* incentives he faces as controlled by the Principal. Where the present model departs from the standard framework is in the incorporation of a second source of benefit from the RU's services: an external demand for the RU's innovation efforts from the DU's competitors. This provides an *implicit* incentive that is out of the control of the DU. Now the RU's efforts produce a new technology that has value not just to the DU, but also other actors in the DU's output market. So to completely describe the contracting environment, the assignment of the residual control rights, or the ownership of the IPR, will need to be considered. If the DU delegates IPR to the RU, then the RU faces both an *explicit* incentive through the incentive contract from the DU, and an *implicit* incentive through the private return it can receive through licensing of the new technology to consumers other than the DU.

Hence, for the DU, retaining control of the IPR may be inefficient in terms of eliciting effort from the RU. On the other hand, turning control of the IPR over to the RU results in innovations that could be licensed to its own competitor's, thus reducing its profitability in the output market. The results of this chapter show that firstly, whenever an innovation would be profitable for the DU itself to license, ownership of the IPR is irrelevant. The reduction in the DU's expected profit from assigning IPR to the RU is offset by an exact reduction in its expected payment to the RU. Second, whenever the marginal licensing revenue of the innovation is greater than the reduction in the value of the innovation to the DU from allocating IPR to the RU, the DU prefers to retain the IPR for the innovation; in which case it will choose not to license.

The next section of the chapter briefly reviews the related literature. Section 3 sets up the incentive contracting model and derives the optimal wage contract. Section 4 presents the main ownership equivalence result when the DU would leak the innovation through licensing itself, and section 5 demonstrates the conditions under which ownership matters. Section 6 provides an example of the results in the context of a product differentiation model, and the last section concludes.
4.2 Relation to the Literature

Much of the vast literature on innovation contracting and management is dedicated to assessing the impact of the legal strength and ownership of IPR on incentives under different informational environments. Broadly speaking, the innovation literature that deals with incentive theoretic issues can be divided into two streams: one where the informed party takes the initiative, and one where the uninformed party takes the initiative.6

The informed party takes the initiative in the signalling models of Bhattacharya and Ritter (1983), Gallini and Wright (1990), Anton and Yao (2002, 2004) and its extension Bhattacharya and Guriev (2006). The form that the signal can take differs among these papers. For example, the innovator’s decision to publicly disclose knowledge through patenting in Bhattacharya and Ritter signals the economic value of the innovation to an external financier, while at the same time eroding its advantage over its competitors through leakage and imitation. In fact, the decision to disclose knowledge through a formal patent system or rely on common law and trade secrets depends on the legal IPR strength that patenting confers to the innovator. This was studied in Anton and Yao (2002, 2004) and Bhattacharya and Guriev (2006). In Gallini and Wright, the structure of the licensing contract under asymmetric information serves as a signal of the licensor’s pre-contractual information about the value of the innovation, a feature that is also examined in Martimort, Poudou and Sand-Zantman (2010). It is the possibility of imitation, or the strength of IPR that generate the results in all these papers.

The uninformed party takes the initiative through use of incentive contracts in the models of Bhattacharya, Glazer and Sappington (1992), Veraevel and Vencatchellum (2009), Lai, Riezman and Wang (2009) and Martimort, Poudou and Sand-Zantman (2010). The latter of these papers is a model of double sided asymmetric information. On one side, the innovator has private knowledge of the value of its innovation which it attempts to signal through contract form to a developer who must be incentivised to exert efficient effort by the same contract.7 The optimal mode of licensing is studied in

---

6Such a distinction is used by Salanie (2005) to classify families of contract theory models (pg. 4).

7Hence, Martimort, Poudou and Sand-Zantman (2010) really have both informed and uniformed parties taking initiative.
Bhattacharya, Glazer and Sappington, where entry fees into a research joint venture with simultaneous innovators can influence the degree of information sharing and effort levels. In their model, the innovators first produce an innovation, then compete with each other in a final output market. This is similar to Veraevel and Ventachellum's model where duopolists competing in an output market simultaneously innovate, however in their model the innovation is outsourced to a common R&D laboratory, which is compared to the benchmark of in-house innovation. In both cases, the degree to which the informed party can benefit from spill-overs determines to some extent the organisational form of the industry. In contrast, this chapter starts out assuming that innovation is delegated to a researcher, and evaluates how assignment of IPR is relevant to the contract. The study that bears the most similarity to this endeavour is that of Lai, Riezman and Wang's (2009) model of innovation outsourcing.

At the heart of Lai et al (2009) lies the agency problem of leakage, where the researcher cannot commit (in general) to not exploit their knowledge. However, leakage does not emerge as an agency problem in this chapter. Rather, the problem is analysed here as the external source of benefit in the contracting environment, and access to this external demand is through leaking the innovation through the legitimate means of licensing. Like Lai et al (2009), this chapter uses an incentive theory approach where the developer has all the bargaining power. They ask whether a developer wishing to develop a cost-reducing innovation should organise their innovation activities using an in-house research team, or outsource to an independent research team. So the ownership of the IPR is implicitly a choice variable in their problem. In essence, they investigate a make-or-buy decision for R&D contracts. In contrast, the present analysis takes as given a vertical industry structure - there is no question of innovating in-house.

Unverifiable leakage of knowledge leads to Lai et al (2009)'s central trade-off: the cost of outsourcing R&D, or the erosion of the developer's profit through leakage of the innovation knowledge, is weighed against the increase in efficiency of innovation from employing a specialist researcher. As a result, Lai et al (2009) find that the optimal incentives take the form of a revenue sharing contract. This arises because it is a mechanism that aligns the researcher's incentives with the developer's: the researcher is less
4.2. RELATION TO THE LITERATURE

likely to leak the innovation knowledge if it has a claim to a share of the developer’s profit if leakage is harmful to the developer. If the balance of the effects falls in favour of the efficiency of the researcher, then outsourcing is the equilibrium outcome. Otherwise, the researcher prefers an in-house arrangement.

Unlike Lai et al (2009), this chapter begins by assuming that the developer is in the business of licensing its new technology to competitors. Hence, the literature exemplified by Wang (1998) on insider patentees comes to bear on the investigation. Whenever it is optimal to license the innovation, the developer would choose to do it. IPRs are therefore formalised in this chapter and play a fundamental role in determining who benefits directly from licensing, and the magnitude and determinants of the optimal incentive contract. Both the results of this chapter and Lai et al (2009) imply a dependency of the optimal marginal incentive coefficient on the total revenue generated by the innovation. Hence, the incentives theory framework of both studies shed light on the determinants of the optimal incentive contract and the most efficient IPR ownership arrangement. The studies are however limited in situations where the researcher has some bargaining power. Aghion and Tirole (1994) employ an incomplete contracting approach to deal with that case.

The results of this chapter actually verify a specific case of an earlier result in Aghion and Tirole (1994). Their paper on managing innovations uses incomplete contract theory to generate results on efficient IPR ownership. One of their results replicates Grossman and Hart (1986) in requiring efficient ownership to be one where the ex-ante owner of the IPR is the party whose marginal efficiency is greater, relative to the other parties’ investment.

In this chapter, however, the developer has all the bargaining power and so the researcher can always be kept to its reservation utility level. In Aghion and Tirole (1994), both parties undertake a non-verifiable action that affects the probability of an innovation being created. In the chapter here, innovation size depends on the non-verifiable effort of the researcher, and the developer only offers a lumpy action of offering the contract or not. Nevertheless, the results obtained bear some resemblance: the party with the greater marginal benefit of its action relative to the action of the other party should
retain ownership of the asset. If the reduction in the developer's direct profit as a result of the innovation being licensed is interpreted as a marginal investment, then the Aghion and Tirole (1994) result is analogous.

Finally, the agency problem of this chapter and Aghion and Tirole (1994) is the standard moral hazard problem of unobservability and non-verifiability of effort. Aghion and Tirole (1994) keep their analysis very general. In this chapter, however, the impact of IPR ownership with leakage is investigated in a model with linear incentive contracts. Specifically, the Linear-Exponential-Normal framework of Holmstrom and Milgrom (1987), (1991), as augmented by Laffont and Martimort (2002), is used here to outline a Principal-Agent problem. Their models feature a linear compensation wage that the Principal (developer) offers to the Agent (researcher). Moreover, the functional form of the researcher's utility is assumed to exhibit Constant Absolute Risk Aversion (C.A.R.A), with \( r \) defined to be the Arrow-Pratt absolute measure of risk aversion. Finally, the developer receives a noisy signal about the researcher's effort level, \( e \), rendering effort a non-contractible variable. The noise is additive and assumed normally distributed. This framework requires a linearisation of the researcher's payoff in order to obtain closed form solutions for the contract.

4.3 Analytical Framework

This section employs a Principal-Agent model to establish a framework for analysing the delegation of the IPR. The framework allows for a comparison of the structure of the incentive contract that the DU offers the RU under the different ownership arrangements. First, the model is set up carefully to include the internal and external benefits from the creation of an innovation. Then individual rational and incentive compatible constraints are formalised taking into account the two sources of benefit. Finally, the optimal incentive contract is derived, and its properties evaluated in the conclusion to the section.

