Evans, Lewis; Guthrie, Graeme; Quigley, Neil

Working Paper

Contemporary Microeconomic Foundations for the Structure and Management of the Public Sector

New Zealand Treasury Working Paper, No. 12/01

Provided in Cooperation with:
The Treasury, New Zealand Government

Suggested Citation: Evans, Lewis; Guthrie, Graeme; Quigley, Neil (2012) : Contemporary Microeconomic Foundations for the Structure and Management of the Public Sector, New Zealand Treasury Working Paper, No. 12/01, ISBN 978-0-478-39639-3, New Zealand Government, The Treasury, Wellington

This Version is available at:
https://hdl.handle.net/10419/205620

Contemporary Microeconomic Foundations for the Structure and Management of the Public Sector

Lewis Evans, Graeme Guthrie
and Neil Quigley

NEW ZEALAND TREASURY

WORKING PAPER 12/01

MAY 2012

New Zealand Government

NZ TREASURY WORKING PAPER 12/01

Contemporary Microeconomic Foundations for the Structure and Management of the Public Sector

MONTH/YEAR

May 2012

AUTHORS

Lewis Evans
Victoria University of Wellington
PO Box 600
Wellington
New Zealand

Email lew.evans@vuw.ac.nz
Telephone 64 4 463-5560
Fax 64 4 463-6798

Graeme Guthrie
Victoria University of Wellington
PO Box 600
Wellington
New Zealand

Email graeme.guthrie@vuw.ac.nz
Telephone 64 4 463-5763
Fax 64 4 463-6798

Neil Quigley
Victoria University of Wellington
PO Box 600
Wellington
New Zealand

Email neil.quigley@vuw.ac.nz
Telephone 64 4 463-5083
Fax 64 4 463-6798

ISBN(Online)

978-0-478-39639-3

URL

Treasury website at May 2012:
http://www.treasury.govt.nz/publications/research-policy/wp/2012/12-01
Persistent URL: http://purl.oclc.org/nzt/p-1446

ACKNOWLEDGEMENTS

The authors acknowledge the assistance they have received from participants in seminars at the Treasury, and referee reports from Peter Bushnell, Ig Horstmann, Kirsten Jensen and Graham Scott.

NZ TREASURY

New Zealand Treasury
PO Box 3724
Wellington 6008
NEW ZEALAND

Email information@treasury.govt.nz
Telephone 64 4 472-2733
Website www.treasury.govt.nz

Abstract

The new public management of the 1980s was based in part on a range of important new insights about the role of transaction and agency costs arising from contractual incompleteness in defining the boundaries of the firm and the governance relationships within it. In this paper, we consider the literature of the last 25 years which extends our understanding of allocations of ownership rights and the boundaries of the firm as responses to contractual incompleteness. From this perspective, ownership represents an allocation of control rights to those with the potential to make the most important (value-enhancing) relationship-specific investments. We provide an outline of this modern approach to contractual incompleteness, illustrate its application to a range of issues in public and private ownership, investment, governance and decision-making, and provide suggestions about the impact that this approach might have on the scope, structure and management of the public sector in the 21st century.

JEL CLASSIFICATION H40, L20, L33, G31

KEYWORDS Incomplete contracts; investment decision-making; public ownership, management and governance

Summary

The new public management of the 1980s was based in part on a range of important new insights developed by academic economists over the preceding 15 years. Those insights were based on thinking about transaction and agency costs as central to defining the boundaries of the firm and the governance relationships within them, separation of purchase and ownership interests, and the benefits of competitive delivery of commercial services that had previously been provided by government monopolies. This literature focused on contractual incompleteness (the inability to write contracts that explicitly address every contingency relevant to the contract) as a driver of transaction costs, and on transaction cost minimisation as a driver of organisational boundaries.

In this paper, we explore the theoretical literature developed in the last 25 years which focuses on allocation of residual decision rights through ownership decisions as the fundamental response to contractual incompleteness. The ability to exercise residual control rights increases incentives to make relationship-specific investments by improving the *ex post* bargaining position of an asset owner. Asset ownership should therefore be assigned to those with the potential to make the most important (value-enhancing), relationship-specific investments.

As an example, consider the choice between public and private ownership. This choice defines allocations of residual decision rights and control over all of those aspects of any activity which cannot (efficiently) be specified explicitly in a contract. Debates that focus on issues such as who collects the dividends generated by the firm completely miss the point that decisions on ownership should be driven by the desire to allocate residual decision rights to the party most able to make value-maximising decisions about the operation of the firm.

Decisions to undertake investment projects will often be incomplete in the sense that different aspects of the implementation, including the timing of the investment and the phasing of the investment, will often be at the discretion of the party implementing the investment rather than the party providing overall approval for the project. The residual decision rights are often captured in the ability to exercise “real options,” that is, options relating to the timing or phasing of investment that are extinguished once investment proceeds. Public infrastructure decisions are often driven by political imperatives to specify start and completion dates *ex ante*, which carries with it two dangers; that the removal of flexibility about start dates will undermine value or increase costs associated with the project, and that the project will be undertaken immediately, when a delay in starting would maximise efficiency.

The importance of allocating incomplete contractual rights over real options carries over to public-private partnerships. Where these require the private infrastructure provider to start construction on a particular day or remove from the private infrastructure provider the residual decision rights to determine whether investment in research on innovations will take place, and whether the innovation will be implemented, much of the value of involving a private party in the project may be lost. The incomplete contracts literature provides the potential to get past what had become a rather sterile and ideological debate on the role of the private sector in the ownership and management of facilities delivering public services by refocusing the debate on the allocation of residual decision rights that will add the most value to the project.

For service firms, in particular, the customer base is a critical asset that will determine the success of the firm, and this is especially true where long-term customer relationships are important. In this context, the literature on incomplete contracts suggests that the allocation of residual ownership rights over assets such as the customer base of a service firm may be a critical determinant of the success of the firm. Government is involved in wide range of contracting for services, and these should be organised so as to allocate residual decision rights in the customer relationship to the parties that add the most value to the relationship. From this perspective:

- The importance of the relationship between general practitioners (GPs) and their patients provides an explanation for the fact that GPs are the part of the health sector that in almost every country is privatised (that is, GPs, not the government, "own" the relationship with the patient).

- Community-based lending programmes (such as microfinance) may be effective in utilising relationships and moral suasion within communities to replace high cost monitoring and enforcement mechanisms used by loan companies and loan sharks.

- Whānau Ora and other community-based initiatives can perhaps be seen as about providing a community organisation with ownership of the relationship with the client/patient.

In a world where contracts are incomplete, corporate governance defines the rights to make *ex post* decisions in relation to unspecified contingencies shifting the focus of the analysis of governance away from agency (incentive) problems and towards decision and investment problems. A common justification for creating Crown entities is the substance, or perception, of greater independence from ministerial direction by comparison with departmental heads. But where a minister has ultimate responsibility for the performance and strategic direction of the organisation, the minister will wish to exercise residual decision rights on most matters of importance. These problems could be resolved in three complementary ways:

- Separate those entities in which the boards serve primarily as advisory boards rather than as governance boards, and reconstitute them as such. This would make it clear that for those entities, the minister rather than the board had the responsibility for addressing contractual incompleteness.[1]

- Establish a much clearer distinction between public-sector organisations where governance can be effectively delegated to a board of directors, and enshrine that separation from ministers much more clearly in the constitutions of those entities. Where contracts are highly incomplete, then governance would be improved by making the boards advisory boards only.

- Change the structure of the boards that remain, making them smaller to ensure that the whole board can be informed of the minister's views, and leave them to make decisions.

A key issue for the public sector in New Zealand is the need to generate greater coordination between a large number of public-sector agencies. Much of the academic literature addressing the issue of coordination is in the context of consideration of a choice between a "divisional structure" and a "functional structure," or a centralised or a

[1] A range of possible approaches to achieving this present themselves, including a dramatic reduction in the number of Crown entities, and even a questioning of the whole governance and reporting framework for Crown entities.

decentralised allocation of residual decision rights within a hierarchical structure. In the public sector, this literature may be translated as suggesting that specialised departments and agencies are better at identifying projects (have more local knowledge), worse at selecting which projects to invest in, and more efficient in implementing the projects that are chosen. If the identification of projects *ex ante* is not critical (for example, if this is driven by ministers rather than by the specialised knowledge of civil servants), but decisions about which projects to invest in are critical, then large hierarchical government departments with multiple divisions may be preferred.

The more the public sector focuses on policy, as opposed to service delivery and asset ownership, then the more likely it is that coordination will be important. Thus it is likely that a smaller number of larger public-sector organisations will be preferred, because they may provide individual units with incentives to develop specialised expertise and collect information while also providing high level coordination of the investment and resource-allocation decisions made.

The literature on incomplete contracts, therefore, suggests new ways of thinking about problems, and directions for organisational change that might result in a 21st century public sector being structured around:

- Fewer public-sector organisations and further thinking about contestability in policy advice and service delivery both within and between large public-sector organisations.

- Greater clarity in monitoring, with only one central monitoring agency being given "ownership" of public-sector performance.

- A clearer distinction between organisations where governance can be delegated to a board, and organisations where the minister retains residual decision rights.

- Less public ownership of service delivery, with wider delegation of responsibility for investment in outcomes and customer relationships to private or community service delivery organisations.

- Increased levels of private ownership of state infrastructure assets, with the contracts taking advantage of the recent literature on optimal contractual arrangements.

- Greater attention to the structure of compensation and reward, both to change incentives and to change the type of people that are attracted to public-sector organisations.

Each of these areas of policy development will of course need to be the subject of more detailed consideration and application of the framework that we have provided before explicit policy recommendations could be provided.

Contents

Contemporary Microeconomic Foundations for the Structure and Management of the Public Sector

1 Introduction

In the decade after 1984, New Zealand implemented a programme of state sector reform that placed it at the forefront of international debate about the scope, structure and management of public-sector activities in modern developed economies. Today, the New Zealand public sector continues to operate within the framework for public management laid down in that period.

New Zealand's model of public management had its origins in the ideas of reform-minded politicians and policy advisers who found support for change in the academic literature in economics, accounting and management that had emerged during the 1970s and early 1980s (Scott, 2001). In respect of economics, many of those ideas related to the development of institutional economics, and in particular its focus on:

- Transaction costs as explanations for the organisation of firms and markets.
- Principal-agent theories of economic interaction, particularly the cost of acquiring, and the asymmetric distribution of, information, and the pervasiveness of moral hazard and adverse selection in management and contracting problems.
- Contestability and a rethinking of the importance of competitive markets in maximising consumer surplus and social welfare.

These developments in the academic economics literature provided theoretical support for advice that supported a large-scale reduction in government ownership of commercial activities, and substantial changes to the way in which what remained in the public sector was managed.

This study considers what has changed in the academic literature in microeconomics over the past 25 years, and how those changes in the literature might inform new thinking about the role, structure, and management of the public sector. One would hope that, as a result of the research efforts of tens of thousands of academic economists around the world, the answer would be that much has changed—and indeed it has.

To keep the task manageable, we have focused our analysis on developments in those fields that relate directly to contemporary policy issues. In particular, we focus on developments in three broad areas of microeconomics:

- incomplete contracts, ownership rights, and the theory of the firm
- agency relationships, governance and incentives inside the firm, and
- real options and investment.

Each of these areas of literature has implications for a wide range of contemporary policy issues including public ownership, investment and divestment, the effectiveness of governance and management in the state sector, and alternative models of state service delivery. Inevitably, a range of active research programmes in microeconomics, including those in game theory, behavioural and experimental economics, and law and economics, are not considered, or are considered only tangentially as they touch on the literatures in the core areas identified above.

In each case, the task of this paper is to identify those developments in the literature that have relevance to current policy issues including interest in reducing the scale of, and increasing the return from, public expenditure, and the rationale for public ownership of a wide range of activities. Where possible, we use elements of the current organisation of the public and private sector to illustrate the points made by the (largely) theoretical literature. However, we have not attempted to provide advice about particular policy issues, since one of the lessons obtained from the increasingly rich and diverse literature in microeconomics is that careful analysis of each issue rather than vanilla policy prescriptions is indicated at every level.

The paper follows the literature in economics and management in focusing on questions of efficiency and social welfare or total surplus maximisation as the benchmark against which organisation and decision structures are assessed. Much of the work that we consider is drawn from a theoretical literature that uses the language of private-sector management and decision-making, but this does not reduce its relevance to issues in public management—indeed it is an important strength. The rigour of academic economics derives in part from its methodological precision in the analysis of the issues being studied, and its ability to separate questions of efficiency from questions about political objectives and income redistribution. Questions about the complexity of the public sector and the valuation of different outcomes normally reflect political and distributional issues about which economics and management have little to say. The benefit to be obtained from the application of economics to public management is to obtain for the public sector the benefits of efficiency and social welfare maximisation, creating wider degrees of freedom to consider political valuations and redistribution, as well as a clearer understanding of the implications of those choices for the wealth of society as a whole.

Following a brief outline of the scope, structure and management of the public sector in New Zealand as it has emerged from the post-1984 reforms, we begin in Chapter 2 with an analysis of recent developments in the theory of the firm and ownership. We point out that the incomplete contracts literature which is now the focus of attention in this literature provides a compelling theory of ownership which is a substantial advance on the literature of the 1970s and early 1980s.

In Chapter 3, we summarise the literature on real options and investment. A contribution of our work is to point out that the exercise of real options is a substantial example of contractual incompleteness, where the allocation of ownership rights to a party who can exercise the real options is critical for efficient investment decision-making.

In Chapters 4 and 5, we utilise the insights of the incomplete contracts and real options literature to consider aspects of public and private-sector ownership, service delivery and procurement. Chapter 4 focuses on the case where services are associated with the construction of a specific capital intensive facility (prisons being a commonly used example), while Chapter 5 considers public and private service delivery in the absence of a requirement for specific capital investment.

In Chapter 6, we consider governance in the public sector, utilising the incomplete contracts framework. We point out that the role of governance is to allocate residual decision rights that allow resolution of contractual incompleteness, and we provide an assessment of whether the current governance and monitoring arrangements for in the public sector are consistent with this.

In Chapter 7, we consider the current literature on personnel management, particularly internal labour market tournaments, teams, and compensation, and some applications of this to the public sector.

In Chapter 8, we consider the literature on coordination and decentralisation, and show how the findings of this literature may assist analysis of the optimal scope and level of decentralisation in individual public-sector organisations and in the public sector as a whole.

Our conclusions are in Chapter 9.

2 Microeconomic foundations of state-sector reform in New Zealand after 1984

2.1 Introduction

Good public policy draws on ideas from a wide range of disciplines and practical experience, among which economics and management will naturally feature. Boston *et al* (1996:16) argued that ". . . one of the distinctive and most striking features of New Zealand's public management reforms was the way they were shaped by certain bodies of economic and administrative theory." In this chapter, we briefly consider those economic theories that were important in the formulation and implementation of New Zealand's public management framework after 1984.

2.2 The theoretical framework

The new public management (NPM) models developed in New Zealand in the late 1980s relied on the application of general management principles, new approaches to accounting (accrual accounting and output-based budgeting), and developments in the literature in economics (Scott and Gorringe, 1989). The key developments in economics that underlay NPM models were associated with institutional economics—a branch of economics concerned with considering the impact of institutional design on performance (Scott, 2001:26). In particular, the literature in institutional economics of the late 1970s and early 1980s was concerned with the theory of the firm, which at that time was built around ideas about transaction costs, principal-agent relationships, and information.

Since the issue was first raised by Ronald Coase in the 1930s, academic economists have developed increasingly sophisticated theories of why firms exist, why some economic activity is organised within the market and some is organised within firms, and how the efficient boundaries of firms are determined. While the majority of this work has been developed in relation to private-sector corporate forms, it has been applied to the public sector with sufficient frequency to make its applicability in this domain clear. In particular, public-sector organisations may be thought of as a type of not-for-profit firm generating policy advice and/or public services, and much of the thinking about the boundaries of the firm is also applicable to the analyses of the scope and organisation of the public sector.

The academic work on principal-agent relationships examined the relationship between those with authority or ownership rights (the principal), and those agents required to undertake activities that maximise value as it is defined by the principal. From this perspective, firms are viewed as mechanisms for reducing moral hazard and adverse selection problems by providing the principal of a firm with the ability to provide incentives to, and engage in active monitoring of, agents (for example, Jensen and Meckling, 1976). The importance of this work for the development of thinking about public management in New Zealand was outlined by Scott (2001:33):

> The path to greater efficiency and effectiveness often involves delegation of authority, within constraints, to lower levels within government. In these circumstances, agency theory is helpful in thinking about improvements. This theory is concerned with optimising the transactions involved between the superior and subordinate levels of public management when the goal is to allocate the decision rights to the party with the best capability, information and incentives to achieve performance goals.

The analysis of transaction costs, especially those relating to the governance of relationship-specific investments, was also central to the theory of the firm. Williamson (1979) argued that the transaction costs associated with writing, monitoring and enforcing an external contract determined whether activity was undertaken within the firm or through contracts with external providers. The more specific the asset, and the higher the transaction costs of contracting, the more efficient it would be to undertake activities within the firm. The analysis of transaction costs was central to the analysis of the costs and benefits of contracting out government services versus keeping them in government ownership. Horn (1995) adopts this approach in his treatise on public administration.

Where neoclassical economic models had assumed that information was freely available and costlessly acquired, economists began to consider the implications of building more realistic models in which information was costly to acquire and asymmetrically distributed. The availability of information was recognised as a critical part of decision-making, while at the same time the inevitability of parties not having access to the same information relating to a decision or negotiation was also formally incorporated into economic models.

An increasing academic interest in the role of competition in generating efficiency and in shaping the organisation of markets also influenced the theory of the firm in the 1980s. Recognition that the benefits of competition cannot be replicated by government monopoly, and that it is not possible for the private sector to compete against government departments that can hide costs or run at a loss, resulted in wide-ranging questions about the efficiency of state monopoly provision of goods and services. Where private-sector ownership of activities was feasible, the benefits of competition as a driver of both static and dynamic efficiency overtook traditional theories that had supported the creation of state monopolies, resulting in a wave of privatisation of entities and outsourcing of activities.[2]

The academic literature on which those developments were based would now be characterised as foundational, but unsophisticated. For example, the academic literature of the 1980s provided those interested in management applications with very little assistance in the consideration of:

- Approaches to addressing complex agency and information problems that could not be resolved by incentive and performance management schemes.

- The role of ownership in allocating decision rights, and the implications of this for efficiency in a world where all contracts are incomplete (that is, where not all contingencies relating to the performance of the contract are explicitly addressed).

- The difference between private and government approaches to the analysis of potential investment, particularly in relation to more complex investment issues such as information and investment timing.

- Complex combinations of private and public ownership, management and investment such as those that have been explored in the past two decades under the banner of public-private partnerships, mixed ownership models, private finance initiatives, and private or community delivery of publicly-funded services.

[2] It is noteworthy that the benefits of competition within the public sector (that is, between public entities), and the benefits of competition in government-funded activities, were not explicitly considered.

2.3 Putting the theory into practice

The Treasury (1987:37–39) set out a transaction-cost and incentive-based theory of the limitations of state ownership. It motivated the benefits of private ownership by drawing attention to the agency problems associated with information acquisition and performance management under state ownership given the complex objectives of state entities and the absence of market monitoring and competition.

Under New Zealand's new model of public management, the distinction between public and private has been conceived of primarily in terms of ownership, and in defining commercial and non-commercial activity. Activities were deemed to be in the public sector if they were government-owned and if they could not viably be provided on a commercial basis; if not, they were in the private sector.[3]

To clarify the interests of the state, the Public Finance Act 1989 established the distinction between the government's purchase interests (quantity, quality and price, or cost, of the services produced) and ownership interests (the human, intellectual and other assets and liabilities available to the department for the production of current and future services) within departments (Scott, 2001:17).

New Zealand's public-sector reforms were designed, wherever possible, to eliminate multiple accountability relationships, based on the idea that agents should never be accountable to more than one principal. Dual accountability was thought to place agents in the invidious position of having to balance the often conflicting demands and expectations of their respective principals, and exacerbate the problems faced by principals with regard to contract specification, monitoring and enforcement. Though in practice there were rarely neat ways to resolve these agency issues, New Zealand's public-sector reforms did attempt to clarify and order agency relationships, and provide a thorough analysis of the allocation of monitoring and decision rights for public-sector managers and organisations (Scott, 2001:34). This meant that in designing public-sector organisations there was a preference for separation of responsibilities, particularly a separation of policy and operations, and a separation of funder, purchaser and provider.