The risk neutral principal, or Development Unit (DU) and risk averse agent, or Research Unit (RU) enter a contractual arrangement to produce an innovation of magnitude indexed by \( \theta \in \mathbb{R} \). The units of the measure of the index depend on the context of the
innovation. For instance, for technological innovations, $\theta$ may measure the reduction in marginal cost when using the innovation. For product quality-improving innovations, $\theta$ may be a parameter in the consumers’ utility function.

The innovation may be valuable to both DU and its competitors, the latter to whom it can be sold, or patented and licensed by the holder of the Intellectual Property Rights (IPR) of the innovation. The parameter $\tau$, $0 \leq \tau \leq 1$ reflects the degree to which the innovation can be passed on to the DU’s competitors. Equivalently, $\tau$ is a measure of the specificity of the innovation to the DU. If the nature of the innovation that is contracted for is firm-specific, then $\tau = 0$, and clearly none of the DU’s competitors would purchase a license to use it. On the other hand, if the innovation is of a very general nature or common to the technology that each of DU’s competitors have installed, then $\tau = 1$ and the innovation is valuable to the DU’s competitors. For all other values of $\tau$, only part of the full innovation that the DU can use can be implemented by its competitors.8

Innovation IPR is a valuable by-product of the RU-DU relationship. The IPR can be used to generate revenue in two ways: directly as a result of the DU’s increased competitiveness in the output market, and indirectly through extracting rents created from competing firms using the innovation to produce in the output market. IPR, being a knowledge good, is durable. Hence, the RU team members who create the innovation retain the knowledge of the process of inventing the innovation. Hence, while the product of the knowledge can be legally defined as a property right, it is possible that the information from the IPR can be transferred both formally and informally through the RU’s other activities.

To be explicit, even in the case that the RU does not retain the IPR from its relationship with the DU, it may use its experience to develop an innovation for one of the DU’s downstream competitors at a small cost. Alternatively, the RU may overtly re-sell the innovation to the DU’s competitors, whether it infringes on a patent, or imitates

---

8A justification for use of this device can be found in Lai et al (2009). Their paper identifies adaptability of the outsourced innovation to the production firms’ environment as a feature of R&D activity. Although in the context of their paper, adaptability refers to the relative ease with which an informed in-house team can adapt an innovation compared to an outsourced R&D team. The parameter $\tau$ can be thought of as a measure of relationship-specific investment, although it is taken to be exogenous in the model.
the technology in a way that does not infringe on IPR. In light of this, it is reasonably assumed that the only way the DU can protect its IPR, or appropriate the rents from its licensing, is if it incurs a cost of identifying infringements and enforcing the IPR. Moreover, the RU's expertise allows it to engage in the same activities for a negligible cost. Because of this, the RU can never credibly commit to not leaking the IPR to the downstream competitors, unless the DU patents and enforces the IPR of the innovation.

Hence the IPR can be patented at some cost and licensed, or kept as a trade secret and leaked through the channels mentioned above. In either case, the revenue from the transactions over the new technology will be referred to as licensing. License revenue is denoted $l(\theta, \tau)$. In order to compute a closed form solution to the incentive problem, it is necessary to use a first order approximation to the license revenue:

$$l(\theta, \tau) \approx \kappa(\tau)\theta, \quad \kappa(\tau) \geq 0 \quad (4.1)$$

where $\kappa(\tau)$ is a measure of the aggregate incremental value of a unit of innovation to the DU's competitors: the marginal licensing revenue. The marginal license revenue is non-decreasing in the degree of firm specificity of the innovation, $\kappa'(\tau) > 0$.

The RU can only produce non-drastic innovations. A drastic innovation is defined as one where, in maximising its profit, the innovator can drive the other firms out of the market with its new technology. It follows that a non-drastic innovation does not result in all firms being driven from the market. While it is not essential to the results presented here, it will be assumed that no firm is driven from the market as a result of adoption by one or many competitors of an innovation technology.

It is conceivable that large innovations should emerge in different industry structures than the one considered here. For example, RUs can be contracted by a DU after already developing an innovation, or having some R&D at interim stages of development. In those situations, typically a bargaining framework is used, and the mechanics of incomplete contract theory are relied on to determine optimal ownership arrangements. While ruling out drastic-innovations may reduce the scope of this analysis, it is in accord with the other assumptions of the model. In particular, the modeling framework requires a degree of
4.3. ANALYTICAL FRAMEWORK

linearity that is best implemented by considering only small deviations from the initial optimum.

The RU has a linear innovation technology, where every unit of effort that it exerts, $e$, translates into a unit of innovation, up to the realisation of a random outcome:

$$\theta = e + \varepsilon$$  \hfill (4.2)

where $\varepsilon$ is a random variable normally distributed with mean zero, $\varepsilon \sim N(0, \sigma^2)$. As a result of the random variable in the innovation technology, the realised size of the innovation is only a noisy signal of the agent's effort level. Hence, effort is unobservable and non-verifiable by a third party so that no contracts can be written directly on the RU's effort. However, since effort controls the mean of the innovation technology, contracts can be written on the innovation signal $\theta$, which is positively correlated with effort.

The terms of the contract specify a money payment to the RU contingent on the size of the innovation. The compensation is linear in the innovation: $w(\theta) = \alpha \theta + \beta$, where $\alpha$ is the strength of the marginal effort incentives provided by the DU, and $\beta$ is a fixed income to insure the RU against the uncertainty in the innovation technology. The strength of the $\alpha$ coefficient, which is the DU's control variable, determines the power of the incentive provided for the RU's task. The total money cost of producing an innovation is $C(e)$, which is assumed strictly increasing and convex in effort level. It is also assumed that $C(0) = 0$, $C'(0) = 0$ and $\lim_{e \to \infty} C(e) = +\infty$. The expected value of the RU's payoff is then $w(\bar{\theta}) + \lambda(\bar{\theta}, \tau) - C(e)$, with the expectation taken over $\varepsilon$. The binary variable $\lambda$ indicates ownership of the IPR of the innovation. If the RU owns the IPR then $\lambda = 1$, otherwise the DU owns the IPR and $\lambda = 0$.

The DU's ex post payoff from R&D is denoted as $V(\theta, \tau)$. The benefit to the DU consists of the incremental direct profit, $\Delta \hat{\pi}(\theta, \tau)$, that it makes from selling in the output market with its new innovation-augmented technology, and the license revenue if it is the owner of the IPR. Regardless of who licenses the innovation, if the innovation is of value to the DU's competitors then it is assumed to erode the direct profit of the

---

$^9$These conditions ensure a positive finite amount of effort is exerted at the optimum.
DU. Later on, to obtain a closed form solution for the incentive coefficients, the first order approximation for the DU’s ex post benefit from producing with the innovation is employed: $\Delta \pi(\theta, \tau) = \Delta \pi(\tau) \theta$ for any $\tau \in [0, 1]$. The risk neutral DU’s expected profit becomes:

$$E \bar{V}(\theta, \tau) = \int_{-\infty}^{+\infty} (\Delta \bar{\pi}(\theta + \varepsilon, \tau) + (1 - \lambda)\bar{l}(\theta + \varepsilon, \tau) - w(\theta + \varepsilon))dF(\varepsilon)$$

(4.3)

The timing of the model is illustrated in Figure (4.1) and summarised as follows: at date 0, the DU offers the RU a linear compensation wage, $w(\theta)$ and specifies the ownership mode of any IPR that is created; at date 1, the RU accepts or rejects the offer; at date 2, the RU exerts an effort $\varepsilon$ to produce an innovation according to its innovation technology; at date 3 the innovation outcome $\theta$ is realised; at date 4 the owner of the IPR chooses whether to license the innovation; at date 5 all output market transfers and production decisions are made, and the contract is executed. To operationalise the

$$u(w_{CE}) = E \bar{V}(w(\theta) - C(\theta) + \lambda l(\theta, \tau))$$

(4.4)
where the agent’s utility function \( u(.) \) defined over wealth, \( w \), is C.A.R.A, with \( u(w) = -e^{-rw} \), where \( r \) is the Arrow-Pratt measure of absolute risk aversion. Note that the agent’s effort cost is defined in money terms, and so this is a case of non-separable utility (see Laffont and Martimort (2002), Ch. 5.2.3).

In order to induce the RU to accept the contract, the DU must offer it a level of expected utility equivalent to its reservation utility, or its opportunity cost. This is equivalent to guaranteeing the corresponding certainty-equivalent.