Reflecting the focus on agency problems in the academic literature of the time, the public management regime was designed to increase accountability by facilitating the monitoring of performance information for each public-sector organisation. The creation of a large number of public-sector organisations each with relatively homogeneous internal functions was designed to simplify the collection and interpretation of information on performance. It was expected that central agencies (Department of Prime Minister and Cabinet, State Services Commission, and the Treasury) would use performance information to improve resource allocation and hold chief executives accountable for organisational performance.

2.4 Concerns and unresolved questions in public management

New Zealand's public-sector reforms in the decade after 1984 established it at the forefront of international thinking about public management and the role of the state. That level of international recognition was well justified at the time, but after 25 years it is not

[3] Boston *et al* (1996: 8) suggest that whether this constitutes an adequate criterion for differentiation is a matter of debate. Other possible criteria include legal status, the receipt of public funds, the nature of the service being provided, and the public effects of an activity.

surprising that there is an increasing awareness of the challenges in public management that were not resolved by the reforms of that period. The continuing dilemmas include:

- The boundaries between the state and the private sector, including:
  - the case for public investment where the private sector is unwilling to invest, and
  - the allocation of ownership and service delivery between the private and public sectors.
- The place of competition in the public sector and, in particular:
  - the role of competition in promoting greater efficiency in the delivery of services and the management of assets within the public sector, and between the public and private sectors, and
  - the balance between competitive discovery of efficient solutions to operational and organisational problems, and single national approaches to investment and public-sector infrastructure.
- The need for stronger individual and organisational incentives for performance, and more effective mechanisms for the measurement and monitoring of that performance. Gill and Hitchener (2010:498) argue that while the vertical structures of accountability created under the Public Finance Act and the State Sector Act were designed to allow greater scrutiny of performance of ministers, chief executives and their departments or agencies, in practice there is relatively little use of performance information by central agencies, other than as a measure of bottom-line performance when things go wrong, and that this has tended to reinforce rather than mitigate the "well known bureaucratic pathologies of public organisations, in particular risk-averse, rule-driven behaviour."

- The effectiveness of the governance and management of the public sector as a whole, including the role of advisory and governance boards, the central monitoring agencies, and the challenge of producing more effective mechanisms for solving problems and developing innovative new approaches to policy where policy issues span the mandates of multiple teams and multiple government organisations. Scott *et al* (2010) point out that there have been consistent concerns about the ability of the public sector to deliver quality and innovative policy advice on the big issues that are of relevance to multiple departments and entities.

2.5 Conclusion

All ideas are shaped by the intellectual context in which they were formed, and New Zealand's post-1984 public management reforms were no exception. Those reforms were built on innovative applications of the academic literature of the time, and were successful in addressing many of the biggest problems associated with the public sector that had emerged from the earlier period of widespread government ownership and control of the economy. Many aspects of the performance of what remained within the public sector were improved by the reforms. But a wide range of challenges in public management and performance remains to be addressed, while some aspects of the post-1984 reforms appear not to have worked as anticipated or to have not worked at all. In the remainder of this paper we consider whether a range of new developments in the academic literature suggests opportunities for alternative and potentially superior approaches to addressing some of these issues in public management.

3 Incomplete contracts, ownership and the boundaries of the firm

3.1 Introduction

The modern literature on the theory of the firm began when Coase (1937) asked which transactions are more efficiently conducted in a firm than a market. Contractual incompleteness was at the core of the issues that Coase, Oliver Williamson (1979) and others explored, particularly as they were apparent in the distinction between the highly incomplete contracts within the firm and the necessarily more complete contracts with third-party providers of goods or services. While outsourced contracts with third parties are necessarily incomplete, the choice to outsource must reflect a greater value in the *ex ante* specification of aspects of the contractual relationship than is true for those activities conducted inside the firm.

Since the 1970s there has been a profusion of work on this question, but in many respects a unified theory of the firm is more rather than less elusive (Gibbons, 2005; Garrouste and Saussier, 2005). The reason is that a unified theory of the firm must be able to cope with external and internal organisational choices. In particular, a theory of the firm must be capable of providing a unified analysis of the costs and benefits of integration of activity within a firm.

Our focus in this chapter is on elaborating the "incomplete contracts" or "property rights" approach to the theory of the firm, which is currently the most popular paradigm for investigating the boundaries of the firm. But to provide some context for understanding why this approach is considered to be so important, we begin with a survey of the transaction-cost-based theories of the firm developed in the 1970s and 1980s. Applications of the incomplete contracts approach to a variety of private and public-sector ownership issues, including public-private partnerships, are reserved for subsequent sections of this paper.

3.2 Coordination and transaction costs

In the late 1970s and 1980s, much attention was focused on transaction costs as an explanation for the organisation of activity within or outside firms. This work focused on the firm as an entity able to govern transactions which markets could not provide. In particular, this work focused on the problems created by asset-specificity and the potential for opportunistic behaviour (hold-up) created when specific assets were owned by contractors independent of the firm. However, this form of organisation came at the expense of high costs (management costs) by comparison with the simpler incentive systems that could be used when sourcing non-specific inputs from the market (Williamson, 1979, 1999).

An extension of this transaction-costs approach focuses on the costs of conflict over the "appropriable quasi-rents" resulting from the activities of one or more firms (the costs of attempting to avoid such conflict being one form of transaction cost). Where these rents are large, conflict will be more intense, and the benefits of integration of activity within the firm will be large.

To the extent that transaction-cost approaches provided a theory of who owns the firm, it was developed within the "nexus of contracts" theory of the firm that underlay the work of Jensen and Meckling. Here the firm is "a legal fiction which serves as a focus for the complex process in which the conflicting objectives of individuals . . . are brought in equilibrium within a framework of contractual relationships" (1976:312). In this view, each party is fully protected by its contract within the nexus with the exception of shareholders, who accept a residual return because of their comparative advantage in diversifying risk, and residual control rights to guard against managerial appropriation of this residual return. But this view is inconsistent with the observation that contracts are never complete.

Despite its predominance in the literature at the time that the NPM models of the 1980s were being developed, transaction-cost approaches are now regarded as providing limited insight into the most interesting questions about the firm. Transaction-cost explanations for contractual incompleteness are unsatisfactory, because there is more incompleteness than can be accounted for by transaction costs (specifically, because there are many elements where contracting is not possible rather than just more costly than the alternative). Examples include situations where information is symmetric, but key contractible elements are not verifiable by either party. Even when transaction costs are zero, incomplete contracts may arise because parties cannot observe relevant economic variables, cannot verify those variables to a legal standard of proof, or prefer not to disclose information about themselves that would be required for a complete contract (Schwartz, 1998).

3.3 Incomplete contracts

In a paper that continues to have a profound impact on the literature on contracting and the theory of the firm, Grossman and Hart (1986) argued that transaction costs provided an incomplete theory of the nature and boundaries of the firm. Transaction costs could account for the benefits of integration of activities within a firm so long as the costs of monitoring by management were lower, but transaction costs provided only a partial approach to the analysis of the costs and incentive effects of different allocations of ownership and different configurations of the principal-agent relationship. Grossman and Hart argued that the boundaries of the firm were actually determined by the efficiency of allocating the residual rights of control to different parties rather than by transaction costs and management costs *per se*. So, while transaction costs provide a partial theory of the costs of ownership in different hands, a much broader conception of the costs and benefits of ownership was required to understand the boundaries of the firm.

The starting point for this approach to the theory of the firm is the incompleteness of contracts. Since humans are boundedly rational, not all issues of relevance to a contract can be anticipated at the time of writing the contract. Since information is costly to acquire, it may not be efficient to acquire some of the information that could be used to address contractual incompleteness even in situations where that information might be available. Further, not all future actions can be specified *ex ante* or verified *ex post*. Consequently, not all factors that are relevant to the creation of surplus in economic activity can or should be written into an enforceable contract. When contracts are incomplete, and consequently not all aspects of performance relevant to the contract are specified in advance, the contractor retains some discretion over those aspects of performance that are not explicitly identified in the contract.

Since their first identification of the importance of contractual incompleteness, developments in different parts of the economics literature have reinforced the importance of ownership as a means of solving contractual incompleteness. Management of risk often requires flexibility in the timing of investment and in the investment options chosen—a key component of the literature on real options which we consider in later chapters of this paper. In addition, modern understandings of the importance of dynamic efficiency focus on the fact that it is the process of decision-making under competitive pressure that is important for maximising the present value of social welfare over time. Neither the timing of investment nor the precise nature of competitive responses can be specific *ex ante*, suggesting that contractual incompleteness is both necessary and desirable in maximising dynamic efficiency.

Grossman and Hart (1986) identify two types of contracts: contracts that specify particular rights over the assets of another party and contracts that allocate residual rights. Ownership is the possession of those residual rights of control. "When it is too costly for one party to specify a long list of the particular rights it desires over the party's assets, it may be optimal for that party to purchase all the rights except those explicitly mentioned in the contract" (Grossman and Hart, 1986:692). From the perspective of this theory of the firm, the efficiency of different structures for the organisation of production will hinge on whether residual rights of control (ownership) are correctly allocated. For example, backward integration through the purchase of a supplier may be inefficient if the removal of residual rights of control over the production of the input distorts the decisions of the manager of the supplier in a way that reduces total surplus derived from the vertical supply chain. In this case, even if there were substantial transaction costs associated with the writing of a contract with the supplier, if non-contractible elements of supplier performance were critical in determining the contribution of the supplier to total surplus, the efficient allocation of ownership and production may require that residual ownership rights for supply be left with an independent supplier.

The analysis of Grossman and Hart (1986) focuses on incentives for investment (in any of assets, knowledge or effort) under different ownership structures. In particular, they observe that ownership confers the rights to the surplus generated by any investment. Thus, if ownership rights are conferred on an entity that does not have the ability to control the investments that have the greatest impact on the generation of surplus, then investments and total surplus will be below their optimal levels. The allocation of ownership rights, and thus the boundaries of the firm and of contractual relationships, will be determined by the value that is generated by allocating residual control rights to the parties whose non-contractible investments (effort) in production have the greatest impact on total surplus. From the perspective of this theory of the firm, the difference between employee and contractor relationships with a principal lies in the fact that the principal retains discretion over all non-contractible elements of the work, while the contractor retains residual rights of control over all those elements of the work not specified in the contract (Grossman and Hart, 1986:717). This example usefully illustrates the way in which the incomplete contracts theory of ownership extends earlier principal-agent and transaction cost frameworks for understanding the organisation of production.

Thus, in Grossman and Hart (1986) and Hart and Moore (1990) the firm is a collection of assets that are jointly owned. In the absence of complete contracts, ownership matters because it provides the right to make decisions on all unspecified contingencies. Rajan and Zingales (1998) extended this approach by proposing that the firm is a network of specific investments (including in human and physical capital) that cannot be replicated by the market.

3.4 The incomplete-contracts approach: An empirical example

To illustrate the approach of Grossman and Hart, consider the following elaboration of the numerical example provided by Aghion and Holden (2011:183 -184) in which there is a B(uyer) and S(eller) of an intermediate good that is used to produce a final good that B sells to consumers. Both parties can make private, non-contractible investments[4] relevant to this production process:

- S can make an investment costing $5 that will make the intermediate good cheaper to produce. This investment will reduce the cost of the intermediate good from $16 to $10.

- B can make an investment costing $5 that will make the final good more valuable for consumers. This investment will increase the value of the final good to consumers from $32 to $40.

Table 1: Incomplete contracts and ownership – A numerical example

(i) Impact of investment on cost and value to customers

	Pre-investment value		Investment cost	Post-investment value	
	S	B		S (input)	B (final product)
Production cost of intermediate good	16		5	10	
Value to customers of final good		32	5		40

(ii) Surplus generated by different forms of ownership

	Investment		Sale price	Cost	Total surplus
	S	B			
Incomplete contracts – ownership assigned to B	0	5	40	16	19 (40 – 0 – 5 – 16 = 19)
Incomplete contracts – ownership assigned to S	5	0	32	10	17 (32 – 10 – 5 – 0 = 17)
Incomplete contracts, S and B share surplus 50:50	0	0	32	16	16 (32 – 16 - 0 – 0 = 16)
Complete contracts	5	5	40	10	20 (40 – 10 - 5 – 5 = 20)

The results of different allocations of residual control rights for this production problem are summarised in Table 1. If contracts were complete, B and S would write a contract that specified both types of investment *ex ante*, with the result that the total surplus generated would be $20 (sale price of $40, less investment cost of $5 for S and $5 for B, less the

[4] Private and non-contractible means that only S can make the cost-reducing investment in production of the intermediate good and only B can make the value-enhancing investment in value to consumers.

cost of the intermediate good of $10). However, where the investments are non-contractible there are three possibilities:

- S and B continue as separate firms. Contractual bargaining about the price that B pays to S results in an agreement to split the surplus 50:50. In this case, if B invests but S does not, B will bear a private cost of $5 but obtains half of the increase in surplus generated = ($40 - $32)/2 = $8/2 = $4. B therefore will not invest since $4 is less than the investment cost. Similarly, if S invests but B does not, it will bear a private cost of $5 and obtain half of the increase in surplus of ($16 - $10)/2 = $6/2 = $3. S will therefore not invest, since $3 is less than the investment cost. When neither party invests, we have the base case for incomplete contracts where total surplus is $16.

- S and B are integrated with S acquiring residual control rights (ownership) over the entire project. In this case, the investment by S will be made, but the investment by B will not, resulting in a total surplus of $17 (Table 1) which is superior to separation of S and B.

- S and B are integrated with B acquiring residual control rights (ownership) over the entire project. In this case, the investment by B will be made, but the investment by S will not, resulting in a total surplus of $19 (Table 1) which is both superior to separation of S and B and superior to residual control rights being allocated to S.

In this example, incomplete contracts mean that production will most efficiently take place in an integrated firm owned by B.

The incomplete contracting perspective embodied in this example represents a sharp break with the earlier transaction cost-based literature on the firm. Whereas incomplete contracts imply that inefficiencies arise because it was hard to foresee and contract about the uncertain future, earlier literature tended to take a “complete contracts” perspective in which imperfections arise as a result of moral hazard and asymmetric information. The shortcoming in the latter approach is in its lack of attention to the impact that ownership and other aspects of organisational form have on the efficiency of the allocation of residual decision rights: these aspects of ownership were subsumed to the focus on the moral hazard and asymmetric information problems faced by all owners employing professional managers. Relying on transaction costs alone to generate differences in organisational form under this complete contracts approach appears highly unsatisfactory. In contrast, ownership matters where contracts are incomplete, because the rights of the owner provide the power to take decisions in relation to those matters not explicitly covered in contracts covering the use of assets, supply of inputs or service to customers (Hart, 2003:296).

In subsequent chapters of this paper, we will explore the wide range of issues to which this framework is applicable. Appreciation of the breadth of the potential applications flows from the fact that it may be applicable whenever relationship specific investments are important in the determination of economic efficiency. In its most general form, it implies that the party whose non-contractible effort, knowledge, or investment makes the greatest contribution to total surplus should have residual control rights. Further, it provides a framework for thinking about both the boundaries of state-sector ownership of assets and service delivery, but also about the potential to redefine the boundaries of state ownership of activities to allocate residual control rights for particular tasks to public

or private entities depending on the value added by different ownership structures for the activity. Each of these lines of inquiry is explored below.

3.5 Linking transaction cost- and incentive-based theories of the firm with incomplete contracts

Understanding the contribution that Grossman and Hart (1986) and Hart and Moore (1990) made to the theory of the firm requires an explanation of its relationship to the other approaches to the definition of the firm used by economists in the past 40 years. Here we consider explicitly incentive-based theories of the firm, and transaction cost-based theories of the firm.[5]

Incentive-based theories of the firm

Incentive-based theories of the firm have their foundation in the analysis of the incentive problem between a principal and an agent. This approach assumes that there are many tasks and many instruments associated with the agency problems in a firm, and asset ownership is merely one of the instruments. Papers in this paradigm consider two ways to structure the agency problem: (i) where the agent does not own the asset (is an employee) and therefore has incentives provided by being paid on measured performance, and (ii) where the agent does own the asset (is an independent contractor) and receives both a payment based on measured performance and the value of the asset after production occurs.

This approach to the theory of the firm has in effect focused on the claimed distinction between the low-powered incentives associated with employment, and the high-powered incentives associated with contracting. Employees require low-powered incentives because they are not distracted by the contractor's incentives to increase the value of the assets used for production. More generally, joint optimisation over asset ownership and contract parameters determines whether to conduct activity within the firm or outside.

The incentive-system theory of the firm is therefore related to the incomplete contracts literature, both in its use of ownership as an instrument and in its ability to provide a unified account of the costs and benefits of integration. But this literature has one major shortcoming by comparison with the incomplete contracts literature: who owns the asset affects the payoff of agents, but not the behaviour of the agents. This means that incentive-system theories do not easily capture the ability of agents to respond to the same set of incentives with two different types of behaviour, only one of which is optimal for the principal.

Transaction cost-based theories of the firm

Both the transaction cost-based approach of Williamson (1979) and the incomplete contracts approach of Grossman and Hart (1986) and Hart and Moore (1990) rely on the potential for relationship-specific investments to create hold-up. Investment in relationship-specific assets generates *ex post* quasi-rents, and any contingency not explicitly specified within a contract will create opportunities for bargaining about the division of those rents. Williamson focuses on the way in which parties to a contract attempting to capture (*ex post*) a greater share of these quasi-rents generated transaction

[5] The authors acknowledge the contribution to this section made by the referee report of Professor Ig Horstmann, Rotman School of Management, University of Toronto.

costs, whereas Grossman, Hart and Moore focus on the potential for this incompleteness to affect the incentives to make the initial investment in the specific asset. Williamson focuses on the ability of firms to avoid these transaction costs associated with *ex post* bargaining by bringing activities within the firm, while Grossman, Hart and Moore focus on the way in which allocation of control (ownership) rights to the party that is able to add the most value to the contract produces the best investment outcome.

The incomplete contracts approach adopted in this paper may therefore be thought of as an alternative lens on same issues that motivated the work of Coase and Williamson. The two approaches are, at least to some extent complementary, but they may also lead to quite different conclusions about the optimal structure and boundaries of a firm. The reason is that the focus of Grossman, Hart and Moore is on the potential for allocations of ownership rights to overcome the problems created by *ex post* bargaining about quasi-rents. Their focus on the ownership of activity, and its impact on the *ex ante* efficiency of the investment decision has two advantages over the transaction cost approach: it is consistent with competition between different potential owners of an activity rather than just focussing on the "inside the firm versus outside the firm" decision of one firm, and it has more direct application to the contemporary literature on investment decision-making, public ownership and governance than was the earlier transaction-cost approach. Those applications are the subject of the remainder of this paper.

3.6 Conclusion

Like the principal-agent and transaction-cost frameworks that preceded it, the theory of incomplete contracts provides a framework for thinking about efficient approaches to organisational structures and decision-making. As it has been elaborated in the literature over the past two decades, the theory of incomplete contracts explains ownership as the allocation of decision rights, explains decision rights as critical in the context of contractual incompleteness (which includes the inability to write complete incentive contracts) and explains the boundaries of the firm as determined by the efficient scope of *ex ante* decision rights, not by transaction costs created by *ex post* bargaining about the allocation of returns from specific investment. The approach lends itself well to comparative analysis of institutions and decision-making frameworks, but is less easily linked to empirical tests, and does not provide simple off-the-shelf solutions to complex questions of policy and management.

4 Real options and investment

4.1 Introduction

Contemporary microeconomic theory reflects the progress that has been made in incorporating time, expectations and uncertainty into more realistic models of decision making. In the last 20 years, the theory of investment decision-making and our understanding of the costs and benefits of investment have been revolutionised by analysis of investment timing, the value of the flexibility to delay investment, and the cost that is incurred by giving up flexibility at the point where the decision to invest is actually made. This broad class of problems in investment is addressed by the literature on real options (Guthrie, 2009a).

An important source of real option values derives from firms' flexibility regarding the timing of investment, since this allows them to "wait and see" how economic conditions evolve, for example, before committing to large irreversible investments. While real options are commonly discussed in the context of irreversible risky investments in network industries, the concept is applicable to a very wide range of situations, including those in workably competitive markets. Real options analysis may assist understanding of a wide range of choices that relate to irrecoverable expenditure.

A key insight of the literature on real options is that the total economic cost of an individual project is not just the capital expenditure involved, but also includes the reduction in the value of the firm's growth options due to investment. This manifests itself in decision-making in two equivalent ways: (i) investment is optimal only when the value of the completed project exceeds the required capital expenditure by at least the amount of the reduction in the value of the growth-option; and (ii) investment is optimal only when the internal rate of return from investment exceeds the project's weighted average cost of capital by some strictly positive premium.

While there are few explicit links between them in the published literature, real options may be understood as a key element of contractual incompleteness. For example, large projects which require and receive approval may still require the exercise of management discretion about the optimal time to invest if the return on the investment is to be maximised. In other words, and as we explain below, since the optimal timing of investment cannot be specified *ex ante,* the decision on exactly when to invest may be viewed as a component of contractual incompleteness. The allocation of decision rights in respect of investment timing should therefore be thought of as a response to contractual incompleteness.

In this chapter we provide a survey of the literature on real options, and consider its implications for public-sector ownership and investment. Subsequent sections provide further examples of the relevance of this approach for the choice between private and public ownership.

4.2 The value of flexibility in decision-making

A standard cost-benefit approach to project evaluation recognises that capital devoted to one project cannot be used for other projects: that opportunity cost of capital is factored into the calculation. And the cost of capital highlights starkly that a benefit 10 years ahead is much less valuable than the same-sized benefit that can be achieved a year from now.

However, an important limitation of standard cost-benefit analysis is that it is static. To the extent that it incorporates uncertainty about the future, it will do this by valuing different scenarios, but it contains no formal mechanism for incorporating the implications of the uncertainty that is reflected in those scenarios into the analysis.

In contrast to standard cost-benefit analysis, a prudent investor will incorporate into their investment analysis the fact that uncertainty about the future is pervasive, and that at any point in time the investor has the option to act today, or to delay the action until a future time and wait and see what happens in the meantime. Where uncertainty is pervasive, flexibility will be valuable. Much of the value of flexibility stems from the ability to avoid making large irreversible investments in projects that turn out to be not worthwhile.

When there is substantial uncertainty surrounding the outcome, an alternative to acting immediately can be to delay acting and instead invest resources in learning more about the outcome of particular actions.[6] For example, a roading project can be delayed while research is conducted to reduce the uncertainty surrounding the cost of building the road and/or the uncertainty surrounding demand for the road. If construction begins immediately, then the ability to exercise this so-called learning option is destroyed.

If the present value of benefits from a new road equals B and the present value of the required expenditure equals I, then the net payoff from building the road equals B - I. If the option value of waiting and learning more about the cost of construction equals W, then the payoff from waiting equals W. It is optimal to build the road immediately if and only if B - I is greater than W. Equivalently, B must be greater than I + W.[7] That is, immediate construction is optimal if and only if the benefits of the project, B, are greater than the total cost, I + W. The key insight here is that the cost of constructing the road immediately is the sum of the (present value of) the required capital expenditure and the learning option that is destroyed upon investment. In order for decision-making to result in an optimal allocation of resources, the cost of projects needs to incorporate the full opportunity cost—in this case, capital expenditure *and* the value of learning options destroyed by investment.

If a firm delays investment and then receives new information that indicates it would have regretted earlier investment (perhaps because demand is insufficient or new technology supporting a different approach to investment has emerged), then the decision to delay has allowed the firm to avoid incurring wasteful expenditure. In contrast, if it delays investment and then receives new information that indicates it would not have regretted earlier investment, then it still has the option to invest. That is, delaying investment does not prevent the firm from undertaking good investments, but it helps it to avoid making bad ones. This is an established result in investment theory—the so-called "bad news principle."[8] The bad news principle requires that decisions be wholly or largely irreversible (such as in infrastructure), and that there be uncertainty about the future path of costs, technology or demand for the investment.

If the option value of waiting is ignored, irreversible capital investment proposals will consistently underestimate the costs of investment. The private sector incorporates in its

---

6 For example, Flyvbjerg *et al* (2005) estimate that the forecast error for actual traffic using large roading projects in the first year after construction is approximately 40 percent of the forecast level. Flyvbjerg *et al* (2002)) report that the forecast error for the construction cost of similar projects is approximately 30 percent of the forecast level.

7 It is possible that an investment may create options, in which case W is negative.

8 See, for example, Bernanke (1983).

analysis of investment the cost of real options destroyed (either through a formal analysis of the real options or "rules of thumb" such as the use of a higher cost of capital in the analysis), but ignoring the value of flexibility will also lead officials consistently to recommend public-sector investment when private-sector investors would wait.

4.3 Sequential investment

There will often be an intermediate route between investing immediately and delaying all investment in order to learn more about the payoff. For example, it might be possible to undertake limited investment in the meantime, while learning more about the project payoff before undertaking the rest of the investment. Indeed, the outcome of the partial investment might reveal information about the desirability of the full project. For example, the early stages of a gradual rollout of ultra-fast broadband will reveal information about the level of demand for the services it provides. If this demand is high, then the full network can be built, with users experiencing slightly delayed delivery. In contrast, if the demand revealed by the early stages is low, then the rollout of the remainder of the network can be abandoned or delayed until demand rises, allowing the government to avoid (or delay) uneconomic investment. However, investing in stages has a cost that may outweigh these benefits. Specifically, building a large network initially will allow it to fully exploit economies of scale in network construction. The decision to build the network in stages or all at once must trade off the benefits of quick, low-cost construction against the cost of investing without first gathering potentially valuable information.

As in the case considered above, for decision making to result in an optimal allocation of resources, the cost of projects needs to incorporate the full opportunity cost. In this case, the full opportunity cost includes the incremental capital expenditure plus the value of the learning option destroyed by investment.

The value of a learning option destroyed by investment depends on the amount of information that can be learned if the investment is delayed (more precisely, the amount by which uncertainty can be reduced by further investigations). Real options analysis provides a rigorous approach to estimating the value of learning options. Various models of learning options appear in the academic literature and have been used to identify the sources of option value and the implications for resource allocation.[9] These can be used to evaluate staged investment—allowing learning to occur before the full resources are committed to a project—against alternative policies.

Phased investment has another potential benefit, which is subtly different from the one just considered. As in the case just discussed, there is an option value in delaying (or slowing down) investment. In this case, however, the source of the option value is the flexibility to respond to different market conditions. For example, by conducting staged investment a decision-maker can suspend construction, if the price of inputs rises, if recession dampens demand, or if new technology supporting a different approach to investment has emerged, and accelerate it, if input costs fall, if demand rises, or if a new technology does not emerge. The difference is more than one of semantics. In the situation considered previously, the decision-maker could use the opportunity to delay to learn more about a project: if nothing was done while waiting, then nothing would change while waiting. Here, there can be value in the option to simply wait and do nothing and wait for market conditions to improve.

[9] See, for example, Epstein *et al* (1999), Childs *et al* (2001, 2002a, b), and Guthrie (2007).

The payoff to typical infrastructure investment is sensitive to variations in market conditions, whether it be fluctuations in demand for ultra-fast broadband or longer-term variation in electricity carried over the national grid. The cost of building such infrastructure will typically exhibit economies of scale. In such circumstances, investment involves a trade-off between exploiting the economies of scale by investing in a small number of large expansions and retaining flexibility by investing in small increments as and when they are needed so as to avoid unplanned investment in excess capacity.

Greater volatility means that investment should involve larger, relatively infrequent, steps.[10] By raising the value of delaying investment, greater volatility motivates the decision-maker to set a more demanding investment test. This in turn changes the trade-off between scale and flexibility, because it reduces the costs of "overinvesting" in additional capacity—that is, even if demand growth is lower than expected after expansion, the fact that the investment threshold is so high at the time of investment means that the decision-maker may still be able to recover the capital expenditure incurred. The optimal policy is to take greater advantage of economies of scale and raise the scale of each expansion of capacity.

Volatility can vary between industries, so that the appropriate scale and timing of the ultra-fast broad band roll-out, for example, can differ from the rate at which the national electricity grid should be upgraded, which can differ from the rate of investment in the roading network. Furthermore, as we will see below, the nature of contractual arrangements between the funding authority and the party undertaking the investment influences the allocation of demand risk, which affects volatility. Thus, the precise form of a contract can affect the scale and timing of investment that occurs as a result of the contractual arrangement.[11]

4.4 Quantifying the value of flexibility

There is now an extensive academic and practitioner literature setting out rigorous methods for estimating the value of flexibility and using the estimated values to inform decision-making.[12] The existing literature mainly, but not exclusively, looks at this from a corporate perspective, so that the decision-maker is a manager who maximizes the present value of the flow of surplus received by a firm's owners. As we explain below, this approach needs to be modified for public-sector project evaluation, where the ultimate objective is to maximize the present value of the flow of total surplus.

In principle, real options analysis is implemented by calculating the present value of the flow of a suitable "net benefit function" for each possible policy that can be adopted by the decision-maker. Of course, there are a very large number of possible policies in all but the simplest multi-decision problems, so in practice dynamic programming is used instead. This technique breaks a complicated multi-period optimization problem into a sequence of simpler two-period optimization problems. Provided this technique is implemented correctly, it yields exactly the same outcome as the more primitive approach of evaluating each possible policy in turn. Furthermore, in many applications the information requirements are not much more stringent than static cost-benefit analysis

10 This result is proven in a corporate setting in Guthrie (2011), and also analyzed by Kort *et al* (2010).

11 For example, Evans and Guthrie (2011) show that the level of price cap imposed on a regulated infrastructure provider alters the trade-off between scale and flexibility, and that the level of cap can have a material impact on overall welfare.

12 See, for example, Dixit and Pindyck (1994), Copeland and Antikarov (2003), and Guthrie (2009a).

even though cost-benefit analysis has real limitations in considering risk by comparison with real options analysis.[13]

Even in straightforward commercial operations—where the costs might be clear, and the interests of the owner are also clear—cost-benefit analysis is not always easy. In public-sector projects where direct price signals are often unavailable, the challenges are even greater. But accepting that such analysis is difficult does not change the importance of doing it well, including testing and evaluating credible alternative scenarios, and exposing the analysis and assumptions as far as possible to public scrutiny.

4.5 Real options and public-sector investment decisions

Real options analysis was developed to analyze the investment behaviour of value-maximizing private-sector firms. However, in principle, it is easily modified to analyze investment (and other) behaviour by the public sector. Whereas a private-sector firm maximizes its market value, which is taken to be the present value of the flow of surplus to the firm's owners, a public-sector organisation can be charged with maximizing the present value of the flow of total surplus. The mechanics of real options analysis are unchanged: the flow of producer surplus is simply replaced with the flow of total surplus.[14]

In a static setting, this will mean that a private-sector entity will tend to under-invest relative to the public sector. This will happen whenever the required investment expenditure is greater than the present value of the flow of that part of the surplus received by the entity's owners (so that the private-sector entity would not invest) but less than the present value of the flow of total surplus (so that the public sector would invest). A similar result holds when there is timing flexibility, except now instead of "under-investment" the problem is "investing too late." That is, the investment policy that maximizes the present value of the public investment payoff will feature earlier investment than the policy that maximizes the present value of the private payoff.

The situation becomes more complex when the public and private sectors interact. For example, infrastructure might be owned and operated by a regulated monopolist, in which case some decisions are made in the public sector (e.g. the level at which prices are capped) and other decisions are made by the firm (e.g. the timing and scale of investment and the level of spending on maintenance). Even if infrastructure is owned and operated by a public-sector organisation, particular tasks (and the associated decisions) might be outsourced to the private sector. Real options analysis can still be applied in such situations, but now at each decision node the decision-maker chooses an action that maximizes its own objective function shaped by the allocation of ownership. Thus, decisions that are the responsibility of the public-sector organisation are made in order to maximize the present value of total surplus, and decisions that are the responsibility of the private-sector organisation are made in order to maximize the present value of the surplus received by its owners: in each case the actual objectives will be shaped by contracted terms and the allocation of ownership.[15]

---

13 For example, Guthrie (2009b) demonstrates the real options analysis of a multi-stage commercial real estate project that uses only the inputs into a static discounted cash flow analysis of the problem, plus an estimate of the volatility of property prices.

14 For example, Pennings (2004) uses this approach to determine socially-optimal policies for investing in quality improvements.

15 For example, Turnbull (2010) uses this approach to evaluate the development incentives of private-sector landowners when the government can seize land needed to provide a public good. A similar procedure is used in the corporate finance literature to examine the effects of the conflict of interest between managers, who decide the timing of investment by firms, and

Allowance for flexible decision-making (as occurs with real options analysis) is especially important in situations where the private and public sectors interact. The incomplete nature of the contracts involved means that both parties retain some flexibility in how they carry out their assigned tasks. For example, while a private contractor may be required to complete construction of a project within a particular time frame, crucial decisions during the construction process are delegated to it. It will have some freedom regarding the rate at which particular stages of the project are completed, the materials used, the quality of the work done, and so on.

The actions taken by public and private-sector decision-makers will depend on the value of flexibility, such as the option value of delaying taking an action. These option values depend on the volatility of the decision-maker's payoff, and the level of this volatility will depend on how risk is allocated. That is, when evaluating projects in which the public and private sector interact, we need to be concerned not just about project risk, but also about how that risk is allocated. Below we consider three different types of interaction between the public and private sectors through the lens of real options analysis.

Risk aversion

Ultimately the actions undertaken by organisations will be determined by the decisions made by individuals (or groups of individuals) within those organisations. For example, while the objective function of a private-sector firm may be the present value of the flow of the surplus received by its owners, the decision maker(s) will maximize their own utility function. Similarly, while a public-sector organisation may have the present value of the flow of total surplus as its objective function, the decision-makers—politicians or bureaucrats—will maximize their own utility function. The overarching objective of private-sector managers could be argued to be narrower and thus admit clearer incentives for managers. In contrast, in public-sector organisations the combination of multiple goals and risk associated with political as well as financial capital means that the risk considered in relation to public-sector investment will be different, and possibly also that public-sector decision-makers will be more risk averse than private-sector decision-makers.

Private-sector managers can be incentivized by basing their compensation on project outcomes. The corporate finance literature makes this assumption, and treats the individual's utility function as a concave function of this compensation (or of the consumption that it can purchase) when it examines the effect of risk aversion on investment behaviour. This approach is more problematic when applied to public-sector decision-makers since such explicit sharing rules are ruled out for public-sector agents. However, the lessons of the corporate finance literature can be applied, if public-sector managers are risk averse regarding project outcomes: for example, they might perceive that their career prospects will suffer a relatively large negative shock from a worse-than-expected project outcome and a relatively small positive shock from a corresponding better-than-expected project.

All else equal, greater risk aversion lowers the present value of a risky flow of future surplus. In a static setting (that is, one in which investment is "now-or-never"), this means that some risky investments with the potential to provide quality-enhancing and/or cost-reducing innovations will not be approved by a public entity even though they would maximize the present value of the flow of total surplus. For such projects, the public-sector decision-maker attaches a (personal) value to the flow of benefits that is less than

shareholders, who ultimately bear the consequences of the managers' decisions. See, for example, Grenadier and Weng (2005), Hugonnier and Morellec (2007), and Shibata (2009).

the required expenditure, even though the present value of the flow of total surplus is greater than the (same) required expenditure. Similarly, there will be situations in which investments will not be approved by a private entity even though they would maximize the present value of the flow of surplus to the entity's owners.

However, this result can change when there is flexibility regarding the timing of investment. In this case greater risk aversion has been shown to lead to accelerated investment.[16] This happens because, although greater risk aversion reduces the payoff from immediate investment, it reduces the value of the option to wait by a larger amount. The decision-maker is able to reduce its exposure to risk by investing immediately, since this eliminates the possibility that conditions will change while it waits to invest: post-investment uncertainty remains, but pre-investment uncertainty has been eliminated.

A similar result holds when investment involves exercising learning options. Specifically, when the cost of completing a project is uncertain and this uncertainty is only resolved by completing construction of the project, Whalley (2010) shows that greater risk aversion leads to earlier abandonment of the project.

It follows that if a manager is risk averse with respect to his or her employing organisation's investment payoff, then investment will generally start too soon, and be abandoned too quickly, to maximize the present value of that payoff.

Financing constraints

Standard capital budgeting theory assumes that there are no frictions involved in raising the financial capital needed to undertake investment. This assumption is reasonable at government level and for private-sector firms with strong credit ratings. However, public-sector entities with restrictions on the transfer of unspent funds from one financial year to the next and private-sector firms carrying high debt loads, may find it unusually costly (or even impossible) to raise the capital needed to undertake projects, even if those projects have a positive net payoff. In such situations, the standard results from capital budgeting theory need to be modified.

In a static setting, the presence of frictions in external capital markets can lead to underinvestment. However, when there is flexibility regarding the timing of investment, such frictions can actually lead to accelerated investment.[17] The intuition for this result is that delaying investment exposes the entity to the risk that when it does eventually want to invest in the project it might not be able to access sufficient external capital to make up for any shortfall of capital from internal sources. This risk lowers the value of waiting and, all else equal, means that entities with temporal funding constraints such as those in the public-sector budget process may invest too soon to maximize overall welfare.

Limited liability

The interests of bondholders and shareholders in a private-sector firm diverge because shareholders are exposed to both the upside risk and the downside risk of investment, but due to limited liability their downside risk is capped. In contrast, bondholders have no exposure to upside risk (their return is fixed) but they bear a share of the downside risk (where losses exceed the capital of the firm). This conflict can lead to asset substitution (where a manager working in shareholders' best interests adopts high-risk projects,

16 See, for example, Henderson (2007), Hugonnier and Morellec (2007), and Miao and Wang (2007).

17 See Boyle and Guthrie (2003).

transferring wealth from bondholders to shareholders) and debt overhang (where a manager working in shareholders' best interests forgoes value-increasing investments that transfer wealth from shareholders to bondholders).

When there is flexibility regarding the timing of investment, debt financing induces the manager to accelerate investment relative to the investment policy that maximizes the value of the firm as a whole (that is, the sum of the values of debt and equity).[18] This allows shareholders to start receiving dividends sooner, but exposes the firm to the risk of poor future outcomes—which may impose losses on bondholders.[19] The New Zealand government may not in practice have limited liability in respect of the financing of individual projects (it is unlikely to be politically feasible to leave private bondholders with losses after it had induced them to invest in a public project), so it does not have the incentive to accelerate investment because some of the losses will be carried by bondholders. However, soft budget constraints; (the power to tax to raise funds to subsidise projects) may induce early investment by providing a softer constraint on early investment than is true for private-sector firms.

4.6 Example: Public and private investment in state housing

Recent analysis of the provision of low-income housing in New Zealand highlights the importance of increasing housing supply (Housing Shareholders Advisory Group, 2010). Real options analysis can help understand the challenges in increasing supply and the issues that would be confronted, if private investment in state housing were to be considered.

The supply of housing is dependent on the rate at which land is developed (whether it be rural land developed into housing or existing urban land redeveloped to provide newer, and perhaps higher density, housing). All the ingredients for valuable real options are present in the land-development decision. First, the owner of a block of land has flexibility over the date at which development occurs: the existence of pieces of vacant land in city centres demonstrates the importance of the real option to delay development. Second, the payoff from delaying development is uncertain: when the housing market was booming, landowners could (and did) delay development and receive a much high development payoff, while others delayed past the market peak and would have received a much lower payoff had they gone ahead with development. Third, developing land is costly to reverse.[20]

Any landowner deciding whether or not to develop land for housing should calculate the payoff from developing the land immediately and compare it with the payoff from delaying the development decision. The development payoff equals the amount by which the

18 See Mauer and Sarkar (2005).

19 However, the acceleration of investment is reduced, if debt can be renegotiated in the event that the firm becomes financially distressed, and the reduction is greatest when the shareholders' bargaining power is strongest. See Sundaresan and Wang (2007).

20 The landowner is able to sell the developed land, but this does not reverse the development decision (and allow the landowner to recover the development expenditure). The sale price will equal the present value of the rental income from the developed land and in the absence of other frictions, the landowner will be indifferent between holding onto the land (and receiving a flow of rental income) or selling it (and receiving an equivalent lump sum). If the housing market declines after development, the ability to sell the land at the market-clearing price does not protect the landowner. See Pindyck (2007, Section 6.3) for a discussion of this issue in the context of the unbundling of telecommunications networks.

present value of the flow of net benefits exceeds the required development expenditure, with the definition of the relevant net benefits depending on the identity of the owner. For a private-sector owner, the net benefits will equal the rental income generated by the property, plus any subsidies provided by public-sector agencies, minus ongoing expenses such as insurance, maintenance, local authority rates, and so on. For a public-sector owner, the benefits will be broader, reflecting the value to society as a whole in addition to those of the house's occupants. The payoff from delaying the development decision equals the option value of waiting, and—as it depends on the distribution of the eventual payoff from future development—also depends on the identity of the owner.