**Lemma 1 (Individual Rationality)** Given the RU’s utility function is C.A.R.A and defined over wealth, \( w \), the RU’s certainty equivalent defined in units of wealth is:

\[
w_{CE} = \beta + (\alpha + \lambda \kappa(\tau))e - \frac{1}{2}e^2 - \frac{5}{2}(\alpha + \lambda \kappa(\tau))^2 \sigma^2
\]

(4.5)

Individual rationality requires \( w_{CE} \geq \bar{w} \), where \( \bar{w} \) is the RU’s outside opportunity.

The individual rationality constraint alone does not provide incentives for the RU to exert effort. Since effort is unobservable to the DU and costly for the RU, the RU selects it in a way that optimises its utility from the R&D activity. This involves a balancing of the sources of revenue from the innovation and the personal effort cost incurred by the RU, while mitigating the risk inherent in the innovation technology. To this end, the RU will select a level of effort on its task to maximise the certainty equivalent payoff in equation (4.5). The RU’s problem is:

\[
e(\lambda) \in \arg \max_{e \geq 0} \{ \beta + (\alpha + \lambda \kappa(\tau))e^* - \frac{1}{2}e^{*2} - \frac{5}{2}(\alpha + \lambda \kappa(\tau))^2 \sigma^2 \}
\]

(4.6)

The solution to the RU’s problem gives the incentive compatibility constraint for the DU. The DU anticipates the behaviour of the RU in choosing its level of effort, and should select the power of its incentive payment accordingly. Recognition of this leads to the following Lemma.

**Lemma 2 (Incentive Compatibility)** In implementing the incentive feasible linear wage schedule, the DU is bound by the following constraint: \( \alpha(\lambda) + \lambda \kappa(\tau) = e(\lambda), \lambda = 1, 0. \)
The intuition for the incentive constraint is straightforward: the RU selects a level of effort where its marginal benefit in units of wealth is equal to the marginal cost of exerting effort, e, measured in units of wealth. When the RU retains the IPR of the innovation, it is rewarded at the margin for the last unit of effort exerted partly through the reward scheme, α, and partly through its private return from license revenues, κ(τ).

The DU’s problem can be written as:

$$\max_{\{a, \beta, e\}} \left[ E_e(\Delta \pi(\tau)\theta + (1 - \lambda)l(\theta, \tau)) - w(\theta) \right], \quad \text{s.t Lemma (1) and (2)} \quad (4.7)$$

In the Principal-Agent relationship, the DU has all the bargaining power which means the IR constraint will always bind. If the DU leaves the RU with any wealth over its certainty equivalent from entering the relationship, then the RU’s expected utility is higher than the utility of its certainty equivalent. Then it would always be possible for the DU to reduce its payment to \(w(\theta) - \epsilon\), for an \(\epsilon > 0\) and still induce the RU to sign the contract. Using the fact that the IR and IC are binding at the optimum solution for effort profiles, the problem can be rewritten as a maximisation of the choice of incentive coefficients \(\alpha\) only. First, using the IC condition in Lemma (2) to eliminate effort choices from the certainty-equivalent wage in Lemma (1) yields:

$$w_{CE} = \beta + \frac{1}{2}(\alpha + \lambda \kappa(\tau))^2(1 - \tau \sigma^2) \geq \bar{w}$$

Since this equation binds at the reservation certainty-equivalent level \(\bar{w}\), an expression of the fixed income component of the linear compensation wage, \(\beta\), can be recovered. Now replacing the new expression for the certainty equivalent wage in program (4.7) yields:

$$\max_{\{\alpha\}} \left[ E_e(\Delta \pi(\tau)(\alpha + \lambda \kappa(\tau)) + (1 - \lambda)l(\theta, \tau)) - \bar{w} + \frac{1}{2}(\alpha + \lambda \kappa(\tau))^2 (1 - \tau \sigma^2) - \alpha(\alpha + \lambda \kappa(\tau)) \right]$$

(4.8)

The DU’s program in (4.8) is concave. The unique solution gives the optimal incentive coefficient \(\alpha^*(\lambda)\), from which the general linear compensation wage can be constructed:
4.3. ANALYTICAL FRAMEWORK

Proposition 1 (Optimal Linear Compensation Wage) The unique solution to the DU's program in (4.8) yields a linear incentive compensation wage given by: \( \{ \alpha^*(\lambda), \beta^* \} \), where:

\[
\alpha^*(\lambda) := \max \left\{ \frac{\Delta \pi(\tau) + (1 - \lambda) \kappa(\tau) - \lambda \kappa(\tau) r \sigma^2}{1 + r \sigma^2}, 0 \right\}, \quad \beta^* := \tilde{w} + \frac{(r \sigma^2 - 1)}{2} \left( \frac{\Delta \pi(\tau) + \kappa(\tau)}{1 + r \sigma^2} \right)^2
\]

where \( \lambda = 1 \) for RU IPR ownership, and \( \lambda = 0 \) for DU IPR ownership.

The optimal incentive coefficient \( \alpha^*(\lambda) \) depends on both the internal and external demands for the innovation, and the risk aversion parameters of the RU's utility. As usual, when the risk aversion parameter \( r \) and the variance of the innovation shock, \( \sigma^2 \) are large, the DU optimally shades the strength of the marginal incentive coefficient. This is to shield the risk averse RU from the full uncertainty in the innovation production, while still forcing the RU to bear some of the risk so as to align their incentives with those of the DU. The risk aversion parameters can be assumed to take values such that \( r \sigma^2 > 1 \), without impacting on the results. Then an increase in \( r \) or \( \sigma^2 \) also raises the fixed component of the wage, \( \beta \). This is the standard insurance versus incentives trade-off in the moral hazard literature.

The magnitude of the marginal incentive coefficient, \( \alpha^*(\lambda) \), reflects the value to the DU from the RU's effort. If the IPR is retained by the DU, then it appropriates the internal value \( \Delta \pi(\tau) \) as well as the external value \( \kappa(\tau) \) from leaking the IPR through licensing. As such, to optimally align the explicit incentives of the RU with the DU, \( \alpha \) incorporates the full value of both these sources of benefit, \( \Delta \pi(\tau) + \kappa(\tau) \), appropriately shaded for risk by \( 1 + r \sigma^2 \): \( \alpha^*(0) = \frac{\Delta \pi(\tau) + \kappa(\tau)}{1 + r \sigma^2} \). In contrast, when the IPR of the innovation is delegated to the RU, the DU only appropriates the internal benefit from its direct profit \( \Delta \pi(\tau) \). Hence, the DU only explicitly incentivises the RU's effort for that value, and not for the license revenue. Moreover, since the RU now faces an implicit incentive through leaking the innovation to the external market, the DU can lower the marginal incentive coefficient: \( \alpha^*(1) = \frac{\Delta \pi(\tau) - r \sigma^2 \kappa(\tau)}{1 + r \sigma^2} \). So under IPR delegation, the RU bears the risk for the effort to satisfy the internal demand for its services, for which it is partially insured and rewarded explicitly by the DU. It is also bears some risk itself for the effort that it exerts to satisfy the external demand, for which it is not insured, but is rewarded implicitly.
through the marginal licensing revenue it receives.

Proposition (1) also reports that the optimal marginal incentive coefficient cannot be negative. A negative marginal incentive implies that the RU would have to provide payments to the DU for provision of its own efforts.\textsuperscript{10} To this end, incentive contracting is defined as \textit{feasible} only if $\alpha(\lambda) > 0$ under the relevant ownership mode. Later in this chapter the key results are studied in a product differentiation model. In that setting, feasibility of incentive contracting yields an IPR ownership result where only innovations whose value to the DU’s competitors are small enough will be delegated.

In summary, it can be seen that the IPR delegation decision will affect the optimal incentive contract that the DU can offer the RU. Leakage through licensing is explicitly rewarded if the IPR belongs to the DU. Under delegation, leakage through licensing is not explicitly rewarded, but rather subsidises the DU’s marginal wage bill. The next section draws out the implications of this subsidy on the total effort level of the RU, the wage bill for the DU under both ownership arrangements, and the impact in the DU’s expected profit.

4.4 Weak Equivalence of Ownership

This section presents the main result of the chapter using the optimal incentive contract derived in Proposition (1) in the previous section. Whenever it would be profitable for the DU to license an innovation to its own competitors, an incentive contract to obtain such an innovation results in an invariance of the DU’s expected payoff to delegation of IPR. This result is underpinned by three key assumptions: first, that the RU has superior information regarding the innovation technology. Second, the DU has all the bargaining power in the contract. Third, the DU is neutral to the risk in the creation of the innovation, whereas the RU is risk averse. Under these conditions, an IPR ownership invariance results.

The first key assumption gives rise to the agency problem. If the DU has perfect and complete information, then it would have no problem in implementing the first-best level of effort from an in-house research team with a simple discrete contract. In practice,

\textsuperscript{10}Note that payments from the RU to the DU are ruled out by assuming that the RU is cash-constrained.
4.4. WEAK EQUIVALENCE OF OWNERSHIP

R&D divisions use a large number of performance measures to monitor their workers' efforts. The literature on performance measures and R&D projects, (for example, see Bergmann and Friedl (2008)), identifies a number of non-financial performance measures for R&D managers. These include functionality points (particularly relevant for software companies), achievement of milestones, and achievement of efficiency and quality standards. This non-exhaustive list includes measures that can be directed at individual, team or workplace levels. Moreover, there is empirical support for the link between the use of such measures and successful innovation as documented in Kressens-van Drongelen and Bilderbeek (1999). The use of these performance measures in innovative activities is evidence that an agency problem exists.