Equilibrium in the markets for developed and undeveloped land makes the development decision relatively straightforward when the decision-maker is a private-sector landowner.[21] The present value of the flow of net benefits received by the owner equals the market-clearing price of developed land, and the option value of waiting equals the market-clearing price of undeveloped land. Therefore, the landowner develops his land as soon as the market price of developed land exceeds the market price of undeveloped land by an amount that is greater than or equal to the required development expenditure. That is, the equilibrium process does the job of calculating present values for the landowner.

In contrast, public-sector landowners must estimate the public benefits of housing, calculate the present value of the flow of public benefits from developed land directly, and calculate the option value of leaving the land undeveloped. In practice, this is a sufficiently complicated problem requiring real options analysis that, even if their aim is to maximize overall welfare, public-sector decision-makers will resort to simpler—and therefore potentially suboptimal—decision-making rules. For example, Housing Shareholders Advisory Group (2010:40) highlights some ways in which an expectation that HNZC meets a specific target number of state houses has distorted its decision making. Politically-directed investment may take a broader view of the payoff from a project via the ability to "capture" consumer surplus in political benefits, but public-investment decisions may also reflect narrower objectives such as votes in particular locations or with particular constituencies.

From the discussion above, provision of low-income housing by private-sector entities with strong balance sheets will likely lead to more efficient use of timing flexibility being made. The decision-makers in such entities are likely to be less risk averse than their counterparts in the public sector, and so will be more willing to delay investment until market conditions improve and/or more information about the location and strength of demand for low-income housing is known. A strong balance sheet will reduce the risk that insufficient capital will be available at the time of investment, if that investment is delayed, which will also encourage efficient use of investment timing flexibility. Taking advantage of this flexibility allows the provider to reduce the risk that new houses will be stranded in the future by changing market conditions.

The literature on regulatory economics shows that welfare can be improved by capping the price that the firm can charge its customers and that the best price cap outperforms the best revenue cap.[22] Since this will lead to some demand not being met, the price

21 See Guthrie (2010) for an equilibrium model of land values and development decisions that incorporates stochastic demand, development irreversibility, and competition between property developers.

22 See, for example, Dobbs (2004), Evans and Guthrie (2011), Hausman (1997), Hori and Mizuno (2006), Pindyck (2007), and Roques and Savva (2009).

regulation should be accompanied by quantity regulation, which puts a cap on the amount of rationing that can occur. The price levels of the price- and rationing-caps depend on factors such as the risk of stranding and the extent of economies of scale in investment. When there are economies of scale, the government faces a dynamic-inconsistency problem, because once the firm has undertaken its initial investment, overall welfare could be increased by raising the price cap and lowering the rationing cap. Thus, regulatory arrangements should make it possible for the regulator to commit to pre-investment settings. In the limiting case with constant returns to scale, the dynamic inconsistency disappears, but it is still socially optimal to impose both a cap on the price that the firm can charge and a cap on the extent to which it can ration demand.[23]

A private provider of subsidised low-income housing would bear some similarities to a regulated natural monopolist infrastructure-provider with the option to make a series of investments that generate services of value to the public. The two situations have much in common:

- Much of the investment decision-making is delegated to the firm, which has flexibility regarding when it invests and how large its investments are.
- Investment is largely irreversible.
- There is uncertainty regarding the future levels of revenue that will result from those investments.
- The firm is exposed to the risk that some or all of its investment will be stranded
- The regulation/subsidy regime must be designed in a way that avoids creating perverse incentives, such as "gold plating" the housing stock (so that the base on which any rate–of-return floor is calculated is higher) or building infrastructure/houses where they are cheap to build rather than where they are needed, knowing that a minimum level of revenue will be received anyway.

The main difference between the two situations is that in the case of price regulation welfare can be improved by placing some restriction on the upside of the firm's revenue, whereas in the housing case it can be improved by placing some restriction on the downside of the firm's revenue.

As far as we are aware, corresponding analysis for the subsidized provision of low-income housing has not been undertaken, but our conjecture is that the regulatory example is likely to have direct relevance. It seems likely that the best subsidy would place a floor on the rent that the firm could charge for each house (or perhaps each occupant) rather than a floor on its overall revenue or rate of return, and that some minimum quantity requirement on the private housing provider would be necessary to maximize welfare.[24]

---

23 The constant-returns-to-scale case is of limited interest in the regulatory setting, since investment returns to scale are typical for the types of firms that are regulated. However, it is more relevant in the case of housing provision because in some locations the amount of low-income housing required will be sufficiently small that few economies of scale are to be expected.

24 The precise form of the socially-optimal subsidy regime would also depend on whether there are any economies of scale in providing low-income housing (and if there are, then on the degree of economies of scale).

4.7 Conclusion

Decisions to undertake investment projects will often be incomplete in the sense that different aspects of the implementation, including the timing of the investment and the phasing of the investment, will be at the discretion of the party implementing the investment rather than the party making the original decision. The literature on real options represents a substantial advance on the earlier literature on cost-benefit analysis and the weighted average cost of capital. Perhaps more importantly, the growing literature on real options allows more sophisticated analysis of complex investment scenarios, including those where ownership interests or responsibilities for implementation are shared between the private sector and the public sector. We have shown that while the differences in the objectives and constraints between public-sector and private-sector investors may be material, these may be incorporated into real option analysis in a straightforward way.

Real-option analysis has provided formal methods for capturing the differences between investments that involve large fixed and sunk infrastructure projects, and those where the scale, location and use of the investment are more flexible. Further, real-option analysis may provide assistance in thinking about the circumstances in which private investment may be utilised to achieve public purposes, and the extent to which the terms of the contract will need to explicitly recognise characteristics of the investment (including the variability in demand and the extent to which it is sunk) if efficient outcomes are to be achieved.

5 The ownership of facilities that deliver public services

5.1 Introduction

While there is a wide range of potential options for government to contract out the provision of public services to private profit-maximising firms, the most complex are those that require substantial investment in facilities associated with the provision of those services. In countries such as the UK, private investment in the infrastructure associated with public services makes up a significant portion of the total investment in public infrastructure, including investment in prisons, hospitals, schools, electricity, water and defence facilities.

Advocates of private ownership and management traditionally argued that the public sector tends to be wasteful and inefficient—characterised by rigid work practices, excessive staffing levels, and provider capture. Critics of this view tend to point to the problems of specifying the quality dimension of desired outcomes, establishing appropriate performance targets, monitoring subsequent performance and enforcing the relevant contractual obligations. They sometimes also argue that, as a matter of principle, the provision of services on behalf of the community should be undertaken by institutions of a wholly public nature.[25]

The incomplete contracts literature provides the potential to get past what had become a rather sterile and ideological debate on the role of the private sector in the ownership and management of facilities delivering public services. From an incomplete contracts perspective, ownership structures matter, because value is created by research into innovative approaches to carrying out the tasks of constructing and delivering services with the facility, but are not contractible *ex ante* even though they may be verifiable *ex post*. This means that ownership rights are associated with the residual decision rights to determine whether investment in research on innovations will take place, and whether the innovation will be implemented.

There is now a substantial literature looking at incomplete contracts analysis of the construction and operation of such facilities, beginning with the simple analysis of contracting out of management under public and private ownership in Hart *et al* (1997), and extended by numerous papers that develop more complex and realistic models of the costs and benefits of traditional procurement and government ownership, versus private construction and management of public facilities such as prisons. In this chapter we provide an analysis of this literature, and in addition, consider the implications of real options for these analyses of the optimal form of contracting for constructing and managing public facilities.

[25] For example, the traditional debates about the role of public ownership are highlighted by the literature on prison management (see Trebilcock, 1995; Boston *et al,* 1996:8-9).

5.2 Contracting-out under public and private ownership of the facility

Hart *et al* (1997) consider two cases: (i) the government can own the prison and employ a manager to run it, and (ii) the government can contract with a company owned by the prison manager to run the prison for a period of time. They ignore investments on the government side, but suppose that the prison manager can either invest in efficiency enhancements that improve the quality of prison services or invest in finding ways to reduce cost and quality within the letter of the contract with the government. A government employee has little incentive to engage in either activity since they will receive little or no share of the additional surplus created, but a private owner-manager has stronger incentives to do so. "The good news about this is that private ownership encourages the first, innovative, type of investment. The bad news is that private ownership also encourages the second quality-shading type of investment. The choice between public and private ownership depends on which of these effects are more important, and the extent to which the government can limit the extent to which quality-shading can occur"(Hart 2003:C71).

In a related analysis of private ownership, Lulfesmann (2007) extends this analysis by considering the importance of public-sector decision-making in the presence of a soft budget constraint as a rationale for ownership by the private sector. He argues that soft budget constraints may apply to private-sector firms operating important infrastructure facilities because of the proclivity of governments to offer bailouts to private firms to allow the continued operation of such facilities. He then demonstrates that in an incomplete contracts model where either the private-sector owner or the government renegotiate wages with the manager of the facility, once the ongoing viability of an existing technology is known, private ownership will generate the most efficient levels of investment in all states, except that where the existing technology is not viable and closure of the facility would be required if it was privately owned.

5.3 Comparing public procurement and private ownership of facilities and services

Hart (2003) discusses the relative merits of the conventional provision of services, whereby the government contracts with one private party to build a facility and then contracts with another private party to operate it ("unbundling"), and PPP, whereby the government contracts with a single private party to build *and* operate the facility ("bundling"). In either case, the private party responsible for building the facility has the option to spend more on construction. "Productive" investment raises the benefits of the services produced by the facility and also lowers the (operating) cost of providing them. In contrast, "unproductive" investment lowers both the benefits and operating costs of providing these services.

Under conventional provision, the building party simply minimizes the construction cost, and so does not invest in either productive or unproductive investment. In contrast, under a PPP arrangement, the private party internalizes operating costs (but not benefits), which leads it to over-invest in unproductive investment (since this allows it to lower its own operating costs, and it is unconcerned by the associated lowering of benefits in the process) and to under-invest in productive investment (since it is not motivated by the associated raising of benefits).

Hart concludes that unbundling is good, if the quality of the facility can be well specified whereas the quality of the service cannot be, as this avoids the overinvestment in unproductive investment that would result from a PPP arrangement (since there is little the government could do to penalize poor service provision). He also concludes that bundling is good, if the quality of service can be well specified in the initial contract and the quality of building cannot be (since a PPP partner can be incentivized to improve outcomes on the basis of quality of service).

Contractual incompleteness makes it impossible to contract on some things, but this still leaves freedom regarding how to contract over the issues that are contractible. For example, the choice of delivery date determines the amount of timing flexibility the "builder" has. Giving the firm more timing flexibility is giving it a potentially valuable real option, and this should be reflected in a relatively low tender price. This means that allowing a contractor more time to build a road allows it greater flexibility to avoid carrying out construction during periods of labour shortages or when other inputs are expensive. This benefit to the public needs to be measured against the cost of delayed introduction of the services the investment is to provide.

However, there are other issues to consider when deciding how much flexibility to allow in a contract, which are related to Hart's distinction between productive and unproductive investment. The private party responsible for the "building" stage has real options to engage in productive investment and real options to engage in unproductive investment. The terms of the contract will determine how much flexibility the party has, which will determine the value of these two options. Increasing the value of the party's real option to engage in productive investment is socially beneficial, as is reducing the value of the party's real option to engage in unproductive investment. For example, the builder might have the option to adopt emerging technology that lowers operating costs. If use of this technology is not something that can be contracted on, then giving greater flexibility regarding delivery dates will give the builder the option to wait and adopt the technology if it becomes economically viable (perhaps the technology becomes cheaper or the operating costs avoided become greater due to labour shortages, etc). This may be socially beneficial, or not, depending on the effect that adopting the new technology has on the benefits generated by the contracted services. In general, the choice of how much flexibility to include in a contract depends on how the firm will use that flexibility.

Hart's story, in its most basic terms, deals with situations where it is impossible to write a contract that will force a private party to operate in a way chosen by government. However, it is possible to predict how the private provider will behave, and that prediction helps government to decide whether or not to adopt a PPP arrangement and, if so, how to design the underlying contract. Circumstances may change during the course of a contract. For example, initially market conditions may be such that we would predict that the firm prefers bundling; perhaps the technology to undertake "unproductive" investment is currently so expensive that government believes that a PPP arrangement is suitable. If the price of this technology falls over the life of the construction phase of the contract, then the firm might actually undertake the unproductive investment, so that an initially desirable PPP may end up becoming an undesirable one. If this possibility seems likely, then contractual elements such as delivery dates might be tightened to reduce the flexibility available to the private party.

In an extension of Hart (2003), Bennett and Iossa (2006) consider a setting where a public entity delegates to private firms the construction and management of a facility that will be used to supply a public service. They model a situation in which specialisation requires

separate firms undertake construction and management, but where construction and management firms can form a consortium to undertake both aspects of the project, and where there is the potential for investment in welfare-enhancing innovations that could not be anticipated at the time that the initial contract is signed. Within this structure they consider the impact of allocating the residual ownership rights to the government (traditional procurement) or the private sector (under a type of public-private partnership that they term "private finance initiative" (PFI)).

Hence they consider ownership by:

- The construction company.
- The management company.
- The government with separate procurement from a construction company and a management company.
- The government with procurement from a consortium.
- A private consortium.

As with other applications of the literature on incomplete contracts, these ownership structures matter, because value is created by research into innovative approaches to carrying out the tasks of constructing and delivering services with the facility. The discovery and implementation of any innovations are not contractible *ex ante*, but they are verifiable *ex post*. This means that ownership rights are associated with the residual decision rights to determine whether investment in research on innovations will take place, and whether the innovation will be implemented. Under private ownership, and provided that basic requirements in the contract are not violated, the private firm makes that decision. Under traditional procurement with public ownership, renegotiation between the public manager and the private firm must take place before the innovation is implemented.[26]

If there is a positive externality associated with innovation at the construction stage, then it will always be optimal for construction and management to be undertaken by a consortium, whether the residual ownership rights rest with the government (procurement) or the private sector (PFI). This is because separate contracting for construction and management will always create the potential for underinvestment in value-enhancing innovations because the gains will be shared with government (under procurement) and since ownership by either the firm constructing the facility or managing the facility will result in only one of the potential innovations in construction or management being implemented.

Bennett and Iossa (2006) also consider a model in which there are three distinct phases: construction of the facility, management and delivery of services using the facility, and the residual value of the facility at the end of the contract period. This allows them to explore the impact of both positive and negative externalities between construction and service delivery, as well as the impact of residual value, on the project.

[26] This is consistent with the evidence in the UK that 73% of procurement contracts involved renegotiation of a higher contract price during the implementation of the project. In contrast, only 22% of PFI contracts were renegotiated at higher price levels, and where they were renegotiated it was primarily due to changes to the contract required by the public-sector agency.

In their model, PFIs should be used for the construction and operation of public service facilities where:

- The potential externalities from innovation are more strongly positive.
- The effect of innovations on social benefits that cannot be internalised in the contract is small.
- The effect of innovation on residual value is large.
- It is possible either to leave the facility in private hands in the long term, or assess its value as part of any renegotiation of the terms on which the facility would be transferred back to the public sector.
- The higher is the residual value for the private sector created by innovations in the contract.

The model generating these results differs from that of Hart (2003) in that it assumes that investments that result in innovations in building design or service delivery are verifiable. If they were not verifiable this would increase the range over which private ownership was optimal, because it would reduce the potential for firms operating under public procurement contracts to renegotiate contracts to obtain a share of the benefits from any innovation that they developed. Similarly the assumptions of the model that the contracting process is "one time" rather than a repeated game, and that information is symmetric, are unlikely to have strengthened the results in favour of private ownership and provision. The development of reputation in repeated rounds of contracting for similar PFI projects may increase the ability of public entities to identify private-sector owners whose construction and management of facilities will generate welfare benefits, and asymmetric information is more likely to increase the challenges of obtaining welfare-enhancing outcomes where private-sector firms working under procurement contracts have to negotiate with government to obtain a share of the value that their potential innovations will create.

5.4 The trade-off between scale and flexibility in construction and operation of facilities

Unbundling a project into separate construction and operating phases restricts the time during which the construction phase of a project must be completed. This will motivate (perhaps even compel) the builder to favour scale over flexibility, since it will be difficult, if not impossible, for the builder to spread construction of a project over time, building successive stages, if and when they are necessary. In contrast, a PPP-style arrangement would open up the possibility of staged investment, allowing the decision maker to use all available information when managing the construction phase.

One situation when the trade-off between scale and flexibility arises is when services can be provided using various combinations of (low cost) irreversible and (high cost) reversible investment (in their simplest form, "capital" and "labour"). In the early stages of a project, for example, it might be optimal to undertake reversible investment because, even though this will be more costly in the short-run, it can be redeployed if demand growth is lower than expected. If strong demand growth does occur, then the cheaper irreversible investment can be made without the risk that it will be stranded. When such investment strategies are feasible and attractive, bundling the construction and operating stages

together via a PPP is attractive, as it allows the building party to delay committing to irreversible investment until well into the operating phase. This is most likely to occur in situations where the real option to delay undertaking irreversible investment is most valuable, which will tend to occur in situations when demand volatility (or other sources of uncertainty) is relatively high.[27]

5.5 Conclusion

The literature on incomplete contracts has made a significant contribution to thinking about public-private partnerships. Studies of private construction and ownership of facilities associated with public services initially focussed on a simple trade-off between private investments in cost reduction and private investments in quality improvement. In this literature, the challenge with the use of the private sector in these initiatives was the design of contracts that reduced the incentives for the private sector to reduce costs excessively (resulting in unacceptable reductions in quality) and/or increased the incentives of public-sector managers to implement innovations that reduce costs or improve quality given that they personally would not benefit from that initiative.[28] The most recent contributions to this literature have added complexity to their models by building in renegotiation and residual value.

There is risk in public-private partnerships, but that risk is different from rather than greater than the risks associated with public procurement and management. The challenge is to determine how the contract should be structured to provide incentives that ensure that private-sector decision-making delivers benefits that are socially desirable as well as privately profitable. This means:

- Identifying those areas in which private investment in innovation is likely to have the strongest positive externalities for the construction and operation of a facility associated with a public service, including those areas in which uncertainties about the level of demand makes the value of the real option to delay or stagger investment high.

- Finding ways to internalise those aspects of performance which provide social benefit, such as making it profitable for private owners of prisons to manage prisons in ways that reduce re-offending rates, or retaining state control, where effective means of internalising these issues cannot be found.

- Maximising the incentives for private-sector value creation by avoiding build, own, operate and transfer (BOOT) schemes with compulsory hand-back to the public sector, since this truncates the benefits that the private sector obtains from value-enhancing innovations that have long-term positive impacts on the value of the facility (impacting on residual value).

---

27 Kandel and Pearson (2002) present a model of investment using a mix of reversible and irreversible technologies, and use it to examine the role that demand volatility plays in the optimal investment policy and the overall cost of investment.

28 This need not be interpreted as being critical of public sector managers. For example, it can be interpreted as a way to capture the fact that the return to cost-reducing or quality-improving innovations is uncertain, so that both in the private-sector and the public-sector context there are questions about how to provide the appropriate incentives for the party with the residual decision rights to make the appropriate levels of investment.

- Providing options for renegotiation which will allow private owners to capture benefits from innovations not anticipated when the contract was first written. Renegotiation may be around payments during the course of the contract, or around the residual value at which the facility would be transferred back to private ownership and management (if there is a BOOT scheme or similar).

The focus of the literature reviewed in this section on the investment decisions at the construction stage abstracts from the "control versus lack of control" issue that will arise in relation to the delivery of services where contracts are incomplete. In the next section, we consider those issues relating to the delivery of services.

6 Public vs private delivery of public services

6.1 Introduction

In Chapter 4, we considered the literature on the ownership of facilities that deliver public services. In this chapter, we consider the literature that has focused only on the question of private versus public delivery of services where it is assumed that the party delivering the service does not need to consider substantial facilities investment.