The strong bargaining power assumption can be qualified by focusing attention on small innovations, like process innovations to reduce the marginal cost of a developed product, or quality innovations that capture small increments in market share. In these cases it is reasonable to assume that there are many RUs that are equally capable of undertaking the R&D, and so competition prevents them from earning rents. Notwithstanding, the bargaining power assumption is pertinent to the results that follow. Allowing the RU to have the ability to bargain, or making bargaining power endogenous to the contract is left for future research.

The validity of the assumptions on attitudes to risk can be justified on grounds of scale: typically, DUs are much larger entities than RUs, and have a correspondingly larger and more diverse asset base. As the innovations provided for in the contract are small relative to the size of the DU, they comprise only a small part of the risk profile of the DU, whereas, the RU is considered as a specialist in the area it is contracted for and is therefore not diversified. Even if the RU firm itself is diversified in an array of other projects, the Research manager is assumed to care about the success of the innovation, not just its expected value.

These three key assumptions permit a closer examination of the determinants of IPR ownership on R&D incentives. In agency problems with only an internal benefit of the type described above, the power of the incentive contract is determined by a balance of provision of incentive for exertion of effort on one hand, and the need to insure the agent
against the risk inherent in the contracted activity on the other. The same is true in
the set up here: as has been noted, the optimal incentive contract involves a marginal
incentive coefficient \( \alpha(\lambda) \) that rewards the RU’s marginal effort, and a fixed payment \( \beta \)
that the RU receives regardless of the outcome of their activities (from Proposition (1)).
However, unlike standard agency problems, both of these coefficients depend upon the
internal and external demand for the innovation.

The degree to which the internal and external demands affect the marginal incentive
depend on the delegation of IPR decision. Consequently, the magnitude of the optimal
incentive coefficient, and the elicitation of RU effort through Lemma (2) both depend
upon the delegation decision. To see this more clearly, note that ex ante, the DU has
two control variables: it can choose to delegate IPR to the RU, or retain it. Secondly, it
selects a level for the marginal incentive coefficient, or \textit{explicit} incentives, conditional on
the assignment of IPR.

It follows that if the DU wishes to raise the effort level of the RU, it has two feasible
options. The first is to raise the value of \( \alpha \). Through Lemma (2), the RU optimally
raises its effort in response, since for every dollar that the innovation generates in total
to the DU, the RU receives a greater proportion. The second option is to delegate IPR to
the RU. In doing this, the DU exposes the RU to the implicit incentive channel through
the possibility to license the IPR in the output market. Then the RU receives a share
of the profit realised by the DU in its output market, and all of the licensing revenue
it can generate in the output market. Since it is only capturing the rent from its own
profit, the internal benefit, the DU lowers the level of the marginal incentive coefficient.
Inspection of the incentive coefficient in Proposition (1) confirms that delegation of IPR
ownership results in a lowering of the power of incentives, since \( \alpha(1) - \alpha(0) = -\kappa(\tau) \).
This expression points toward the result on the impact on the overall incentives facing
the RU.

IPR delegation results in two competing effects on the overall level of incentives
facing the RU. There is a reduction in the explicit incentive from the marginal incentive
coefficient offered by the DU, by the full amount of the marginal licensing revenue.
Simultaneously, there is an increase in the implicit incentive offered through the private
market for the innovation. What then, is the overall impact on the effort level of the agent? Proposition (2) provides the answer:

**Proposition 2** With a linear incentive compensation wage, the RU's effort level is identical regardless of who owns the IPR: \( e(\lambda = 1) = e(\lambda = 0) = e^* \).

This Proposition shows that optimal effort level \( e^* \) that the RU exerts is invariant to the IPR ownership mode. This occurs because the DU has all the bargaining power in the contract relationship. As a result, the DU can hold the RU to its reservation certainty-equivalent. Delegation transfers licensing revenue from the DU's competitors away from the DU to the RU, but in doing so, accomplishes two things: (i) delegation perfectly crowds out the explicit marginal effort incentive, and (ii) delegation lowers the DU's total wage bill since it is paying less per unit of effort exerted than when it retains the IPR, and the same amount of effort is exerted. This means that under delegation, the overall incentives facing the RU are still determined by the internal and external demands. The difference is that the implicit external demand effect, which is positive under delegation, is completely offset by the reduction in the DU's explicit incentives. This intuition for Proposition (2) is the critical result of the analysis. Using this invariance of effort to IPR ownership, it is straightforward to prove that the DU's wage bill is lower under RU IPR. This is demonstrated below in Lemma (3):

**Lemma 3** The total wage bill for the DU is larger if it retains the IPR ownership.

The reduction in the total wage bill at first points to the conclusion that the DU prefers to delegate IPR ownership. However, while the license market subsidises the explicit provision of incentives to the RU, the DU still both loses the licensing revenue, and is exposed to the erosion of its profits through the licensing. This means that in deciding whether to delegate IPR ownership, the DU must weigh up the reduction in the wage bill against the loss of total profit. The effort invariance result of Proposition (2) implies that the DU can impute that the expected size of the innovation is the same under either ownership mode. The implication of this is gathered in Proposition (3) below. In a situation where the DU would license if it owned the innovation, the following Proposition provides an equivalence of IPR ownership.
Proposition 3 (Equivalence) Whenever the DU would license the innovation to its competitors, its expected profit is identical under either ownership mode.

Proposition (3) can be understood by a marginal analysis of the explicit and implicit incentives in the problem. An increase in the marginal incentive coefficient by \( da \) raises the RU's effort. This has three impacts: (i) the DU's direct profit increases by \( \Delta \pi(\tau)da \), (ii) an increase in the licensing revenue by \( \kappa(\tau)da \), and (iii) a reduction in the total wage bill by \( \kappa(\tau)da \). Delegation of the IPR to the RU results in changes in the magnitudes and distribution of these marginal effects. To see this more clearly, first note that because the optimal effort level does not change, the DU's profit is reduced by exactly the amount of the licensing revenue under delegation. Since the DU does not accrue the revenue from licensing under delegation, it reduces the incentives to exert effort by just enough to still induce the RU to participate in the project. As a result, the total wage bill reduces under delegation by exactly the same amount as the loss of licensing revenue to the DU. Since the DU is risk neutral and the expected return under both ownership modes is the same, it is indifferent to both. The RU receives the reservation utility in both cases. Under DU IPR ownership the certainty-equivalent is made up by a compensation wage whose marginal incentive coefficient depends positively on the DU's direct profit and the licensing revenue. However, under the delegation, the marginal incentive coefficient depends on the DU's direct profit, but is reduced exactly by the marginal licensing revenue. The RU still exerts the same amount of effort because it has private incentives provided by the return from licensing the innovation in the output market.

This equivalence result suggests that when the DU wishes to be an insider patentee, leakage of the innovation is not the critical factor in determining ownership of IPR as in Lai et al (2009) and Bhattacharya et al (2006). However, by considering the costs and benefits of formalising innovation more carefully, a tie-breaking rule may be fruitful in predicting ownership assignment.

For example, suppose the RU by the nature of their expertise and experience in IP markets are able to enforce or detect infringement of their IPR at a low cost. Alternatively, suppose the DU has to incur a non-trivial cost to protect, enforce and detect an infringement on their IPR which is much greater relative to the RU, since they would
have to hire a team of experts to verify infringement. Then on balance the DU may always prefer to delegate IPR. Hence, in this framework, it is factors like transaction costs of enforcement rather than the provision of incentives that would give rise to efficient ownership assignment of IPR. This is in contrast to the analysis of Lai et al (2009) and Bhattacharyya et al (2006).

Issues of transaction costs aside, the hypothesis of Proposition (2) contains a restriction: for it to apply, the DU would have to want to leak the innovation through licensing on its own account. Essentially this implies that there is no agency problem with leakage. However, in the literature on patent licensing, there is usually a threshold innovation size beyond which the innovator prefers not to license. The next section relaxes the requirement that the DU would wish to license the innovation to establish under what conditions ownership of the IPR matters to the DU.

4.5 When IPR Ownership Matters

In the previous section, the DU was indifferent to delegating IPR ownership to the RU and retaining it. This section discusses the circumstances under which the ownership of the IPR is going to matter to the DU. The hypothesis of Proposition (2) requires that the DU would license the innovation itself. In that case, the DU is indifferent to the mode of IPR ownership. At what point then does the DU prefer to retain IPR?