The literature on the allocation of ownership rights in the presence of incomplete contracts has provided a number of important explanations for the organisation of services firms. Given that contracts are incomplete, direct and/or complete incentive contracts will not produce the optimal level of effort in providing the required services. For service firms in particular, the customer base is a critical asset that will determine the success of the firm, especially where long-term customer relationships are important. In this context, the literature on incomplete contracts suggests that the allocation of residual ownership rights over assets such as the customer base of a service firm may be a critical determinant of the success of the firm.

In this chapter, we explore the allocation of ownership rights in the context of services delivered to customers, and explain how this literature may be relevant to the delivery of services by the public sector. The fact that services are publicly funded does not necessarily make delivery of those services by government departments or Crown entities the optimal form of organisation of service delivery. Where it is possible to increase investment in the provision of public services by making a private, community or other NGO entity the residual claimant in the value obtained from the delivery of those services, then traditional public-sector service delivery may be inferior.

6.2 Foundations: Grossman and Hart on insurance companies

In their foundation paper on this subject, Grossman and Hart (1986) used the example of the incomplete contracts for profit-enhancing investments by participants in the insurance industry to explain the allocation of ownership rights to clients. Grossman and Hart consider the choice between ownership of the client list by the insurance company, and ownership of the client list by the insurance sales agent. In their model (and in practice) the insurance company determines the type of insurance policies available for sale, and their retail price. Substantial cost is incurred in the identification of clients and in selling them policies. Insurance sales agents are paid via an up-front commission at the point of sale, and a (smaller) renewal commission when the customer renews their policy in subsequent years. The up-front commission rewards the effort expended on identifying and selling to customers. The presence of the renewal commission incentivizes the agent to select clients who will be persistent (that is, to avoid forced selling) and to provide post-sale support to the client so that the company can cover the substantial fixed costs of making the sale.

In general (motor vehicle, fire, and accident) insurance, policies are usually written for one year, whereas most forms of life insurance involve a long-term contract. This means that general insurance customers have a much greater tendency to switch companies than do life insurance companies. It also means that for life insurance the critical sales effort is in

initial selection of the client and matching of the appropriate policy, whereas in general insurance post-sale support to the client is much more important. It also means that for life insurance changes to the competitiveness of the company will not affect retention of existing clients, whereas changes in the competitiveness of the prices of a general insurance company will affect the retention of existing clients. In other words, complete contracts on agent effort and on insurance company competitiveness cannot be written.

Grossman and Hart (1986) show that in the portion of the insurance industry where agent effort in post-sale service to the client is most important, independent agents (agents who sell for a range of different insurance companies and own their client base independent of the insurance firm) dominate sales. In other words, the more important is post-sale effort from the agent, the more likely are the agents to own their client lists. In contrast, in life insurance where post-sale service to the client is least valuable, exclusive agents of individual insurance companies dominate sales, and this in turn means that the client list is owned by the firm. Thus, the structure of insurance firms, and in this case the ownership residual rights of control over the client list, is determined by whether the ownership by the insurance company or the sales agent delivers the largest *ex post* surplus from the production and sale of insurance.

Rebitzer and Taylor (2007) provide an analysis that is similar to that provided by Grossman and Hart (1986), but their example is law firms, and in particular, explaining why law firms are structured as partnerships with "up or out" rules for promotion from associate to partner. Their explanation is based around the idea that partners of a law firm control critical assets of the firm–specific knowledge about the business needs and interests of long-term customers. The specificity of this knowledge gives individual lawyers an incentive to try to increase their share of firm profits by threatening to leave and take their clients with them. Law firms respond to this threat in two ways. First, they fire lawyers who are not promoted to partner within a specified period of time to ensure that staff below the level of partner do not have the ability to develop long-term specific relationships with key clients of the firm. Second, the partnership structure provides incentives for each partner to refer the client to other specialist partners when the needs of the client require that, and reduces the incentive for any individual partner to leave and take with them the clients with whom they have built specific relationships, by creating an ownership structure in which all partners (and only partners) share in the profits of the firm.

Other professional services firms rely on long-term customer relationships, but as Rebitzer and Taylor (2007:225) acknowledge, not all of them use the same "up or out" tenure and partnership structure used by large law firms. The variation can be accounted for by a number of factors, including the ability of the firm to protect aspects of the relationship with the client through patents on key intellectual property, and the effectiveness of non-compete clauses in the contracts of senior staff. The importance of the reputation of the firm vis-à-vis the reputation of the individual partner or client-relationship manager, and the ease with which clients distinguish the reputation of the individual from the reputation of the firm, are also likely to be critical.

A related literature considers the relationship between delegation, information, and the contribution of different parties to the generation of total surplus in any activity. Short of allocation of ownership to a particular party, delegation of decision rights and compensation packages linked closely to the outcomes of those decisions may provide mechanisms that reduce the inefficiency created by incomplete contracting. When information is publicly available or otherwise easily measured, it is efficient to centralise

decision-making. But in the presence of factors that will make performance contracts incomplete, such as the need to adapt to a local environment or employees with valuable private information, delegation of decision-making together with pay for performance may represent an optimal solution (Aghion and Tirole, 1997). Among empirical investigations of this issue, Wulf (2007) finds that the compensation of managers who have broader authority is more sensitive to firm-level performance, and in a survey of sales force personnel, Ghosh *et al* (2011) find a positive relationship between performance pay and the level of delegation.

6.3 Contracting and procurement by the state

When non-contractible quality dimensions (or tasks) are important, letting suppliers or agents compete on price or other contractible dimensions may lead to a very inefficient outcome for the buyer/principal. Transactions, however, are often regularly repeated. Reputational forces may then help, and non-contractible dimensions can be governed by self-enforcing relational contracts. A supplier who overstates the quality of a good in circumstances where the quality of the good only becomes apparent after a period of use, a consultant or employee who purposely reduces non-verifiable but *ex post* observable effort below what was promised, or any other agent who behaves opportunistically can then be punished by the principal in future interactions. Cooperation incentives are typically stronger, the higher are future expected payoffs, and so relational-contracting needs may conflict with other important needs of a principal, in particular, that of inducing agents to compete, both to capture more surplus from the relationship and to screen and select more able agents. This means that, when designing a relational contract to govern non-contractible dimensions of the transaction, the principal must also take into account his/her present and future choices on whether, how and when to screen agents competitively, besides the design of any underlying explicit contract.

Private procurement and the importance of the interaction between relational contracts and competitive screening policies lie at the heart of many discussions about Toyota's "relational" procurement policies compared with, say, General Motors' more "competitive" or "arms length" ones. Toyota, along with many other Japanese firms, maintains a small stable set of "highly trusted" dedicated (and often exclusive) suppliers, restricts competition for the various orders to these suppliers, caring for their profitability and rewarding the best performing suppliers with a higher share of orders, while replacing those that fail to deliver the extremely high levels of contractible and non-contractible quality required. The limits to competitive screening have a cost in terms of reduced screening and therefore higher prices, but ensure sufficient weight is placed on the future and a consequent cooperative perspective in the supply relationship. This explains why Japanese firms were so unhappy when political pressure from the United States forced them to make their supply chain more open to competition from US suppliers. The general validity of this well-known story and of this interpretation of it in terms of interdependence between competitive screening and contracting choices has been empirically confirmed by a recent study showing that competitive screening and long-term relationships tend to be substitutes (see Radkevitch *et al,* 2008).

Turning to public procurement, Banfield (1975) draws attention to the reduction in the quality of government procurement linked to accountability rules forcing public buyers to use open auctions for supplier selection. Kelman (1990) followed up on the theme of quality losses linked to too rigid rules forcing open competitive screening of suppliers, and on the importance of taking past performance into proper account. He stressed the

differences between public and private procurement processes, noting that private firms used open auctions much less often, left higher margins to suppliers, switched suppliers less often and were much more satisfied about the quality of the procured goods and service.

Calzolari and Spagnoloz (2009) developed and studied a dynamic model of recurrent exchange with non-contractible dimensions between an infinitely lived principal (the buyer) and a population of heterogeneous and privately informed, infinitely lived agents. The model incorporates both moral hazard (on non-contractible dimensions) and adverse selection (on agents' type). Non-contractible quality can be governed by a relational contract, and should be interpreted in broad sense to capture all value-enhancing decisions, like investments, that a supplier is free to take during the contract execution and which the buyer observes, but cannot regulate with an explicit contract.[29]

Calzolari and Spagnoloz show that in a dynamic procurement process a buyer may optimally restrict the number of potential trading partners at the cost of reduced screening and more expensive procurement to boost non-contractible quality provision. By restricting competition, the buyer loses performance bonuses (since they are not more credible), but leaves firms sufficient future rents, so that they can find it profitable to build reputational commitments for future interactions and prefer to provide the high levels of highly valuable non-contractible quality the buyer requires (refrain from moral hazard). Shortening the duration of the (explicit part of the) contracts is crucial for the relationship. Abstracting from technological aspects such as the rate of obsolescence, a shorter duration of supply implies more frequent re-selection or search. Higher frequency of search makes it easier for a buyer to obtain high non-contractible quality levels from sellers by threatening to withhold the bonus or to exclude the seller from future trade. Indeed, with more frequent contracting, the threat of exclusion is closer in time and gains from "cheating" are smaller, so that higher quality can be expected.[30]

Thus, in this type of model, longer duration of supply contracts—less frequent auctions—together with larger pool of competing suppliers both deter collusion among eligible suppliers but also reduce non-contractible quality levels obtainable from them. Symmetrically, shorter contracts—more frequent auctions—and a smaller pool of suppliers both facilitate suppliers' collusion but also the enforcement of non-contractible quality standards.

[29] Because Calzolzri and Spagnoloz consider both contractible and non-contractible quality or tasks, their model is a dynamic model with multi-tasking in the sense of Holmstrom and Milgrom (1991).

[30] Some recent studies do confirm this intuition. For example, studying a dataset for train operating companies in UK, Afuso and Newbery (2002) show that (discretionary) investment is stimulated by shorter rather than longer contracts. Notwithstanding a standard hold-up problem associated with contract renewal that should point in the opposite, the authors suggest that frequent re-procurement with short contracts disciplines suppliers who care for future re-award of the franchise.

6.4 Who should make the decision on private or public delivery?

The incomplete contracts literature has been extended to consider how, or by whom, the decision on the allocation of residual decision rights should be made in the public sector. This literature is consistent not only with that reviewed in earlier sections of this paper, but also with the idea that the process by which the allocation of rights is made may be very important in reaching an efficient allocation.

Bennedson and Schulz (2011) consider various approaches to delegation of public service decisions in a framework that is general enough to capture delegation within political hierarchies as well as the use of specialist agencies to make decisions about ownership rights. They begin with the trade-off between cost and quality in public-sector service provision identified by Hart *et al* (1997), and the inability to write complete incentive contracts to address this problem. They then consider different approaches to delegation in the context of a decision about production that may occur in-house or be outsourced. The framework of the model allows for renegotiation of the contract once the in-house or outsource service provider has developed a proposal for cost savings.

The view of the outsourcing decision developed by Bennedson and Schulz (2011) is innovative in their identification of two important effects of delegation:

- **The incentive effect**. Delegating the rights to negotiate with a service provider to an agent with known preferences may substitute for incentive contracts. For example, if renegotiation is delegated to an agency whose mandate is focused on achieving cost reductions, this will incentivize investment in cost reductions by the service provider.

- **The bargaining effect**. Delegating the right to determine whether outsourcing will occur will affect the bargaining about the cost of outsourcing options. If the right to decide whether services are outsourced is placed in the hands of an agency with a know preference for in-house provision, this will induce private bidders for outsourced contracts to offer lower prices.

Within this framework, they show that partial delegation, where the principal (government or minister) makes a decision in principle or clearly establishes the preferences for the agency to whom subsequent decisions or bargaining are delegated, is the optimal institution, while full (arms length) delegation is superior to no delegation. The reason is that delegation of decision-making can create incentives in contracting that cannot be replicated in-house or in the absence of delegation, and the bargaining effect means that efficient outsource contracts may be negotiated even when there is limited competition among private providers.

These findings are consistent with a more general literature on the benefits of delegation, and the circumstances in which delegation will be politically feasible. For example, Ludema and Olofsgard (2008) argue that delegation will occur where there is a consensus that time consistency in policy has benefits that may not be achieved because of incentives to attempt to use policy for short-run stabilisation, and the more polarizing or contentious policy decisions are across different constituencies in the electorate. This approach is often used to explain, for example, the delegation of competition policy and monetary policy, but ministerial control of fiscal and taxation policy.

6.5 Conclusion

Government is involved in wide range of contracting for services, from services supplied to government itself to services that government purchases on behalf of the community. A traditional approach to the economic analysis of this type of contracting would have emphasized transaction costs and standard cost-saving versus quality tradeoffs. An incomplete contracts approach provides a different perspective by focusing on these arrangements as allocations of residual decision rights in the customer relationship, and as a mechanism by which to ensure that those rights sit with parties that add the most value to the relationship. This literature may potentially be applied to the public sector in a variety of ways. For example:

- The importance of the relationship between general practitioners (GPs) and their patients provides an explanation for the fact that GPs are the part of the health sector that in almost every country is privatised. Since the GPs' investment in human capital and patient-specific knowledge will both be the key to delivering the best services to patients, government ownership (as opposed to just funding) of the general practice may reduce efficiency.[31] By comparison with GPs, it is not clear that hospitals have long-term customers or customer-specific knowledge to the same extent, which might explain why public hospitals do not "own" the ongoing relationship with their patients.

- Community-based lending programmes (such as microfinance) may be effective in using relationships and moral suasion within communities to replace high cost monitoring and enforcement mechanisms used by loan companies and loan sharks.

- Whānau Ora and other community-based initiatives can perhaps be seen as about providing a community organisation with ownership of the relationship with the client/patient. The more specific the knowledge that the community organisation might have in relation to the provision of service to the patient, the more important that it is the community organisation that owns this.

The value of these types of relationships is reinforced by the literature on "arms-length delegation" of public activities, which demonstrates that a range of problems created by contractual incompleteness may be addressed or ameliorated where the government delegates service provision or other decision-making to a third party.

[31] This comment should not be taken to imply that there is no room for improvement in the current arrangements for delivery of primary health care in New Zealand. For example, there may be ways in which efficiency could be improved by considering more carefully the incentives that are provided for GPs by the way in which government payments to them are structured. However, it does suggest that any move away from GP ownership of the patient relationship would need to take account of the potential reduction of the incentives for GPs to invest in that relationship, and the implications of that for the quality of care provided, as a cost to be balanced against the benefits that any alternative structure might provide.

7 Governance

7.1 Introduction

We pointed out above that, as a matter of analytical convenience, the existing literature on incomplete contracts tends to abstract from the problems of governance and agency that are central to the management of large firms (and large public-sector organisations). However, this does not mean that the incomplete contracts literature is irrelevant to questions of governance; indeed, quite the opposite is true. In a world where complete contracts could be written (that is, where it is feasible and efficient to formalise all future contingencies in a contract), there would be no role for governance because all material decisions would be made *ex ante*. But in a world where contracts are incomplete, corporate governance may be defined as the set of conditions or the institutions that shapes the rights to make *ex post* decisions in relation to unspecified contingencies, and in particular to determine the ultimate unspecified contingency—the allocation of the surplus between owners and management.[32] The importance of this literature is that it shifts the focus of the analysis of governance and boards of directors away from agency (incentive) problems and towards decision and investment problems.

We begin by defining governance. We then consider the difference between the agency and the incomplete contracts view of governance. We outline the lessons from the recent literature on governance, before considering how those lessons might relate to the governance of public-sector organisations.

7.2 Agency and incomplete contracts perspectives on governance

Governance is important in any circumstance where the management of an organisation is separate from those who provide the funding for it. In fact, the critical role of governance is to provide suppliers of funds to the organisation with sufficient confidence that the management of the organisation is consistent with their interests so that they are willing to go on providing funding for the organisation. In these generic terms, governance is important whether funding is sourced from taxpayers and provided by Parliament for a government department or Crown entity, or is sourced from shareholders and creditors and provided to a corporation.

The role of governance (Zingales, 1998) is to:

- increase the incentives for value-enhancing investments, while reducing inefficient power-seeking. A governance system will provide a check that management is investing in activities that are value-enhancing for the firm rather than for the managers, including encouraging firm-specific investments and discouraging activities that add more value for management than they do for the firm

- minimise *ex post* inefficiency in bargaining about the allocation of surplus. In particular, governance structures reduce the inefficiency associated with free-riding by individual owners when ownership is widely dispersed, and reduce information asymmetries between management and owners

[32] Or, as we explain below, where there are no owners allocating surplus, however defined, among stakeholders.

- allocate the residual risk to the least risk-averse party, or in other words, the party that is best able to diversify the risk associated with the activities of the firm.

This definition of governance establishes a requirement for any governance framework to provide for both monitoring opportunistic behaviour in relation to incentive contracts, and provides a mechanism for making decisions in relation to matters about which contracts were not and/or could not be written *ex ante*.

Until the development of the literature on incomplete contracts, the focus of the literature in economics was on explaining governance in terms of agency theory. For Schleifer and Vishny (1997:738), their "perspective on corporate governance is a straightforward agency perspective, sometimes referred to as separation of ownership and control." This literature focused on explaining agency problems that arise because ownership is widely dispersed (or otherwise distinct from management) and the incentives of managers are not perfectly aligned with those of the owners. For example, managers may focus on projects in which they are personally interested in rather than on projects that maximise the value of the firm, may avoid difficult decisions or risky projects that would increase the wealth of owners, may seek out or oppose organisational changes that maximise value, and may focus on recruiting staff who fit with an existing (sub-optimal) culture rather than staff who will challenge existing ways of thinking. In agency theory, the primary role of the board is to assure shareholders that agency conflicts of interest are being controlled.

One of the reasons that the agency perspective is limited emerged from the literature on incentive contracts. Incentive contracts solve agency problems where it is possible to anticipate *ex ante* what the owners of an organisation want the management to maximise, and how they want management to react to different contingencies. But even in respect of a for-profit corporation, it is not possible to anticipate all contingencies in advance, or to anticipate the managerial responses that owners wish to incentivise in different situations. To take one example, relational contracts, which are typically very incomplete and work on the basis of a wide range of decisions made by the parties to the contract interacting over time, are commonly found in the private sector. In respect of a public-sector entity, multiple objectives and intangible output measures make satisfactory *ex ante* specification of complete incentive contracts even more unlikely.

From an incomplete contracts perspective, the focus of governance shifts from monitoring designed to minimise opportunistic behaviour within management incentive contracts to the necessity for an effective mechanism for *ex post* decisions about matters that are inefficiently costly or impossible to specify in a contract against future observable variables. In other words, the existence of incomplete contracts is what makes corporate governance distinct from contractual governance (Zingales, 1998:499). When contracts are incomplete, it is necessary to allocate the right to make *ex post* decisions in relation to contingencies that cannot be specified in advance, and the allocation of those rights to a board of directors provides an efficient form or delegated decision-making on behalf of the owners of the firm as a whole.

7.3 Functions of boards of directors

Private-sector boards of directors are the representatives of the shareholders who are the residual claimants in the activity of the organisation, and public-sector entities are similar, if it is considered that ministers are representatives of the public who are the residual claimants in that sense. The overarching function of a board is in acting for shareholders

in respect of most decisions, excepting that of ownership itself. Adams, Hermalin and Weisbach (2010) list board functions as including:

- CEO selection, monitoring, evaluation and compensation
- setting firm strategy, and
- providing critical resources to the firm, such as external networks and perspectives that may not be present in the management team.

These roles require that the board has access to significant information that is not provided or generated by the firm.

A focus on the role of the board in monitoring the actions of management implicitly assumes that all decisions are delegated to management, and that in the extreme the primary power of the board comes from its appointment and evaluation of the chief executive. In practice, however, such complete delegation is rare. Most CEOs work under constraints that some decisions require the approval of the board of directors.[33] The role of the board in key decisions goes directly to the heart of the incomplete contracts perspective on governance: boards of directors exist not just to guard against opportunism in respect of the existing incentive contracts, but to provide for completeness in those areas where it is impossible for the owners to give the management appropriate direction (via their incentive contract) *ex ante*. It follows that board functions necessarily involve influencing (being responsible for) strategy as well as management monitoring and key final decisions.

In the private sector, goods and services are provided by for-profit and not-for-profit entities that are governed by boards. Not-for-profit entities are further divisible into owned firms (eg, as observed in the health sector) and non-owned firms (eg, universities). Hansmann (1996) contends that the ownership and control of firms will evolve to the entity being owned by the group of stakeholders whose ownership results in the least combined costs of ownership and contracting. In this, Hansmann has a broad definition of stakeholders that includes suppliers to the firm (including suppliers of raw materials, labour and finance—equity (shareholders) —and its customers. The costs of "ownership" include the costs of co-ordinating (eg, with relevant stakeholders in for-profit firms: shareholders) contracting, decision-making and motivation (ie, inducing management to run the firm efficiently). These contracting costs include transaction costs, costs of market power imbalances, and contractual hold-up costs. According to this theory, the board of directors oversees the operation of the organisation in the context of the functions that the organisation assumes. The board will be engaged in determining strategy, monitoring and completing incomplete contracts among its stakeholders.

The board is a team (Bolton and Dewatripont, 2005) and the agency factors associated with teams are relevant to the performance of boards of directors.[34] The performance of teams is affected by their size (Jensen, 1993) and, while all studies do not reach the same conclusion, many indicate that smaller boards of directors are associated with superior firm performance (Hermalin and Weisbach, 2003). The advantages of small boards

[33] Indeed, there will be some level of decision that requires a resolution of a majority of the shareholders at a general meeting, indicating that even the existence of a board does not comprehensively address contractual incompleteness.

[34] Interestingly, peer reporting on director performance, which is part of the operation of some boards, has recently been shown to offer a general approach to the agency problem of teams. See Kim (2011).

include clearer accountability for the decisions taken by the board. Small boards also imply a commitment to focus board appointments on monitoring and decision-making capability, since smallness rules out allocating a seat on the board to all stakeholders. Small boards also have the advantage of facilitating higher payments to the individual members of the board, and this in turn is likely to attract individuals with higher opportunity cost of time and larger reputational capital at risk in the operation of the board—both factors that are likely to be positively correlated with good governance (Adams, Hermalin and Weisbach, 2010:91 – 96).

The boards of directors of not-for-profit firms differ from the foregoing, because these entities cannot summarise the outcome of their endeavour in a single dimension (profit or financial value added). Without a unifying aggregator (profit), these organisations have multiple activities that the board must reconcile in strategy and monitoring. Also, not-for-profit status alters the form of agency issues, because it rules out opportunistic behaviour with respect to the *ex post* allocation of rents between owners and managers (Glaeser and Shleifer, 2001) although quasi-rents may still be appropriated by management through perquisites. Not-for-profit organisations include consumer cooperatives[35] and firms that provide educational, health and social services based upon fee, donation and grant-funding. The multiple dimensions and service orientation of these activities require frequent monitoring[36] and service-specific knowledge by the board and CEO; which suggests larger boards and the inclusion of stakeholder directors. The not-for-profit status admits funding (donations) that would not be present, were the organisation for profit, and this too is often reflected in board expertise. These general differences from for-profit boards are confirmed in O'Regan and Oster (2005), which reports a survey illustrating that non-profit boards include donors, suppliers and recipients of the entity's services.

Boards of directors then may be considered a team of information providers, and decision-makers, among whom there are different interests, and incentives for participation and decisions. These differences can be expected to be at least as wide, where the board consists of non-equity stakeholders as is commonly the case in not-for-profit organisations. While it might be considered that the incentives for board opportunistic behaviour are strongest in for-profit firms, incentives for this behaviour in not-for profits may also be very substantial among stakeholder directors. In both sets of organisations, key agency issues concern interaction among the board and the CEO.

The position of CEOs relative to boards of directors and the proportions of independent directors[37] vary widely across countries (Tirole, 2006:30-31). We confine attention to boards of (largely) independent directors and non-director CEOs, but this changes the nature, rather than the quantum, of agency issues that arise. The limited theoretical literature tends to analyse each agency issue separately and the vast empirical literature is plagued with difficulty in controlling for factors so that the empirical import of a subset of these factors is hard to distinguish (Adams, Hermalin and Weisbach, 2010). Sources of conflict—and synergies that may be at the expense of shareholders—among inside/outside and independent directors in for-profit firms are discussed in Milgrom and Roberts (1992:Ch.15). Sources of opportunism include the managers having more

35 Consumer cooperatives can be organized as though for-profit with certain distribution constraints. They also have more sharply defined and obvious objectives than the other not-for-profits cited. These predispose different organisational processes—a less diverse board than more general not-for-profits, for example.

36 See Brock and Evans (1996).

37 Independent directors we define to have no interest in the entity other than their engagement as a director. Inside directors we define to be employed in some way in the entity, and outside directors to be limited to a shareholding interest.

information and control of an entity's resources, different incentives and appetite for risk, where they have weak interests as residual claimants. Even in for-profit entities, the board and management have mixes of conflict and coordination that significantly affect the performance of the entity. In not-for-profit organisations, the CEO agency issues remain, but differ by the implications of the different objective function and concomitant board composition. For example, the absence of a residual claimant in not-for-profits removes the separation of ownership and control and the concomitant opportunity for risk and management to be assigned to different parties (Fama and Jensen, 1983).

7.4 Governance in the New Zealand public sector

The public sector in New Zealand is made up of a wide range of different organisations, including departments of state, Crown entities, state-owned enterprises (SOEs), and *sui generis* organisations such as the Reserve Bank of New Zealand. Crown Entities are of three types: Crown agents (such as -Accident Compensation Corporation (ACC), New Zealand Qualifications Authority (NZQA) and New Zealand Trade and Enterprise (NZTE)), autonomous Crown entities (such as Te Papa) and independent Crown entities (such as the Commerce Commission).[38]

For SOEs, the dominant Crown interest is ownership. In this case, it is relatively easy for the Crown to clarify its objectives, for these are typically shaped by the fact of these entities operate in industries that have more-or-less competitive markets, or if not, are subject to standard approaches to regulation. Consequently, the government may leave it to the board to ensure that management of the SOE meets its objectives.

But for Crown entities, the situation is more complex, because in many cases the Crown interest involves a more complex mix of ownership, regulation and purchase objectives. Crown entities are of particular interest, because the government owns them and therefore has rights to direct them unless constrained by law. At the same time, Crown entities generally have boards of directors, while also being subject to government policy, monitoring by a relevant government department, and oversight by a responsible minister.[39] The way in which the government's rights of direction are exercised will have important implications for performance—if the structure and principal are incorrect, then responsibility and accountability will be diffused and performance will be relatively poor (Scott, 2001:271).

A common justification for creating Crown entities is the substance, or perception, of greater independence from ministerial direction by comparison with departmental heads. But while the mechanisms for ministerial direction of Crown entities are in theory more formal than those for government departments, in practice there may be little difference. The presence of a board in Crown entities (a governance layer between the minister and the chief executive), contributes to the perception that Crown entities have a greater degree of independence from political intervention in the management of their affairs (Scott, 2001:275). In practice, however, this independence may be more apparent than real, leading to a highly confused and ineffective governance structure.

---

38 Crown Entities Act 2004.

39 The most direct link to government policy is for Crown Agents, which are organisations that give effect to government policy. Independent Crown entities are generally independent of government policy, while autonomous Crown entities must have regard to government policy.

The responsible minister is answerable to Parliament for the actions of the Crown entity. This includes actions taken under delegation by boards and managers. Ministerial accountability for a major failure is therefore unlikely to differ, if it is in a Crown entity or a government department. Further, while Crown entities have greater management freedom than government departments, this is counterbalanced by the ease with which ministers may change appointments (including the appointment of the chair of the board). Although board members of Crown entities have appointment letters with terms of office, they really serve at the pleasure of ministers (Scott, 2001:276).

Ministers are therefore the principal and (in many respects) residual claimant in the operation of both government departments and Crown entities: they carry the risk of non-performance and these risks extend beyond the performance of the entity *per se* to political risks arising from the fact that the minister has ultimate responsibility for the performance and strategic direction of the organisation. The CEOs of these entities are selected by the board, as they would be in the private sector, but they are public servants whose appointment requires confirmation by the minister.

The boards of Crown entities are appointed by the minister, but the boards are in most cases large and distant from the minister. This is appropriate in those situations where the interest of the minister can be expressed in legislation or occasional policy documents, but it is more difficult in Crown agents where the minister has substantial political risk from policy and service delivery, and thus expects to be allocated residual decision rights in all material matters. For Crown agents, statements of intent are cumbersome *ex ante* mechanisms for the specification of objectives that meet with the approval of the minister, consistent with an assumption of complete contracts, but inconsistent with the real world of contractual incompleteness. The problems of contractual incompleteness are compounded by the complex mix of objectives associated with public-sector organisations, and the way in which those objectives will change (both explicitly and implicitly) with changes in minister.

In practice, generic contractual incompleteness is addressed by a high level of interaction between the CEO of the Crown agent, the chair of the board of the entity, and the minister. This means that most important decisions are decided away from board meetings, and brought back to the remainder of the board as a *fait accompli*. In turn, this means that the most that can be said for the accountability of the boards of Crown agents is that they have responsibility for oversight of implementation of the decisions made by the minister, but it is not clear how much value they add in this respect. Effective Crown agent boards may address agency issues such as the level of effort provided by the CEO and senior management, but here the size and incentives of the board will be relevant.

The boards of Crown entities differ from boards found in the private sector as they are inserted between the principal (the minister) and the operation of the entity. The boards of Crown entities, particularly Crown agents implementing government policy, are likely to lack the knowledge or mandate to address problems of contractual incompleteness without reference to the principal. Thus Crown entity boards:

- May serve to confound the implementation of policy or performance via their own actions/acting as a shield for the CEO.

- At best operate to provide external advice to the CEO about their relationship with the minister, and monitor against the statement of corporate intent.

- May default to being stakeholder consultation boards rather than governance boards and the likelihood of this outcome will be positively correlated with the size of the board and the representativeness of its members.

Further, we note that the absence of full delegation of responsibility (eg, the minister for ACC rather than the board sets the levies, and the minister of tertiary education controls government funding and sets ceilings on the tuition fees set by universities) means that senior executives and boards always have ways to explain poor performance as resulting from factors outside their control.

A further complication arises from the existence of multiple central monitoring agencies of government. These agencies are normally identified as the State Services Commission (SSC), Department of Prime Minister and Cabinet (DPMC), and the Treasury. To these should be added, as a minimum, the Office of the Auditor General (OAG). And beyond this, a very wide range of monitoring responsibilities are allocated to different agencies, for example, to the Department of Labour in respect of ACC, and to the Tertiary Education Commission in respect of Universities, Wānanga and Polytechnics and Institutes of Technology. This makes for a convoluted and ineffective public-sector monitoring regime in which the collection, analysis and use of performance monitoring information has not lived up to the expectations associated with the creation of that regime (Gill and Hitchiner, 2010). We consider that this ineffectiveness arises for two reasons. First, multi-agency responsibility will naturally reduce clarity about where true responsibility lies and who is responsible for monitoring what, while also increasing the scope for genuine confusion about the boundaries of responsibility for monitoring. Second, this reduction in clarity reduces accountability for the quality of monitoring and reporting by any individual agency, and a reduction in accountability is likely to produce a concomitant reduction in effort devoted to monitoring.

7.5 Conclusion

The new public management model of the 1980s focused on the potential for public-sector boards to resolve agency (opportunistic behaviour) problems consistent with the academic literature of the time. However, the most recent literature in economics provides grounds for suggesting that the governance models in the New Zealand public sector (both as they were originally conceived, and as they have evolved over time) may need to be reconsidered. In particular, boards of Crown Agents are not close enough to ministers to be delegated authority to make decisions on matters not specified in *ex ante* ministerial instructions, which means that in practice CEOs and (often) board chairs engage directly with ministers on matters of importance. This, in turn, means that these boards are ineffective in undertaking the key role of governance—resolving contractual incompleteness. These boards may serve other roles, including representation of stakeholder groups, but this mixture of roles does not increase the quality of governance, and the Crown entities will often have more effective mechanisms for stakeholder consultation at their disposal.

These problems could be resolved in three complementary ways:

- Separate those entities in which the boards serve primarily as advisory boards rather than as governance boards, and reconstitute them as such. This would make it clear that for those entities, the minister rather than the board had the responsibility for addressing contractual incompleteness.[40]

- Establish a much clearer distinction between public-sector organisations where governance can be effectively delegated to a board of directors, and enshrine that separation from ministers much more clearly in the constitutions of those entities (as has been done for the Reserve Bank, Commerce Commission, and universities). Following the approach that we have set out above, the private board model is appropriate where ministers can delegate to the board residual decision-making rights. These entities would be those where relatively complete contracts relating to the objectives of government could be set out *ex ante*, or directed through the purchase interest of the government.[41] Where contracts are highly incomplete, as they appear to be (at least in the minds of ministers) with many Crown entities, then governance would be improved by making the boards advisory boards only, paying only meeting expenses for those boards, and making them truly representative of stakeholders.

- Change the structure of the boards that remain, making them smaller to ensure that the whole board can be informed of the minister's views, and leave them to make decisions. This would have many advantages in making boards more effective, since smaller boards could pay more to each board member, would provide much stronger individual accountability for board members, and would allow ministers to select individuals with much larger reputational capital. Higher remuneration and reputational capital would mean that individual board members had much more to lose in the event of ineffective governance.[42]

Note that it is likely that these two approaches will be complementary in a variety of respects, including in that smaller boards are likely to provide better governance in every case. Where boards currently serve to provide stakeholder representation, then exploring alternative mechanisms for that representation may be more effective.

---

40 Several possible approaches to achieving this present themselves, including a dramatic reduction in the number of Crown entities, and even a questioning of the whole governance and reporting framework for Crown entities.

41 SOEs are the clearest example.

42 It also seems likely that the savings associated with the reduction in the time and effort required to make the current very large number of board appointments would be material.

8 Public management: personnel economics in the public sector

8.1 Background

The modern literature on "personnel economics" or the economics of human resource management has focused on a range of traditional and new lines of inquiry. Our understanding of the incentives provided by the hierarchy of roles and remuneration within organisations has been advanced by the development of the theory of internal labour market tournaments. Links between performance and remuneration have been explored, particularly in relation to pay for output versus salary or wage-based pay, and the link between the remuneration structure used by a firm and overall organisational performance. Since the explicit use of teams in work environments has increased considerably in the last 25 years, much effort has been devoted to both understanding the nature of team work, and to the complexities of incentives and remuneration for members of teams.

In this chapter, we briefly review the literature in each of these areas, and suggest some possible areas of relevance for the public sector in New Zealand. Developments in the last 25 years to a large extent reinforce and extend the focus on remuneration, incentives, performance and benchmarking against private-sector labour markets that was a feature of the NPM literature in New Zealand.

8.2 Tournaments and promotions

Since the seminal work of Lazear and Rosen (1981) and Lazear (1995), economists have devoted considerable effort to the study of internal labour markets and the promotion "tournaments" that operate within them. The theory of tournaments in internal labour markets provides an integrated theory of compensation at different levels of the management hierarchy within firms, and in particular explains:

- the setting of remuneration primarily on the basis of roles rather than the human capital of the individual (as is demonstrated by the fact that remuneration is assigned to a role before an individual is assigned to the role), and
- the discrete jumps in remuneration that occur when individuals are promoted to higher levels.

Promotion is based more on relative rather than absolute performance, which means that promotion is a tournament in the sense that what matters to workers is whether they outperform their colleagues in the same firm. Further in this promotion tournament, the winner takes all of the prize, which illustrates the point that the compensation at each level is not designed to motivate the individuals at that level, but rather to motivate the individuals below that level to seek promotion.

Dispersion and compression of remuneration levels

The larger the pay spread between different levels of the hierarchy, the stronger the incentives to achieve promotion to the next level in the hierarchy. Large pay spreads may induce high levels of effort. However, large pay spreads may also induce work environments that are unattractive, or that inhibit the participation of certain types of staff or staff of certain ages. The 80-90 hour work week characteristic of financial services and

law firms with very big prizes for those achieving the top jobs or becoming partners are an example. In addition, large pay spreads provide incentives for staff to collude and/or sabotage the promotion prospects of others.

Where collusion and sabotage are feasible and/or likely to be highly destructive of long-term productivity in the firm, pay compression (defined as remuneration levels that are less variable than individual performance) may be a superior alternative to large differentials in remuneration. Pay compression also serves to reduce the incentives to invest in lobbying for promotions rather than investing that effort in increasing the output of the firm.

Finally, the dispersion or compression of remuneration will provide staff with more or less insurance against low-productivity outcomes (some part of which may be exogenous to their effort). For example, larger pay spreads mean that staff know that low-productivity outcomes will be reflected very directly in their remuneration, whereas compressed remuneration levels provide insurance against low-productivity outcomes which will also reduce incentives for effort.

Internal labour markets and external recruitment

This literature emphasises the importance of internal labour markets in shielding workers from fluctuations in external labour markets by having a specific (junior) point of entry with internal promotion to more senior positions. Internal labour markets also promote higher levels of investment in firm and industry-specific capital than are optimal, if workers expect to be required to have regular resort to external labour markets which may allocate them to different firms and industries. However, internationally and (we suspect) in New Zealand, external recruitment has become more rather than less common in many sectors of the economy. The trade-offs associated with external as opposed to internal labour market recruitment are also of increasing interest within the New Zealand public sector (as recent publicity about decisions to advertise publicly for senior and ambassadorial appointments at the Ministry of Foreign Affairs makes clear).

Traditional labour-market models have focused on the requirement for specific skills and characteristics as the rationale for recruiting externally, but suggested that this involves a trade-off in respect of the intensity of the incentives provided by internal promotion tournaments. From this perspective, external recruitment increases the moral hazard and monitoring problems of the firm by reducing the intensity of incentives for workers to strive for promotion. In other words, the intensity of competition for promotion among cohorts of existing workers is reduced, if those workers think that the promotion may go to an external recruit rather than being allocated to the best candidate from the internal cohort.

An alternative perspective which has considerable merit is provided by Chen (2005). Chen notes that workers engaged in a rank-order tournament will have incentives to engage both in productive activities (that enhance the value of the firm) and unproductive activities such as “sabotage” of their opponents’ performance (see Chen 2003 and Lazear 1989) or collusive agreements to shirk. This has two implications:

- When the potential for sabotage is introduced into the internal labour market it means that rank-order tournaments may induce workers to waste resources on unproductive sabotage, but also may mean that those who have the greatest chance of being promoted may not be those workers who add the most value to the firm.

- Collusion to reduce effort, if it is successful, may substantially reduce the payoff to the firm, but result in each individual having no less probability of winning the internal promotion tournament. When work is carried on in teams, the monitoring conditions required for successful collusion may be met, and collusion may be sustainable.

Outside competition for positions addresses both of these problems.

First, external recruitment reduces the payoff to both productive activities and sabotage, but this in itself does not establish net benefits for external recruitment. Second, sabotage has no value in competing against external candidates, but productive effort remains a useful instrument in this tournament. Thus, while external recruitment does reduce the total effort of existing workers, it results in a net increase in productive work and a net decrease in investments in sabotage activities. Thus, "although external recruitment hurts the 'morale' of insiders and reduces their total effort, the output of the workers will actually increase" (Chen, 2005:261).

Second, competition from external applicants ensures that workers' probabilities of promotion are reduced by collusion on low effort. When workers collude to shirk, they increase the probability of an external appointment. If the differences in the probability of promotion with and without shirking are sufficiently large, external recruitment may actually assist the monitoring efforts of firms and reduce moral hazard by ensuring that there is a unique equilibrium where all team members choose the high-effort strategy.

Recent work has also assisted in clarifying the precise cause of the higher levels of effort resulting from participation in tournaments. Tournaments may induce higher levels of effort because of:

- Selection effects: higher pay attracts higher quality competitors, so the bigger the prize, the more each competitor will expect that higher performance is needed to win the tournament.

- Competition effects: competition (the "thrill of victory") alone stimulates higher performance, even in the absence of monetary rewards for the winner.

- Pecuniary incentive effects: each participant in the tournament calculates the marginal cost-benefit of winning, so the higher the prize, the greater the effect on performance.

Coffey and Maloney (2010) show that the tournament model (the pecuniary incentive effect) has predictive power, even when it is possible to control for selection and competition effects. However, they find that all three effects have an impact on performance.

8.3 Production in teams

Academic interest in the incentive and performance aspects of working in teams has been driven by the substantial growth in the number of firms organising problem-solving and analytical workers in teams (Lazear and Shaw, 2007). Overall, the literature suggests that key tasks in large firms are increasingly assigned to problem-solving experts working in team-based environments, and that large firms are currently characterised by increasing wage dispersion as the remuneration of "star" workers with problem-solving skills most

finely attuned to the team environment increases faster than the median employee (Andersson *et al,* 2006; Autor *et al,* 2003).