To help answer this question, it is useful to decompose the DU’s marginal profit from an innovation into the direct impact from operating in the final output market, \( \Delta \pi(0) \), and the external marginal erosion from having its competitors use the innovation, \( \delta(\tau) \): \( \Delta \pi(\tau) := \Delta \pi(0) - \delta(\tau) \). Note that \( \tau = 0 \) in the direct marginal profit since this measures the marginal value of the innovation to the DU gross of the influence of its competitors. In contrast, the magnitude of the marginal erosion depends on how valuable the innovation is to the DU’s competitors, parameterised by \( \tau \). Specifically, it is assumed that when the innovation only has value to the DU there is no erosion: \( \delta(0) = 0 \). Also, as the

---

\[ \text{11See for example Wang (1998) and Hernandez-Murillo and Llobet (2006). Wang (1998) considers a cost-reducing innovation in a homogeneous output market. In his model, for innovations larger than some critical value } \theta^*, \text{ the innovator prefers not to license the innovation. For smaller innovations, the innovator licenses. Hernandez-Murillo and Llobet (2006) consider the case of patent licensing with heterogeneous firms in a product differentiated market.} \]
innovations parametrically become less specific to the DU – \( \tau \) increases – the marginal erosion of the DU's profit increases: \( \delta'(\tau) > 0 \).

Using these definitions, Proposition (4) below characterises the condition for DU IPR ownership. Whenever the marginal erosion of the DU's profit is greater than the marginal extraction of licensing revenue from its competitors, the DU prefers to retain ownership of the IPR. In this case, it would choose not to license. On the other hand, the RU would always choose to license: there is a conflict of interest. This conflict is similar in spirit to the incomplete contracting models, where the action of the agent who controls residual property rights leads to a reduction in the objective of the other agent.

**Proposition 4 (Ownership)** Provided the marginal revenue from licensing exceeds the marginal erosion of the DU's profit, the DU is indifferent to IPR ownership, otherwise the DU prefers to retain the IPR and not license the innovation.

The proof of Proposition (4) is straightforward. It involves evaluating the difference in the expected profit to the DU from delegating IPR ownership to the RU in which case its direct profit is decreased, and retaining the IPR and not licensing. Substituting the optimised linear compensation wage into the DU's expected profit in the situation where it retains IPR ownership and chooses not to license yields the expression \( V_{NL} \):

\[
V_{NL} = \frac{\Delta \pi(0)^2}{2(1 + r\sigma^2)}, \quad V_{RU} = \frac{(\Delta \pi(\tau) + \kappa(\tau))^2}{2(1 + r\sigma^2)}
\]

Note that no licensing corresponds to the case where \( \tau = 0 \). The right hand expression \( V_{RU} \) is the DU's expected profit from delegating IPR whereupon the RU licenses the innovation. The difference between these two expected profits yields:

\[
V_{RU} - V_{NL} = \left( \frac{(\Delta \pi(\tau) + \kappa(\tau)) + \Delta \pi(0)}{2} \right) \left( \frac{\kappa(\tau) - \delta(\tau)}{2(1 + r\sigma^2)} \right)
\]

Since the first factor in the difference in expected profits is the arithmetic average of the total surplus generated under the two arrangements (which is strictly positive), then the sign of the difference in expected profits is governed by the second factor, \( \kappa(\tau) - \delta(\tau) \). Thus, whenever the marginal erosion of the DU’s profit, \( \delta(\tau) \), is larger than the marginal
licensing revenue, $\kappa(\tau)$, the DU prefers to retain IPR ownership and not license the innovation. Otherwise the equivalence result holds: the DU is indifferent to the ownership arrangement.

The result contrasts somewhat to the Lai, Riezman and Wang (2009), to the extent that they can be compared. In their paper, a revenue-sharing contract emerges as the optimal form of contract more often when the RU's benefit from the external demand by leakage is a relatively small fraction of the erosion of the DU's profit from the leakage. So, in their model, the DU is more inclined to have a vertically separated structure, which effectively signs IPR over to the RU, only if it can implicitly control the RU's decision to leak with an incentive contract. In that situation, the DU trades off the probability of leakage occurring against the greater efficiency from the incentives offered to the RU. The opposite is true in Proposition (3). It is only when the erosion in the DU's profit from leakage through licensing dominates the benefit to the RU from leakage that the DU prefers to control the IPR, since in that case it would not wish to license. While it is difficult to make a clear comparison between the two outcomes, the difference in the results does appear to stem from the fact that Lai et al (2009) do not allow that the DU might license the innovation to the external market itself.

It can be seen directly from the condition in Proposition (3) that the degree of firm specificity of the innovation may impact the DU's decision to delegate IPR. The reason is intuitive: innovations which are specific to the DU have low value to the competing firms in the output market, and erode the DU's profit by a smaller amount than general innovations. By imposing some additional structure on $\kappa(\tau)$ and $\delta(\tau)$, it is possible to be more precise about the impact of firm specificity of the innovation. It is reasonable to assume that both the marginal licensing revenue and the erosion of the DU's profit are non-decreasing in innovation specificity. This gives rise to the possibility that on some intervals of the range of specificity the marginal licensing revenue exceeds the erosion of the DU's profit, and on other intervals the reverse is true. This idea is gathered in the Corollary below:
Corollary 1 If $\kappa(\tau)$ and $\delta(\tau)$ are non-decreasing in $\tau$ on $[0, 1]$, then:

(i) if $\kappa(\tau) < \delta(\tau)$ for all $\tau$, then the DU always prefers to retain IPR.

(ii) if $\kappa(\tau) > \delta(\tau)$ for all $\tau$, then the DU is always indifferent to ownership of IPR.

(iii) on any interval $[\tau_0, \tau_1] \subseteq [0, 1]$ where $\kappa(\tau^*) < \delta(\tau^*)$ for $\tau^* \in [\tau_0, \tau_1] \subseteq [0, 1]$, the DU prefers to retain IPR, otherwise it is indifferent to ownership on that interval.

Corollary 1 implies that the DU’s ownership assignment depends on the firm-specificity of the innovation. This is because the relative magnitudes of the marginal licensing revenue and the rate of erosion of the DU’s profit depend on the firm-specificity of the innovation, $\tau$. However, these magnitudes are also industry specific: they will depend on the structural parameters of the industry under study. Different factors, such as the number of firms in the industry, or the nature of the strategic interaction between the firms will impact on the DU’s decision to retain the IPR. Hence, the next section employs a specific output market structure to obtain the sensitivity of the DU’s decision to various market parameters.

4.6 Innovation in a Differentiated Product Market

In this section, a specific market structure is used to explore the IPR ownership issue when the innovation is for cost reduction or quality improving. The exogenous quantity competition model of Vives (1985) provides a useful framework for analysing the optimal ownership arrangement, as it allows for an examination of goods that are gross substitute or complements. Suppose there are $n$ firms, and there is no entry. Each firm produces a quantity of a single product, denoted $x_k$, $k = 1, \ldots, n$. Take firm $i$ to be the DU. For the demand side, suppose that a representative consumer has quasi-linear utility:

$$u(y, x_1, \ldots, x_n) = y + \sum_{j=1}^{n} a_j x_j - \frac{1}{2} \left( \sum_{j=1}^{n} x_j^2 + 2\gamma \sum_{j \neq i} x_i x_j \right)$$

where $y$ is the numeraire good, and $\gamma$ is the degree of substitutability of all the goods, with $-1 \leq \gamma \leq 1$. The goods are strategic complements if $\gamma < 0$, are independent if $\gamma = 0$, and are strategic substitutes if $\gamma > 0$. With this utility representation, the demand
facing each firm is linear in quantities:

\[ p_k = a_k - x_k - \gamma \sum_{j \neq k} x_j, \quad k = 1, \ldots, n \]

where the prices \( p_k, \ k = 1, \ldots, n \) are taken by the consumer to be fixed. The DU competes in the product market by selecting a quantity \( x_i \geq 0 \) to maximise its profits, conditional on the anticipated list of quantities selected by each of its rivals.