The use of teams is of particular interest because they come with two obvious problems: teams are time-consuming to organise and coordinate, and individual performance in teams is difficult for those outside the team to measure, creating the potential for free-riding by low-productivity members of a team. Consequently, there is considerable value in understanding the conditions under which the productivity advantages of teams outweigh these disadvantages.

The literature has identified a range of conditions under which the use of teams will increase productivity. Firms use teams when:

- Hierarchical decision-making is either unnecessary or less effective in solving the relevant class of problem. Teams are often viewed as an alternative to hierarchy in the sense that hierarchy implies sequential consideration of decisions, whereas teams suggest simultaneous and interactive consideration of decisions.
- The cost of hiring staff who embody all of the necessary knowledge and skills to solve problems is very high, but the problem can be addressed by teams of lower paid staff with complementary skills across the required range. In a team environment, the inputs of staff interact multiplicatively, so that each worker's marginal product is enhanced by combining effort with workers with different skills.
- They need to solve complex problems quickly. Teams work when they are delegated the authority to solve a problem, and this is most effective in addressing new issues where the firm has less sunk investment, because the risks associated with a poor decision are lower than those associated with an existing line of business.

This literature raises some potentially interesting questions about the current human resource practices in the public sector. First, the use of teams (both within departments and across departments) appears to be widespread, but is this consistent with the view that teams work best in an environment of delegated authority? Decision-making in the public sector appears still be to be very hierarchical, in part because of the necessity of ministers making the final decision, and in part because of the culture of risk-aversion that is so central to modern media-focused politics. Second, how good is the public sector at constructing teams that actually have complementary skills? Third, can the approaches to recruitment and remuneration in the public sector facilitate the dispersion in remuneration that is necessary to attract and retain the staff with the most high-powered problem-solving skills in a team environment?

8.4 Tournaments and promotions in teams

While the literature on tournaments provides strong support for the proposition that incentives to increase the value of the firm result from providing differential reward for differential effort, the application of this result in teams is less clear-cut. In teams, individual effort may not be observable and the output of each team member may be complementary to that of other team members (each member's output is increasing the output of other team members). This has led to a presumption that rewarding all team members equally may be optimal. The arguments marshalled in support of equal rewards include lower transaction and monitoring costs, increased peer pressure for equal contributions within the team, and avoidance of moral hazard problems (for example, Milgrom and Roberts, 1992).

However, the presumption for equal treatment of team members has long been questioned. Lazear (1989) observed that it was far from obvious that equal pay had these effects, in part because equal pay does not account for heterogeneity in agents' ability and performance (that is, all team members will not be of equal ability even if the manager cannot identify the differences in ability), and the lack of a link between pay and the individual's marginal product can provide incentives for free-riding. Recent work has provided a coherent theoretical framework for Lazear's observations.

Winter (2004), for example, outlined a model of teams in which the maximum level of efficiency is achieved when differential (not symmetrical) rewards are applied to individual team members with complementarity between the outputs of individuals. In contrast, the maximum levels of efficiency are achieved when symmetrical rewards are applied to teams where there is substitutability between the outputs of individuals. Georg and Kube (2010) provide experimental evidence to support this observation.

The intuition behind the findings of these authors is that asymmetric rewards facilitate coordination in teams where outputs are complementary, because the fact that one team member stands to get a large reward means that at least one team member will exert substantial effort. Where effort is complementary, the other individuals anticipate this high level of effort from the highly paid member, and exert higher levels of effort, anticipating that other members of the team will react in the same way. Conversely, with symmetric payoffs team members anticipate that other team members will not exert a high level of effort, and they choose low effort.

In the case of teams where outputs are substitutes and rewards are unequal, all team members expect the team member receiving the lowest reward to shirk. As a result, all team members reduce effort. However, equal rewards result in each team member expecting that the other team members will exert effort.

The general application of this literature is to say that the design of teams and the reward structures may influence the level of output achieved. Unequal rewards for members of teams (for example, if they have different levels of bonus for the same level of output) may not reduce efficiency and will promote efficiency compared with equal rewards in cases where effort in the team is complementary. This appears to be relevant to the public sector in New Zealand, where, as Scott (2001:27) notes, there has been a tendency to argue that individual performance bonuses (as one type of differential reward) might be counter-productive in an environment where individual performance is not easily determined and teamwork is essential for organisational performance.

8.5 The structure of compensation and organisational performance

The most basic choice of remuneration structure faced by a firm is the positioning on the continuum between fixed salary and pay for performance (piece rate remuneration such as commissions for sales being an extreme example). A range of studies have looked at the conditions under which pay for performance may be superior to salary, and identified that:

- Firms will pay for performance when it is cheaper to measure performance. Even where there are elements of performance that are difficult to measure in the short-term, compensation schemes that involve combinations of short-term and long-term

bonuses, or penalties based on *ex post* evidence of poor performance, may be optimal (Horstmann, Mathewson and Quigley, 2005).

- Firms are more likely to pay for performance as the value of a worker in alternative employment approaches their value in their existing employment. If workers with alternative high-paying job options are not paid for what they produce in their current firm, then they are more likely to move to the alternative employment.

- Firms are more likely to pay for performance when they do not have good screening technologies available to identify the most promising potential employees from job applicants.

- Firms are more likely to pay for performance when hiring new workers is relatively low cost, and when it is therefore cost effective to focus investment in performance measurement and management rather than on screening of job applicants.

The structure of compensation packages has two effects. The first is on incentives for action in the short term, and the second is on the selection of staff for the firm (for a summary, see Lazear and Shaw, 2007:100). In the long term, the second effect may be at least as important as the first, because it means that the compensation system will attract those who are best able to respond to the incentives that the firm provides. A wide variety of studies has demonstrated that performance improvement resulting from performance pay arises both from incentives and the selection of staff whose productivity will respond most strongly to those incentives (for example, Lo, Ghosh and Lafontaine, 2011).

More generally, this literature supports the proposition that organisational transformation will require complementary changes in human resource management practices, particularly those relating to remuneration, recruitment and retention. Higher levels of performance or a focus on different types of performance can be obtained by introducing new approaches to human resource management. Moreover, those new approaches are most effective when a complementary set of changes is introduced—for example, team-based environments have higher output when workers are better trained, incentives are team-based, and the selection of team members ensures that individuals in the team have complementary skills.

8.6 Conclusion

The focus of the public sector on providing services to ministers and to the public means that performance in public-sector delivery is very difficult to quantify and therefore involves subjective judgment in its measurement. The public service probably differs from the private sector in that there is lower volatility in the demand for its services. In both circumstances, internal labour market tournaments have advantages: they do not require absolute measurement of performance, but utilize relative rankings instead, while lower volatility means that variations in performance may more certainly be ascribed to productivity differences.

While tournaments provide an efficient means of allocating labour market resources to different positions, their value is increased by the selective introduction of external contestability for positions. Contestability of positions reduces the potential for internal tournaments to be driven by negative or unproductive aspects of labour market culture

such as collusion on low effort, sabotage and narrow views of what constitutes good performance.

The literature in personnel economics suggests that if the public sector aims to achieve a substantial change in its ability to provide ministers with new advice, or to increase productivity, then this will require changes to the structure of compensation and reward both to change incentives and to change the type of people that are attracted to public-sector organisations. As the importance of team-based environments within the public sector increases, the literature suggests that more attention will need to be given to staff training, team-based incentives, the remuneration of those who have the strongest team leadership skills and the ability of the public sector to ensure that individuals in the team display a diversity of complementary skills.

Identifying exactly how to implement those changes to organisation and remuneration is beyond the scope of this paper. Neither have we investigated whether consideration of such changes to public-sector remuneration and employment structures is currently under consideration. But our observation is that if the government wishes to obtain different types of advice, and have alternative types of policy development skills at its disposal, then that work will need to be done.

9 Coordination between public-sector organisations

9.1 Introduction

A key issue for the public sector in New Zealand is the need to generate greater coordination among a large number of public-sector agencies. The difficulties of generating innovative high quality policy advice on the big issues facing government, when that advice necessarily requires input from multiple public-sector organisations, is a theme in two recent reviews of the effectiveness of policy advice in New Zealand (Gill 2010; Scott *et al*, 2010). The tension has been characterised as being between the requirement for vertical lines of accountability (as established under the Public Finance Act and the State Sector Act) required for the machinery of the public sector to operate effectively, and the requirement for leadership and accountability across multiple public-sector organisations that is required to obtain work on the big, long-term issues facing New Zealand. Scott *et al* (2010:58-62) note that the coordination of policy development across multiple small public-sector organisations may be no more difficult than the coordination of policy advice across the divisions of a much smaller number of large organisations. They recommend that these issues be resolved by establishing a more direct mechanism through which policy advice can be driven by existing institutions (Cabinet Strategy Committee and the central agencies).

Much of the academic literature addressing the issue of coordination is in the context of consideration of a choice between a "divisional structure" and a "functional structure," or a centralised or a decentralised allocation of residual decision rights within a hierarchical structure. Much of this literature can be applied to any situation in which organisations can either assign the full responsibility over a set of projects to individual groups or alternatively require groups to cooperate with one another in the execution of projects. To the extent that this literature focuses on the issue of allocation of residual decision rights, it takes a rather different perspective from the focus on accountability, monitoring and performance management that motivated the new public management and recent analyses of the New Zealand public sector.

We begin by considering the literature on divisional and functional structures, and then consider the literature on decision rights in hierarchies, before concluding with some suggestions about the ways in which an incomplete-contracts decision-rights perspective may be applied to thinking about coordination issues in the public sector.

9.2 The costs and benefits of divisional and functional organisational forms

Under the divisional structure, also known as the "M-form," the firm is organised as a collection of self-contained divisions, each of which has full responsibility over a subset of projects, and needs to perform all tasks associated with these projects (eg, production, marketing, finance, human resources, R&D, etc.). Under a functional structure, also known as the "U-form," the firm is organised as a collection of functional departments, each of which specializes in one task and performs it on all projects that the firm undertakes. Therefore, under the functional structure, each project is executed by a team of agents who belong to different functional departments. The study of organisation designs was pioneered by Chandler (1962), who argued that as firms like DuPont, General Motors, Sears, and Standard Oil grew and adopted more diverse product lines, the difficulties in coordinating functions across product lines induced them to switch from

the functional structure (U-form) to the divisional structure (M-form). Chandler concluded that a firm's structure follows its strategy which determines the number and types of its product lines.

Several recent papers examine the choice between the functional and the divisional structures. These papers consider a firm that produces two goods which require two tasks each. The divisional structure corresponds in this framework to the grouping of agents according to products, whereas the functional structure corresponds to the grouping of agents according to tasks.

Aghion and Tirole (1995) consider a model in which the functional structure requires agents to specialize in specific tasks and hence economizes on the cost of training agents, but the divisional structure strengthens the agents' incentives to exert effort by generating better external signals about their talent. They show that as managerial work load increases, the divisional structure becomes more attractive relative to the functional structure since then, the manager relies more often on the agents' decisions and this improves their ability to signal talent to the external job market.

In Rotemberg (1999), the firm can better control agents who perform the same task under the functional structure, but cross-task coordination is more efficient under the divisional structure. He shows that the divisional structure dominates the functional structure when the number of employees is sufficiently large.

Qian *et al* (2006) consider a model in which the divisional structure eliminates the need for costly cross-division communication to coordinate tasks, but the functional structure economizes on the cost of coordination by coordinating tasks on a company-wide basis. The divisional structure is particularly attractive in their model when there is a need for local experimentation of uncertain innovations involving several tasks; such experimentation is inefficient under the functional structure due to the need for costly communication among different divisions that engage in different tasks.

In Maskin *et al* (2000), the functional structure exploits economies of scale by grouping similar tasks in the same division, but the divisional structure provides better incentives because it promotes yardstick competition among similar divisions.

Besanko *et al* (2005) focus on the role of risk aversion: under the divisional structure, the compensation of agents depends only on their own (risky) performance, whereas under the functional structure it also depends on the (risky) performance of other agents. Hence, agents must receive a larger risk premium under the functional structure in order to induce them to exert the same level of effort. This result may be reversed, however, if there are significant asymmetries in the contribution of the tasks to profits, or significant positive externalities across tasks.

Corts (2007) considers a model with two possible configurations of tasks: "individual accountability" (which is akin to the divisional structure), where each agent is compensated on the basis of a single (noisy) performance measure that depends only on the agent's own effort, and "teams" (which is akin to the functional structure), where compensation is based on two (noisy) performance measures which depend on the agents' joint effort. Individual accountability has the advantage of compensating the risk-averse agents on the basis of only one noisy performance measure rather than two. The disadvantage of individual accountability is that the firm uses a single performance measure to evaluate the two tasks that each agent performs, whereas under teams it uses two performance measures.

In Harris and Raviv (2002), the comparison between the divisional and functional structures depends on the likelihood that various cross-task interactions will be realized, as well as on the CEO's cost of co-ordinating company-wide interactions between all tasks. For a wide range of parameters, both the functional and the divisional structures are dominated by either the matrix form where each task is coordinated by two different middle managers, or by a flat hierarchy where only the CEO may coordinate cross-task interactions.

Berkovitch *et al* (2010) consider a firm that consists of a board of directors, a manager, and two agents (mid-level managers, business units or simply employees). Their theory emphasizes the interaction between organisational structure and investment decisions. The manager's role is to select projects and recommend them to the board of directors. If the board accepts the manager's recommendation, the two agents perform tasks like production and marketing on each project. In this setting, the divisional structure corresponds to the case where each agent gets the full responsibility over a subset of projects and performs all tasks on these projects, whereas the functional structure corresponds to the case where each agent specializes in one task and performs it on all selected projects.

In Berkovitch *et al,* the selection of projects by the manager is subject to a moral hazard problem: the manager may have a personal preference for expensive projects, even if more profitable projects available. They show that for a given set of selected projects, the divisional structure is more efficient *ex post* because it enables the firm to offer each agent an incentive contract that ties his compensation directly to his performance. In contrast, under the functional structure, there is a "moral hazard in teams" problem (see Chapter 8) because each project requires the joint effort of two agents. However, the *ex post* inefficiency of the functional structure may render expensive projects unprofitable and hence it may deter the manager from recommending them to the board of directors. Hence, the optimal organisation structure is determined by trading-off its effect on the *ex ante* selection of projects and its effect on the *ex post* implementation of selected projects.

Berkovitch *et al* show that, relative to firms with the functional structure, firms with a divisional structure have less restrictive standards for project evaluation, they adopt more projects, their projects are more likely to succeed and have a higher variance of gross returns, and they pay a higher expected compensation to their agents. They also show that the functional structure is more likely to dominate the divisional structure when (i) expensive projects require a larger initial investment, (ii) conditional on success, projects yield a smaller return, and (iii) the firm's technology exhibits weak economies of scope and strong economies of scale. In addition, Berkovitch *et al* examine how the overall profitability of the divisional and functional structures changes when firms grow and can adopt more projects, when projects become more complex and require more tasks, and when the tasks have asymmetric effects on the probability that projects will succeed.

Berkovitch *et al*'s main insight is that organisational structures which appear to maximize firm value *ex post*, may not be optimal once managerial incentives are taken into account; in many cases, it is optimal to put in place an organisational structure that appears to be *ex post* inefficient in order to restrict the management's ability to manipulate investment decisions in the direction it likes. The idea that a firm may wish to commit itself to an *ex post* inefficient structure in order to enhance *ex ante* efficiency emerges from a model in which the choice of projects to invest in is important. *Ex post* inefficiency of the functional structure could actually induce the firm's management to improve its selection of projects *ex ante*.

9.3 Incomplete contracts and hierarchies

The incomplete contracts approach to issues of coordination focuses on the allocation of decision rights within hierarchies. In other words, it views coordination problems as arising from the allocation of residual decision rights rather than the quality of incentive schemes, and it suggests organisational realignment of decision rights as a response rather than increased effort invested in the design and monitoring of performance under incentive contracts.

Hart and Moore (2005) analyse a hierarchical structure characterized by agents engaged in specialization and agents engaged in coordination. By comparison with Aghion and Tirole (1997), they emphasize *ex post* efficiency, in the sense that they explain the role of co-ordinators (senior management) as being to choose the highest value ideas from those generated by the specialists. Co-ordinators are more senior because they must have control of the assets (residual decision rights) necessary to implement the projects that they choose (ie, they cannot be vetoed by the specialists who generate ideas). Hart and Moore show that when the gains to coordination are large, it is optimal for the organisation to be centralized; if the gains to coordination are moderate, then the organisation should be decentralized, and if the gains to coordination are small, then it is optimal for the organisation to be split into several independent firms (2005:678).

In Hart and Holmstrom (2010), it is assumed that each organisational unit (firm) generates two kinds of benefits: surplus which is transferable with ownership, and private benefits which are not transferable. The private benefits encompass job satisfaction, positive links between the work of the firm and personal values, personal loyalty to co-workers and managers, and human capital linked to the particular production technology of the firm. Private benefits may also capture the fact that performance assessments of workers hinge on their ability to add value to their firm, not to other firms.

Comparing separate firms with integration, Hart and Holmstrom show that the chief executives of separate firms have the right balance between private benefits and profits, but they do not take into account their effect on other units, whereas under integration they have the right balance between units, but will put less weight on private benefits. However, some weight will be accorded private benefits in the integrated firm, because the chief executives must take into account the potential for disenchanted workers to shade effort in response to decision-making that is inconsistent with their private benefits, but this weight is less than with a narrowly defined firm.

With a broader range of activities, the firm's workforce will be more heterogeneous, making the chief executive experience less empathy for any given group within the firm. The reduction in the intensity of the contact between any particular group in the firm and the chief executive will reduce the ability of any individual group to persuade the boss to pursue a path that has benefits for one group, but reduces value across the full range of productive opportunities. ". . . asset ownership is the means for acquiring essential control rights, but the underlying reason that such control rights are acquired in the first place is that activities need to be brought together under the authority of one boss in order to accomplish strategic goals . . ." (Hart and Holmstrom 2010:511).

At its most general level, Hart and Holmstrom (2010) may be summarised as saying that if the only issue is integration versus non-integration of different firms, then the potential mistakes associated with each form are orthogonal. Non-integration can lead to too little coordination; one unit may veto coordination, even though it is collectively beneficial.

Alternatively, an integrated firm may lead to too much coordination, characterised by the reduction in private benefits associated with the loss of independence.

9.4 Conclusion

From a variety of different perspectives, and utilising range of models with different assumptions and setups, the literature has focused on the trade-offs between dispersion and coordination. In particular, it has identified the fact that decentralisation has the advantage of harnessing information about the identification of projects, and motivating agents by maximising the private benefits they obtain from the ability to choose projects. However, decentralisation has the disadvantage that agents may choose excessively expensive projects (since yardstick competition between different agents to have their projects accepted is more limited), and may veto projects that require coordination to maximum social welfare because those projects are less personally attractive. This is important because, no matter how efficient the implementation, social welfare will not be improved, if the wrong projects are often being implemented under the decentralised structure.

Decentralisation will be least costly when the moral hazard problem in the selection of projects is low, and the path of technical change is uncertain (so that allowing different divisions to pursue their own strategies provides benefits of diversification). It is also advantageous for services that are, or could be, commercially provided, since in that case it may be possible to introduce competition for state-provided services that will promote efficiency in the use of resources. Integration has the advantage of providing a mechanism for controlling moral hazard in the selection of projects, and providing senior managers with the ownership rights to implement some of the good ideas identified by individual workers or teams. Integration also has the advantage that the chief executive of the integrated firm will have less personal empathy with individual divisions or projects, thus more effectively focusing resources on those projects proposed by individual units that have the highest potential to increase value.

In the public sector, this may be translated as suggesting that specialised departments and agencies are better at identifying projects (have more local knowledge), worse at selecting which projects to invest in, and more efficient in implementing the projects that are chosen. If the identification of projects *ex ante* is not critical (for example, if this is driven by ministers, rather than by the specialised knowledge of civil servants), but decisions about which projects to invest in are critical, then large hierarchical government departments with multiple divisions may be preferred. A model in which different agencies have ownership rights in respect of different projects and issues is likely to be the worst of all possible worlds (providing neither the benefits of coordination nor the benefits of decentralisation).