Then firm \( k \)'s best-response to the other firm's quantities \( X_{-k} \) is given by:

\[
x_k^*(X_{-k}) \in \arg \max_{x_k \geq 0} \left( a_k - x_k - \gamma \sum_{j \neq k} x_j - c_k \right) x_k, \quad \forall k, \text{ with } \tau = 1 \text{ if } k = i \tag{4.10}
\]

where \( a_k - c_k \) is assumed large enough so that every firm chooses a non-negative quantity of the good. The Nash equilibrium quantity for each firm in the output market at an interior solution is given by:

\[
x_k^* = \frac{(2 + (n - 1)\gamma)(a_k - c_k) - \gamma \sum_k(a_k - c_k)}{(2 - \gamma)(2 + (n - 1)\gamma)}, \quad \forall k \tag{4.11}
\]

Innovations can be either cost-reducing, or quality improving. In either case, it is assumed that innovations are non-drastic. They do not drive other firms from the market. In this linear demand system, cost reducing innovations have the same effect as quality-improving innovations. For a cost-reducing innovation, \( \theta \) is the reduction in the DU's constant marginal cost \( c \), so that its augmented cost per unit produced is: \( c - \theta \). In a symmetric framework, each of the DU's rivals could have access to the innovation, although only \( \tau \) may be installed. Restricting \( a_k = a_j \ \forall k, j \), the profit for firm \( k \) is given by:

\[
\pi_k = (a - x_k - \gamma \sum_{j \neq k} x_j - (c - \tau \theta))x_k = (a + \tau \theta - x_k - \gamma \sum_{j \neq k} x_j - c)x_k
\]

A quality-improving innovation shifts out the demand curve for the innovating firm. To simplify the analysis, the marginal cost is assumed to be constant. That is, suppose \( c_k = c \) for all \( k \). An innovation of size \( \theta \) results in a shift out in the demand by \( \tau \theta \), so that \( a_k = a + \tau \theta \), with \( \tau = 1 \) for \( k = i \). Making these restrictions in (4.11) yields the
same expression for profit as above. Hence, cost-reducing innovations will result in the same optimal output choice as quality-improving innovations in this model. Hence, the optimal output choices for each firm are given by:

\[ x_i^* = \frac{(2-\gamma)(a-c) + (2-\gamma + (n-1)(1-\gamma))\theta}{(2-\gamma)(2+(n-1)\gamma)} \]

\[ x_k^* = \frac{(2-\gamma)(a-c) + (2-\gamma + (n-1)(1-\gamma))\theta}{(2-\gamma)(2+(n-1)\gamma)}, \quad \forall k \neq i \]  

(4.12)

A first order approximation about a zero innovation for the profit of the DU and the license revenue, assuming a fixed fee license are:

\[ \Delta \pi(\theta, \tau, \gamma) := \left( \frac{2(a-c)(2-\gamma + (n-1)(1-\gamma))}{(2-\gamma)(2+(n-1)\gamma)^2} \right), \quad \text{and} \quad \kappa(\theta, \tau, \gamma) := \left( \frac{4(n-1)(a-c)\tau}{(2-\gamma)(2+(n-1)\gamma)^2} \right) \]  

(4.13)

Given this specification, the optimal incentive coefficient and fixed insurance payments are:

\[ \alpha(\lambda) = \left( \frac{2(a-c)(2-\gamma + (n-1)(1-\gamma)(\gamma + (2-\gamma)\tau))}{(2-\gamma)(2+(n-1)\gamma)^2(1+\tau^2/2)} \right) - \lambda \left( \frac{4(n-1)(a-c)\tau}{(2-\gamma)(2+(n-1)\gamma)^2} \right) \]

\[ \beta = \bar{w} + \frac{(\tau^2 - 1)}{2} \left( \frac{2(a-c)(2-\gamma + (n-1)(1-\gamma)(\gamma + (2-\gamma)\tau))}{(2-\gamma)(2+(n-1)\gamma)^2} \right) \]

These expressions can be used to establish necessary conditions on the pairs of values that the innovation specificity and substitutability parameters can take. This is done in Lemma (4):

**Lemma 4** The following restrictions are required on the substitution parameter \( \gamma \) and the degree of innovation specificity \( \tau \) for an innovation incentive contract to exist:

\[ (A) \quad \tau < \min \left\{ \frac{\gamma - 2}{n-1}, \frac{1+\tau^2}{1+2\tau^2} \right\} \equiv \hat{\tau}, \quad (B) \quad \gamma \in I, \quad \text{where} \quad I := \left\{ \gamma | \gamma \in [0, 1] \ or \ -\gamma \leq \frac{2}{(n-1)(1-\gamma)} \right\}. \]

This product differentiation specification can be used to explore the impact of innovation specificity, product substitutability, and the number of firms competing in the same output market as the DU on the IPR ownership decision. In addition, the bearing of the degree of risk aversion and the variance of the innovation technology on incentives can be determined more precisely than in the general case.

It was established after Proposition (1) that \( \alpha(0) > \alpha(1) \), and for an incentive contract to be feasible, \( \alpha(1) > 0 \). That is, the incentive contract must have some “power”,
otherwise the wage payment is simply a fixed fee. Cases where $\alpha(\lambda) < 0$ would mean that the RU must make transfers back to the DU - effectively paying the DU for provision of its own effort. It is not untenable that this could happen. After all, the RU needs the DU in order to create the innovation in the first place that gives rise to the external demand. However, negative marginal incentive coefficients will be ruled out on the grounds that the RU is cash constrained and is unable to obtain external financing (see Aghion and Tirole (1994) for the case where a cash-constrained RU may obtain external finance).

Proposition (5) below establishes that feasibility of RU IPR incentive contracts are bound above by a critical level of innovation specificity. The reason is not because of an agency leakage problem as in Lai et al (2009), but rather that the implicit licensing market subsidisation of the explicit incentive provided by the DU to the RU would require a negative marginal incentive coefficient. For innovation specificity above the critical threshold, DU IPR is always chosen in the production differentiation model. This is because the marginal incentive coefficient $\alpha(0)$ is always strictly positive, whenever the conditions set out in Lemma (4) hold.

**Proposition 5** There exists a $\tau^* \in [0, 1]$ such that for all $\tau \in [0, \tau^*)$, RU IPR incentive contracting is feasible, otherwise the DU prefers to retain ownership of the IPR.

An implication of Proposition (5) is that delegation of IPR would only be observed when the innovation contracted for has little external value. It is more likely that the DU is indifferent to IPR ownership when the consequences for erosion of its profit by licensing is small. Again, other papers have arrived at a similar conclusion, but that is because their analysis trades off an erosion in profit against the efficiency effect of hiring a specialist. Here, the upper bound on innovation specificity arises due to a cash-constrained RU.

Given that there is an innovation specificity upper bound on the feasibility of RU IPR incentive contracting, it is possible to analyse some comparative static effects on the upper bound. In particular, it is possible to examine how the degree of product differentiation, $\gamma$, and the number of firms in the DU's output market influence the critical innovation specificity, $\tau^*$.

Intuitively, an increase in $\gamma$ increases the substitutability of the competing firm's outputs in the eyes of consumers. This lowers the market power that each firm has for
its particular good, hence lowering the external demand effect. Thus, the proportion of licensing revenues in total producer surplus becomes smaller, and so the implicit incentive effect from licensing is smaller, permitting a greater critical pass through of the innovation to the competing firms before a full crowding out of explicit DU incentives occurs.

On the other hand, an increase in the number of firms competing in the output market raises the external demand effect. Therefore, the proportion of licensing revenues in total surplus becomes larger, and the implicit effect becomes larger. The result is the opposite from an increase in product substitutability: the marginal incentive coefficient with RU IPR becomes smaller more rapidly as more firms demand the innovation license, so the critical pass through of the innovation is smaller when full crowding out of explicit DU incentives occurs. Proposition (6) formalises these comparative static effects:

**Proposition 6** In the product differentiation model, an increase in \( n \), a decrease in \( \gamma \) and an increase in \( r \) or \( \sigma^2 \) lower the critical innovation specificity, \( \tau^* \):

\[
\frac{d\tau^*}{dn} < 0, \quad \frac{d\tau^*}{d\gamma} > 0, \quad \frac{d\tau^*}{dr\sigma^2} > 0
\]

The positive relationship between the critical innovation specificity and the size \( r \) and \( \sigma^2 \) arises because the more risk averse the RU is, or the more risky their task is, the lower effort they exert. So on average, the innovation is smaller. Hence, more of the innovation can be passed through to the DU's competitors before the erosion of the DU's profit becomes dominant.

### 4.7 Conclusion

The purpose of this chapter was to investigate the conditions under which a DU would prefer to retain control over the IPR and residual control rights to licensing when contracting for an innovation. The results showed that whenever it would be in the interest of the DU to license an innovation to its competitors, the DU is actually indifferent to delegating IPR ownership to the RU or keeping it. The reason was that the RU still exerts the same amount of effort under both ownership arrangements because its total incentives remain the same: the total marginal incentive under DU IPR is equal to the
sum of the lower explicit incentive and the new implicit incentive from the external source of demand under RU IPR. In addition, the conditions for which the DU would choose to retain the IPR were derived: whenever the rate of erosion of the DU’s profit from the leakage of the innovation knowledge to its competitors was greater than the marginal licensing revenue.

In the context of a product differentiation model, it was established that feasible incentive contracting gives rise to a new ownership result: provided that the innovation specificity, \( \tau^* \) is below a given critical threshold, RU IPR is feasible, otherwise the DU prefers to retain IPR ownership. The critical threshold \( \tau^* \) was shown to be sensitive to the degree of product substitutability, the number of firms that the DU competes with, and the risk aversion parameters of the RU.

The results of this model were generated by imposing some strict linearity conditions to get a closed form solution for the incentive coefficients in the linear compensation wage. The next step is to determine whether the equivalence of IPR ownership result holds more generally. If the result is more general, then the interesting question becomes how the equivalence of IPR ownership can be broken. This line of research is currently under study.

### 4.8 Appendix

#### 4.8.1 Proof of Lemma 1

Expanding the expression for the certainty equivalence of RU, \( u(w_{CE}) = \mathbb{E}_\varepsilon u(w(\theta) - C(\epsilon) + \lambda(\theta, \tau)) \), yields:

\[
-e^{-r w_{CE}} = -\int_{-\infty}^{+\infty} e^{-r(\beta + \alpha e + \alpha \epsilon - \frac{1}{2} \sigma^2 + \lambda \kappa(\tau) e + \lambda \kappa(\tau) \epsilon)} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\theta - \mu)^2}{2\sigma^2}} d\varepsilon
\]

\[
= -e^{-r(\beta + \alpha e + \lambda \kappa(\tau) \epsilon - \frac{1}{2} \sigma^2)} \int_{-\infty}^{+\infty} e^{-r(\alpha + \lambda \kappa(\tau) \epsilon)} e^{-\frac{1}{2} \sigma^2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\theta - \mu)^2}{2\sigma^2}} d\varepsilon
\]

\[
= -e^{-r(\beta + (\alpha + \lambda \kappa(\tau)) \epsilon - \frac{1}{2} \sigma^2)} e^{-r(\frac{1}{2} (\alpha + \lambda \kappa(\tau))^2 \sigma^2)}
\]

\[\iff w_{CE} = \beta + (\alpha + \lambda \kappa(\tau)) e - \frac{1}{2} \epsilon^2 - \frac{r}{2} (\alpha + \lambda \kappa(\tau))^2 \sigma^2\]
where the third equality follows using the moment generating function:

\[
\int_{-\infty}^{+\infty} e^{-r(\alpha+\lambda\kappa(\tau))\varepsilon} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\theta-\varepsilon)^2}{2\sigma^2}} d\varepsilon = e^{-\frac{1}{2}r^2(\alpha+\lambda\kappa(\tau))^2\sigma^2}
\]

Rewriting the last line yields the result. ■

4.8.2 Proof of Proposition 1

Taking the derivative with respect to \(\alpha\) of the maximand of the DU’s program and using (4.8) yields:

\[
\Delta \pi(\tau) + (1 - \lambda)\kappa(\tau) + (\alpha + \lambda\kappa(\tau))(1 - r\sigma^2) - (2\alpha + \lambda\kappa(\tau)) = 0 \quad (4.14)
\]

Solving this equation for \(\alpha\) yields the result. The second order condition is satisfied:

\(- (1 + r\sigma^2) < 0.\) ■

4.8.3 Proof of Proposition 2

The result follows directly from computing the equilibrium effort level from the RU's incentive compatibility constraint (2):

\[
e(\lambda) = \left( \Delta \pi(\tau) + (1 - \lambda)\kappa(\tau) - \lambda\kappa(\tau)r\sigma^2 + \lambda\kappa(\tau)(1 + r\sigma^2) \right) \frac{1}{1 + r\sigma^2} = \frac{\Delta \pi(\tau) + \kappa(\tau)}{1 + r\sigma^2} = e^*
\]

since the second equality shows no \(\lambda\) dependence. ■

4.8.4 Proof of Lemma 3

Starting with: \(w(\theta) = \beta + \alpha(\lambda)e(\lambda),\) where \(\alpha(\lambda)\) is from Proposition (1) and \(e(\lambda)\) is the right hand side of the IC condition of Lemma (2). Writing in full,

\[
w(\theta) = \beta + \left( \frac{\Delta \pi(\tau) + (1 - \lambda)\kappa(\tau) - \lambda\kappa(\tau)r\sigma^2}{1 + r\sigma^2} \right) \left( \frac{\Delta \pi(\tau) + \kappa(\tau)}{1 + r\sigma^2} \right) \quad (4.15)
\]
where it can be seen that $\beta$ is independent of $\lambda$. Hence, the difference in the wage bill is due only to the difference in the expected marginal incentive bill:

$$\Delta w(\theta) = w(\theta)_{\lambda=1} - w(\theta)_{\lambda=0} = \alpha(1)e_1 - \alpha(0)e_0 = -\frac{\kappa(\tau)(\Delta \pi(\tau) + \kappa(\tau))}{1 + \tau \sigma^2} < 0$$

(4.16)

Hence, the total expected wage bill is smaller under RU IPR ownership. ■

4.8.5 Proof of Proposition 3

The proof simply shows that if the DU would choose to license the innovation, then the reduction in its expected wage bill that would occur if it signed over the IPR to the RU is exactly offset by the loss of licensing revenue that it would incur. Define the change in the ex ante expected value to the DU as:

$$\Delta V(\theta, \tau) = E_\theta \left[ \Delta \pi(\tau)(\alpha(1) + \kappa(\tau)) - (\Delta \pi(\tau)\alpha(0) + \kappa(\tau)\alpha(0)) - \Delta w(\theta) \right]$$

From Lemma (2), the optimal amount of effort exerted is the same under either mode of ownership, hence the expected magnitudes of the gross profit effects are identical. Accounting for this and substituting for the linear approximation of the license revenue yields:

$$\Delta V(\theta, \tau) = E_\theta \left[ -\kappa(\tau)\alpha(0) - \Delta w(\theta) \right]$$

The result follows from Lemma (3) and the solution for the optimal incentive coefficient $\alpha(1)$ from Proposition (1). ■

4.8.6 Proof of Lemma 4

For an incentive contract to exist, a necessary condition is for $\Delta \pi(\tau) \geq 0$. Since $a > c$ and $|\gamma| \leq 1$ by assumption, examination of the direct profit expressions show that it is always positive. In comparison, a quality-improving innovation requires condition (B). For condition (B) to hold, its denominator must be positive. This gives the first component of the minimum argument in condition (A) $\tau < \frac{a-2}{\alpha}$. For an incentive contract to have power, it requires $a > 0$. Moreover, discussion of the comparative statics for the optimal incentive coefficient of equation (1) indicate that $\alpha(1) < \alpha(0)$. So, for $\alpha(1) \geq 0$,
CHAPTER 4. INNOVATION CONTRACTS

then in general terms \( \frac{\Delta \pi(\tau)}{\kappa(\tau)} \geq r\sigma^2 \). In the parameters of the product differentiation model, \( \tau \leq \frac{1 + r\sigma^2}{1 + 2\rho\sigma^2} \), which accounts for the second component of condition (A).

4.8.7 Proof of Proposition 5

Whenever \( \alpha(1) < 0 \), even though \( \mathbb{E}_\pi(1) = \mathbb{E}_\pi(0) \), incentive contracting with IR IPR requires the RU to compensate the DU for its own effort, therefore RU IPR incentive contracts are not feasible. On the other hand \( \alpha(0) > 0 \), \( \forall \tau \), which implies that DU IPR would be feasible. The proof of existence of a critical innovation specificity \( \tau^* \) involves finding the value of \( \tau^* \) where \( \alpha(1, \tau^*) = 0 \). It is straightforward to see that \( \alpha(1) \geq 0 \Leftrightarrow \frac{1}{1 + r\sigma^2} \geq \frac{\kappa(\tau)}{\Delta \pi(\tau) + \kappa(\tau)} \). Define \( f(\tau) := \frac{\kappa(\tau)}{\Delta \pi(\tau) + \kappa(\tau)} = \frac{2(n-1)r}{(2+\gamma+(n-1)(\gamma+(2-\gamma)\tau))} \). Then \( \tau^* \) is implicitly defined by \( f(\tau^*) \geq \frac{1}{1 + r\sigma^2} \). Note that:

\[
f'(\tau^*) = \frac{2(n-1)(2+\gamma)}{(2+\gamma+(n-1)(\gamma+(2-\gamma)\tau))^2} > 0, \quad \forall n, \gamma \in I
\]

Hence \( f(\tau) \) is strictly increasing and continuous on \([0,1]\). Also, \( f(0) = 0 \) and \( f(1) = \frac{2(n-1)}{(2+\gamma+2(n-1))} < 1 \). There are two cases: (I) If \( f(1) > \frac{1}{1 + r\sigma^2} \) then RU IPR incentive contract is feasible for all \( \tau \). (II) If \( f(1) < \frac{1}{1 + r\sigma^2} \), then by the IVT, there exists a unique \( \tau^* \in [0,1] \) such that \( f(\tau^*) = \frac{1}{1 + r\sigma^2} \). Then for \( \tau \in [0, \tau^*] : f(\tau) < \frac{1}{1 + r\sigma^2} \) which implies that RU IPR incentive contracting is feasible. For \( \tau \in [\tau^*, 1] : f(\tau) \geq \frac{1}{1 + r\sigma^2} \) which implies that RU IPR incentive contracting is not feasible.

4.8.8 Proof of Proposition 6

The equation \( f(\gamma, n, r^*(\gamma, n)) = \frac{\kappa(r^*)}{\Delta \pi(r^*)} = \frac{1}{1 + r\sigma^2} \) implicitly defines \( r^* \). Totally differentiating this equation with respect to \( i = \gamma, n \) yields:

\[
\frac{\partial f}{\partial \gamma} + \frac{\partial f}{\partial r} \frac{dr^*}{d\gamma} = 0, \quad i = \gamma, n, r\sigma^2
\]

Solving for the comparative static yields: \( \frac{dr^*}{d\gamma} = -\frac{\partial f/\partial \gamma}{\partial f/\partial r} \) for \( i = \gamma, n, r\sigma^2 \). Then the signs of the comparative statics are:

\[
\frac{dr^*}{d\tau} = -\text{sign} \left( \frac{\partial f}{\partial r} \right), \quad \& \quad \frac{dr^*}{dn} = -\text{sign} \left( \frac{\partial f}{\partial n} \right), \quad \& \quad \frac{dr^*}{d\sigma^2} = -\text{sign} \left( \frac{\partial f}{\partial \sigma^2} \right)
\]
Using the definition of $f(\tau)$ from above, the signs of the derivatives of this function with respect to each of the variables are:

\[
\frac{\partial f}{\partial \gamma} = \frac{-2(n-1)\tau(1+(n-1)(1-\tau))}{(2+\gamma+(n-1)(\gamma+(2-\gamma)\tau))^2} < 0
\]

\[
\frac{\partial f}{\partial n} = \frac{2\tau(2+\gamma)}{(2+\gamma+(n-1)(\gamma+(2-\gamma)\tau))^2} > 0
\]

\[
\frac{\partial f}{\partial \sigma^2} = -\frac{1}{(1+r\sigma^2)^2} < 0
\]

Using the signs of these derivatives gives the result. ■
Chapter 5

Concluding Remarks

This thesis examined procurement, franchise and innovation contracts in the shared analytical framework of the Principal-Agent model from the theory of incentives. While the idiosyncrasies of each application are borne out in their respective institutional environment, they are connected through the common need for the principal to resort to contracts to coordinate the agent’s otherwise opportunistic behaviour. Consequently, in each application, the nature of the principal’s contracts reflect the way in which the agent’s incentives are affected by its environment. This section concludes the thesis by highlighting the connection between the principal’s contracts and the wrinkle in the environment that was examined for each application. At the same time, the limitations of the resulting contracts are evaluated, and directions for future research, where applicable, are discussed.

The objective of Chapter 2 was to study the robustness of the linearity property of Laffont and Tirole’s (1986) menu of procurement contracts to an environment where the agent has type-dependent reservation utility. With this modification, the agent’s reservation utility is negatively correlated with its production cost. In this way, an agent that would be more profitable in an external private market has a higher reservation utility. Moreover, the magnitude of the reservation utility is assumed to depend on a private market index – some measure of market power or quality of the agent in the private market. This allows the impact of the agent’s type-dependent reservation utility on the principal’s equilibrium contracts to be metered.
It was shown that in the modified environment the principal could still implement optimal procurement contracts with a menu of linear contracts provided that a convexity condition was met on the agent’s reservation utilities. This condition emerges because the convexity of the agent’s type-dependent reservation utility affects the balance of the marginal incentives facing an agent to truthfully report its cost type to the principal: on one hand, the agent had incentive to under-report its cost to save on disutility of effort – the standard incentive effect. On the other, the agent’s type-dependent reservation utility induces a countervailing incentive effect whereby the agent could receive a larger compensation payment by convincing the government that it has a higher reservation utility.

To minimise its expected procurement cost, the principal’s contracts involved a trade-off in distorting allocative efficiency for rent extraction in order to incentivise the agent to report its true cost type. As in Laffont-Tirole, the principal must still trade-off allocative efficiency for rent extraction; however now the induced allocative efficiency may be an under-provision of optimal cost reducing effort if the direct effort incentive is dominant, or may be an over-provision of effort if the countervailing incentive dominates. Hence the strength of the countervailing incentive effect, as controlled through the private market parameter, has a large impact on the nature of the principal’s incentive contracts.

The connection between the private market parameter and the project that the government procures could be modeled in more detail in an extension of the analysis. As it stands, the private market parameter exists by hypothesis, and was only employed as a device to test the robustness of the linearity property to various strengths of the countervailing incentive effect. However, further study of the connection between the private market and the procurement environment may yield greater insight into the feasibility of the results of Chapter 2.

In Chapter 3, the franchising arrangements between an upstream manufacturer and a downstream retailer were examined. In particular, the commonly used institutional environment of franchising\(^1\) was extended to include the case where (i) the retailer must sign the contract before privately observing the precise level of demand in the final output

\(^1\)Commonly used, in the sense of Rey and Tirole (1986).
market, and (ii) faces an ex-post limited liability constraint on each state of demand. It was shown that in such an institutional franchise environment the manufacturer's usual contracting solution of offering a variable fee schedule would leave the retailer with an expected information rent. However, by including an additional instrument in the form of a fixed franchise fee, the manufacturer could appropriate that expected information rent.

The environment was then allowed to include situations where the retailer could take an unobservable and non-verifiable action to positively influence the probability distribution over future demand states, or have private information about the probability distribution over future demand states. The analysis proceeded to examine the manufacturer's use of the fixed franchise fee in both these cases to elicit the efficient action in the former case, and a truthful report from the retailer in the latter case.

The analytical framework specified that the manufacturer and retailer both be risk neutral. Moreover, it was argued that retailers entering franchise agreements are likely to be bound by limited liability due to, among other things, their size. Consequently, it might reasonably be assumed that the same retailer's attitude to risk may best be described by risk aversion. While this may be a more realistic modeling assumption, it would necessarily complicate the optimal contracts without providing any additional insight. Since the focus was on the use of the fixed franchise fee to appropriate expected information rent and to explore its role in solving a first stage informational asymmetry, the complication of adding risk aversion to the model would only serve to correct a bias in the magnitudes of the power of the incentives offered by the manufacturer.\(^2\)

Another limitation of the model of Chapter 3 was in restricting the institutional environment to only include a once-off interaction between the manufacturer and retailer. In practice, franchise agreements would typically involve repeated contracting, and to the extent that demand realisations are positively correlated across time, the contracts offered in the static setting would in general not be robust to renegotiation or commitment by the manufacturer. Extension of the model to a repeated contracting process would require

---

\(^2\)Rey and Tirole (1986) explored the effect of retailer risk aversion, however their objective was to study the efficacy of different vertical restraints under different types of uncertainty, hence attitude to risk is an essential part of their results.
the tools of dynamic contracting, and is left for future research.\textsuperscript{3}

In Chapter 4, the contractual arrangement for creation of an innovation between a risk neutral developer and a risk averse researcher was evaluated. The institutional environment was extended to include the possibility of licensing the resultant innovation with a third party that competes with the developer – an external source of revenue for either the developer or researcher depending on who owned the intellectual property rights (IPR) of the innovation. As a result, in addition to offering a compensation wage that was contingent on the outcome of the innovation process, the developer had to include an assignment of IPR in the contract. The key result of the model showed that in environments where the developer would choose to license the innovation to its competitor, it is indifferent between retaining IPR or delegating it to the researcher. This occurred because when the researcher was assigned IPR, it became implicitly incentivised by the external source of revenue from licensing. Hence, the reduction in the expected wage paid to the researcher under the terms of the contract, exactly offset the loss the developer faced from having the innovation licensed to its competitors.

In a subsequent result, it was shown that whenever the marginal revenue that could be raised by licensing the innovation was less than the marginal erosion of the developer’s profits from licensing, the developer would then prefer to retain control of the IPR. Finally, by specifying a quantity competition model and parameterising the firm specificity of the innovation, it was demonstrated that the equivalence result on IPR for the developer could be obtained for innovations where the proportion of licensing revenues as a proportion of total producer surplus was small enough. Otherwise, the developer preferred to retain the IPR. In addition, the sensitivity of the assignment of IPR rights to model parameters was explored.

In order to derive a closed form solution for the components of the linear compensation wage in Chapter 4, it was necessary to take linear approximations of the developer’s profit function and the licensing revenue from the innovation. In doing so, the analysis is confined to evaluating only small innovations, and as such, drastic innovations are ruled out. In fact, the equivalence result on IPR ownership may hold in a more general sense.

\textsuperscript{3}See Laffont and Martimort (2002), Salanie (2005) and Bolton and Dewatripont (2005) for expositions of the dynamics of incentive contracts.
In such an analysis, the focus of research would be purely on optimal IPR assignment, and therefore would most likely employ the techniques of incomplete contract theory. Future research may take this direction, using the model of innovation management outlined in Aghion and Tirole (1994), with an extension to include third party licensing effects.

The analysis could also be extended by incorporating features of the institutional environment, such as: researcher bargaining power, or mode of licensing (fixed fee or royalties), or competing research laboratories as a way of breaking the equivalence of IPR ownership.
Bibliography