In general, centralisation of decision-making is preferred when the gains to coordination are large. The more the public sector focuses on policy, as opposed to service delivery and asset ownership, then the more likely it is that coordination will be important. In the New Zealand public sector, coordination of performance and resource allocation by central monitoring agencies has proved difficult, because those monitoring agencies are not residual claimants in the activities that they are charged with monitoring. Thus it is likely that a smaller number of larger public-sector organisations will be preferred, because they may provide individual units with incentives to develop specialised expertise and collect information, while also providing high level coordination of the investment and resource allocation decisions made.

10 Conclusions

10.1 Incomplete contracts and the organisation of the public sector

The new public management of the 1980s was based in part on a range of important new insights developed by academic economists over the preceding 15 years. Those insights were based on thinking about transaction and agency costs as central to defining the boundaries of the firm and the governance relationships within them, separation of purchase and ownerships interests, the benefits of competitive delivery of commercial services that had previously been provided by government monopolies. While the relevance of most of those ideas has not been overturned, they have been supplemented in modern academic analysis by a range of alternative perspectives, some of which imply quite different approaches to public policy, to public-service provision and to the shape and organisation of the public sector.

In particular, we have emphasised that in the last 25 years a theoretical literature has developed that argues that contractual incompleteness affects the boundaries of the firm through allocations of ownership rights. From this perspective, firm boundaries define the allocation of residual control rights, and in the presence of (pervasive) contractual incompleteness, the arrangement of these boundaries is critical for the efficiency of the organisation of economic activity. The ability to exercise residual control rights increases incentives to make relationship-specific investments by improving the *ex post* bargaining position of an asset owner. Asset ownership should therefore be assigned to those with the potential to make the most important (value-enhancing) relationship-specific investments.

As an example, consider the identification of privatisation with competition under the NPM. This relationship is somewhat misleading because in principle it is possible to have several public-sector organisations competing with each other to supply public services or several management teams competing to run a public enterprise. It is also, of course, possible to have a private firm that is a monopoly. The difference between private and public-sector ownership is therefore not able to be characterised simply as the difference between competition and monopoly. Rather, the importance of public and private ownership results from the fact that they represent allocations of residual decision rights and control over all of those aspects of any activity which cannot (efficiently) be specified explicitly in a contract.

Contractual incompleteness has a wide variety of applications, which include:

- Flexibility in the timing and nature of investment (real options).
- The boundaries between the public and private sectors, both in respect of investment in and management of public facilities, and in the allocation of responsibility for the outcomes of public-service delivery.
- Governance arrangements, the institutional organisation of the state sector, and the balance between decentralisation and specialisation of agencies and central coordination.

- The literature on contractual incompleteness assists in understanding a variety of the problems that have arisen with the development and implementation of policy under the current structures within the public sector, including:

  - The challenges associated with obtaining public benefits from private finance (or public-private partnership) initiatives, especially those arising from the costs of contracts that remove flexibility in the timing and implementation of projects, and the incentives for private investors to make the optimal levels of investment in cost savings and quality enhancements.
  - The difficulties of maximising the investment decisions made across the very large number of entities in the public sector, giving that decision rights are very dispersed, central monitoring agencies lack residual decision rights that would give them the incentives to use the performance-monitoring information collected, and governance boards of at least some Crown entities do not have effective residual decision rights consistent with quality investment decisions.
  - The reasons why delegation of public service delivery to private or community entities may result in superior outcomes by making the entities delivering the services the residual claimants in the success of the relationship with the consumer of the services.

10.2 Which elements of the NPM of the 1980s require the most substantial reconsideration?

While many of the theoretical foundations that underlay the development of the NPM of the 1980s continue to be accepted as central to microeconomics, the addition of the incomplete contracts perspective does provide insights that may take policy and public-sector organisation in directions that are different from those pursued over the past 25 years. It is beyond the scope of this paper to suggest specific policy developments based the on theoretical perspectives outlined in this paper. As we have emphasised in discussing different sections of an analysis, the literature on incomplete contracts suggests new ways of thinking about problems, but does not provide policy templates that can be applied in a simplistic way: the value of the approach is the insight that it provides in ways of thinking about and analysing complex organisational, decision and investment problems. With those caveats, however, the analysis provided in the paper does identify some areas in which the conclusions derived from an incomplete contracts approach are likely to lead to some different policy directions from those developed as part of NPM:

- The split between purchase and ownership, and between policy and service delivery, which are based on theories of agency rather than theories of ownership, investment and decision-making.

- The boundary between the private sector and the public sector. An incomplete contracts perspective suggests that the definition of this boundary is more complex than the NPM implied, and that the fundamental decision about the boundary should be based on attempts to achieve the optimal allocation of decision and investment rights, since these are the core of ownership rights.

- Public and private/community ownership of assets and of service delivery, when full or partial government funding of services occurs. While not detracting from the efficiency benefits of competitive private delivery, where that can be achieved, an

incomplete-contracts approach focuses attention on the importance of allocating ownership rights to the party who has the potential to add the most value to the asset or to the customer relationship, and would focus less on the benefits of private delivery arising from competition.

- The proliferation of public-sector entities, and the dispersion of information on, and ownership of, different areas of service delivery and policy development across a very large number of organisations would be reassessed against criteria focused on the effectiveness of investment decision-making.

- The governance and monitoring of state-sector entities would be reassessed based on ability to allocate the residual decision-rights which are necessary to make governance and monitoring arrangements effective.

10.3 What would a 21st century public sector look like?

As we have noted above, the literature surveyed in this paper does not establish a definitive policy prescription for the organisation of the state sector. Rather, it provides an approach to thinking about state-sector organisation, and suggests directions (as opposed to definitive endpoints) in which policy on state-sector organisation is likely to move, if this approach was adopted as the basis for more detailed policy developed. Those policy directions might result in a 21st century public sector being structured around:

- Fewer public-sector organisations with specialised units in each organisation collecting information, but fewer points at which that information is coordinated and investment decisions are made. This is likely to promote further thinking about the use of contestability in policy advice and service delivery both within and between large public-sector organisations.

- Greater clarity in monitoring, with only one central monitoring agency being given "ownership" of public-sector performance. Where Crown entities have governance boards, then monitoring by ministries might be replaced by clearer and stronger accountability for the board in both monitoring and reporting to the minister on performance.

- A clearer distinction between organisations where all decision-making can be delegated to an entity whose objectives are defined by statute, organisations where governance can be delegated to a board, and organisations where the minister retains residual decision rights. In the latter case, advisory boards might assist with stakeholder consultation, but any confusion with governance would be removed. The more substantial is the delegation of residual decision rights by the minister, the stronger should be the focus on performance at the level of private-sector boards, with smaller numbers of board members and higher levels of remuneration per member.

- Less public ownership of service delivery, with wider delegation of responsibility for investment in outcomes and customer relationships to private or community service-delivery organisations. This existing model of the use of private general practices to delivery primary health care, and the Whānau Ora initiative, may be expanded into other areas of service delivery.

- Increased levels of private ownership of state infrastructure assets, with the contracts taking advantage of the recent literature on the ways in which contracts may be structured:
  - To recognise the value of real options and the flexibility in the timing and staging of investment, and
  - To incentivise appropriate levels of investment in quality improvement and cost reduction.
- Greater attention to the structure of compensation and reward, both to change incentives and to change the type of people who are attracted to public-sector organisations. This will include a focus on staff training and incentives in teams, the remuneration of those who have the strongest team-leadership skills, the ability of the public sector to ensure that individuals in the team display a diversity of complementary skills, and selective recruitment to senior positions from outside the public sector.

Each of these areas of policy development will of course need to be the subject of more detailed consideration and application of the framework that we have provided before explicit policy recommendations could be provided.

11 References

Adams, Reneé, Benjamin Hermalin, and Michael Weisbach (2010) “The Role of Boards of Directors in Corporate Governance: A Conceptual Framework and Survey.” *Journal of Economic Literature* 48(1): 58-107.

Aghion, P, M Dewatripont, and P Rey (2002) “On Partial Contracting.” *European Economic Review* 46:745-753.

Aghion, P, M Dewatripont, and P Rey (2004) “Transferable Control.” *Journal of the European Economic Association* 2: 115-138.

Aghion, P, and J Tirole (1995) “Some Implications of Growth for Organisational Form and Ownership Structure.” *European Economic Review* 39: 440-455.

Aghion, P, and J Tirole (1997) “Formal and Real Authority in Organisations.” *Journal of Political Economy* 110(3):564-591.

Alonso-Conde, Ana Belen, Christine Brown, and Javier Rojo-Suarez (2007) “Public private partnerships: Incentives, risk transfer and real options.” *Review of Financial Economics* 16: 335-349.

Alvarez, Luis H R, and Rune Stenbacka (2007) “Partial outsourcing: A real options perspective.” *International Journal of Industrial Organisation* 25: 91-102.

Andersson, Fredrik, Matthew Freedman, John Haltiwanger, Julia Lane, and Kathryn Shaw (2006) “Reaching for the Stars: Who Pays for Talent in Innovative Industries?” NBER Working Paper 12435.

Antelo, Manel, and Lluis Bru (2010) “Outsourcing or Restructuring: The Dynamic Choice.” *International Journal of Production Economics* 123: 1-7.

Antras, Pol (2005) “Incomplete Contracts and the Product Cycle.” *American Economic Review* 95(4): 1054-1073.

Autor, David, Lawrence Katz, and Melissa Kearney (2006) “The Polarization of the US Labour Market.” *American Economic Review Papers and Proceedings* 96(2): 189-194.

Bennedson, Morten, and Christian Schulz (2011) “Arms Length Delegation of Public Services.” *Journal of Public Economics* 95: 543-552.

Bennett, John, and Elisabetta Iossa (2006) “Building and Managing Facilities for Public Services.” *Journal of Public Economics* 90: 2143-2160.

Berkovtich, E, R Israel, and Yossi Spiegel (2010) “A Double Moral Hazard Model of Organization Design.” *Journal of Economics and Management Strategy* 19(1): 55-85.

Bernanke, Ben S (1983) “Irreversibility, uncertainty, and cyclical investment.” *Quarterly Journal of Economics* 98(1): 85-106.

Besanko, D, P Regibeau, and K Rockett (2005) “A Multi-Task Principal-Agent Approach to Organisational Form.” *Journal of Industrial Economics* 53(4): 437-467.

Bolton, Patrick, and Mathias Dewatripont (2005) *Contract Theory*. Cambridge and London: MIT Press.

Brandao, Luiz Eduardo T, and Eduardo Saraiva (2008) “The Option Value of Government Guarantees in Infrastructure Projects.” *Construction Management and Economics* 26: 1171-1180.

Brock, William A, and Lewis T Evans (1996) “Contracts in Continuous Time Asymmetric Information Models: The Importance of Large Information Flows." *Journal of Economic Behaviour and Organisation* 29(3): 523-535.

Brown, Christine (2005) “Financing Transport Infrastructure: For Whom the Road Tolls.” *Australian Economic Review* 38(4): 431-8.

Calzolari, G, and G Spagnoloz (2009) “Relational Procurement.” Unpublished Paper, February.

Chandler, A (1962) *Strategy and Structure: Chapter in the History of the American Industrial Enterprise.* Cambridge, MA: The MIT Press.

Cheah, Charles Y J, and Jicai Liu (2006) “Valuing Governmental Support in Infrastructure Projects as Real Options Using Monte Carlo Simulation.” *Construction Management and Economics* 24: 545-554.

Chen, Kong-Pin (2003) “Sabotage in Promotion Tournaments.” *Journal of Law, Economics and Organisation* 19 (April):119-140.

Chen, Kong-Pin (2005) “External Recruitment as an Incentive Device.” *Journal of Labour Economics* 23(2): 259-277.

Childs, P D, S H Ott, and T J Riddiough (2001) “Valuation and Information Acquisition Policy for Claims Written on Noisy Real Assets.” *Financial Management* 30(2):45-75.

Childs, P D, S H Ott, and T J Riddiough (2002a) “Optimal Valuation of Claims on Noisy Real Assets: Theory and an Application.” *Real Estate Economics* 30(3):415-443.

Childs, P D, S H Ott, and T J Riddiough (2002b) “Optimal Valuation of Noisy Real Assets.” *Real Estate Economics* 30(3): 385-414.

Coase, R H (1937) “The Nature of the Firm.” *Economica* 4: 386 – 405.

Coffey, Bentley, and M T Maloney (2010) “The Thrill of Victory: Measuring the Incentive to Win.” *Journal of Labour Economics* 28(1): 87-112.

Copeland, T, and V Antikarov (2003) *Real Options: A Practitioner’s Guide*. Texere: New York.

Corts, K (2007) “Teams versus Individual Accountability: Solving Multitask Problems Through Job Design.” *Rand Journal of Economics* 38(2): 467-479.

Davis, Kevin (2005) “PPPs and Infrastructure Investment.” *Australian Economic Review* 38(4): 439-444.

Dixit, A K, and R S Pindyck (1994) *Investment Under Uncertainty*. Princeton University Press: Princeton, New Jersey.

Dobbs, Ian M (2004) “Intertemporal Price Cap Regulation Under Uncertainty.” *Economic Journal* 114: 421-440.

Epstein, D, N Mayor, P Schönbucher, E Whalley, and P Wilmott (1999) “The Value of Market Research When a Firm is Learning: Real Options Pricing and Optimal Filtering.” In L Trigeorgis (editor), *Real Options and Business Strategy: Applications to Decision Making*. Risk Books: London.

Evans, Lewis, and Graeme Guthrie (2011) “Price-cap Regulation and the Scale and Timing of Investment.” Working Paper, Victoria University of Wellington. http://ssrn.com/abstract=1826662

Fama, E, and M Jensen (1983) “Agency Problems and Residual Claims.” *Journal of Law and Economics* 26(2): 327-49.

Flyvbjerg, B, M K S Holm, and S L Buhl (2002) “Underestimating Costs in Public Works Projects: Error or Lie?” *Journal of the American Planning Association* 68(3): 279-295.

Flyvbjerg, B, M K S Holm, and S L Buhl. (2005) “How (in)accurate are Demand Forecasts in Public Works Projects?” *Journal of the American Planning Association* 71(2): 131-146.

Garrouste, Pierre, and Stephan Saussier (2005) “Looking for a Theory of the Firm: Future Challenges.” *Journal of Economic Behavior and Organization* 58: 178-199.

Georg, Sebastian, and Sebastian Kube (2010) “Treating Equals Unequally: Incentives in Teams, Workers’ Motivation and Production Technology.” *Journal of Labour Economics* 28(4): 747-772.

Ghosh, M, F Lafontaine, and D Lo (2011) “Delegation and Pay-for-Performance: Evidence from Industrial Sales Force.” Unpublished Manuscript, University of Arizona (April).

Gibbons, Robert (2005) “Four Formal(izable) Theories of the Firm.” *Journal of Economic Behavior and Organization* 58: 200-245.

Gill, Derek and Susan Hitchiner (2010) “Achieving and Step Change.” In Derek Gill (editor), *The Iron Cage Recreated: The Performance Management of State Organisations in New Zealand,* Wellington, Institute of Policy Studies.

Glaeser, E, and A Shleifer (2001) “Not-for-profit Entrepreneurs.” *Journal of Public Economics* 81(1): 99-115.

Grenadier, Steven R, and Neng Weng (2005) “Investment Timing, Agency, and Information.” *Journal of Financial Economics* 75: 493-533.

Grossman, S J, and O D Hart (1986) “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration.” *Journal of Political Economy* 94: 691-719.

Guthrie, Graeme, John Small, and Julian Wright (2006) “Pricing access: Forward-looking Versus Backward-looking Cost Rules.” *European Economic Review* 50(7): 1767-1789.

Guthrie, Graeme (2007) “Missed Opportunities: Optimal Investment Timing When Information is Costly.” *Journal of Financial and Quantitative Analysis* 42(2): 467-488.

Guthrie, Graeme (2009a) *Real Options in Theory and Practice*. Oxford University Press: New York, NY.

Guthrie, Graeme (2009b) “Evaluating Real Estate Development Using Real Options Analysis.” Working Paper, Victoria University of Wellington. http://ssrn.com/abstract=1501946

Guthrie, Graeme (2011) “Uncertainty, and the Trade-off Between Scale and Flexibility in Investment.” Working Paper, Victoria University of Wellington. http://ssrn.com/abstract=1788042

Halonen-Akatwijuka, M (2010) “Organisational Design, Technology and the Boundaries of the Firm.” *Economica* 77(307): 544-564.

Hansmann, H (1996). *The Ownership of Enterprise.* The Belknap Press of Harvard University Press: Cambridge, Massachusetts.

Harris, M, and A Raviv (2002) “Organizational Design.” *Management Science* 48(7):852-865.

Hart, O D (2003) “Incomplete Contracts and Public Ownership: Remarks and an Application to Public-Private Partnerships.” *Economic Journal* 113: C69-C76.

Hart, Oliver, and John Moore (1990) “Property Rights and the Nature of the Firm.” *Journal of Political Economy* 98(6): 1119-1158.

Hart, Oliver D, and J Moore(1999) “Foundations of Incomplete Contracts.” *Review of Economic Studies* 66: 115-138.

Hart, O, A Schleifer, and R W Vishny (1997) “The Proper Scope of Government: Theory and Application to Prisons.” *Quarterly Journal of Economics* 112: 1127-1161.

Hermalin, Benjamin, and Michael Weisbach (2003) “Boards of Directors as an Endogenously Determined Institution: A Survey of the Economic Literature." *Economic Policy Review*, April: 7-26.

Holmstrom, B (1999) “The Firm as a Subeconomy.” *Journal of Law, Economics and Organization* 15: 74-102.

Horn, Murray J (1995), *The Political Economy of Public Administration* Cambridge University Press.

Horstmann, I, F Mathewson, and N Quigley (2005) “Agency Contracts with Long-Term Customer Relationships.” *Journal of Labour Economics* 23 (3): 589-608.

Hsu, Jason C, and Eduardo S Schwartz (2008) “A Model of R&D Valuation and the Design of Research Incentives.” *Insurance: Mathematics and Economics* 43: 350-367.

Hugonnier, Julien, and Erwan Morellec (2007) “Corporate Control and Real Investment in Incomplete Markets.” *Journal of Economic Dynamics and Control* 31: 1781-1800.

Jensen, M (1993) “The Modern Industrial Revolution, Exit, and the Failure of Internal Control Systems.” *Journal of Finance* 48(3): 831-80.

Jensen, Michael C, and William H Meckling (1976) “Theory of the Firm: Managerial Behavior, Agency costs and Ownership Structure.” *Journal of Financial Economics* 3(4): 305-360.

Johnstone, David (2002) “Public-Sector Outsourcing as an Exchange Option.” *Abacus* 38(2): 153-176.

Kim, Jin-Hyuk (2011) “Peer Performance Evaluation: Information Aggregation Approach.” *Journal of Economics and Management Strategy* 20(2): 565-588.

Kort, P M, P Murto, and G Pawlina (2010) “Uncertainty and stepwise investment.” *European Journal of Operational Research* 202: 196–203.

Lai, Rose Neng, Ko Wang, and Yuqing Zhou (2004) “Sale before Completion of Development: Pricing and Strategy.” *Real Estate Economics* 32 (2): 329-357.

Lambrecht, Bart M, Grzegorz Pawlina, and Joao C A Teuxeira (2011) “Outsourcing, Integration and Investment in Costly Capacity.” Working Paper, Lancaster University.

Lazear, Edward P, and Kathryn L Shaw (2007) “Personnel Economics: The Economist’s View of Human Resources.” *Journal of Economic Perspectives* 21(4): 91-114.

Lazear, Edward P, and Sherwin Rosen (1981) “Rank-order Tournaments as an Optimum Labour Contract.” *Journal of Political Economy* 89 (October): 841-864.

Lazear, Edward P (1989) “Pay Equality and Industrial Politics.” *Journal of Political Economy* 97 (June): 561-580.

Lazear, Edward P (2000) “Performance Pay and Productivity.” *American Economic Review* 90(5):1346-1361.

Lo, Desmond, Mrinal Ghosh, and Francine Lafontaine (2011) “The Incentive and Selection Roles of Sales Force Compensation Contracts.” Forthcoming, *Journal of Marketing Research*.

Ludema, Rodney, and Anders Olfsgard (2008) “Delegation Versus Communication in the Organization of Government.” Journal of Public Economics 92: 213-235.

Lulfesmann, Christoph (2007) “On the Virtues of Privatization When Government is Benevolent.” *Journal of Economic Behavior and Organization* 64: 17-34.

Mahajan, V, S Sharma, and R Bettis (1988) “The Adoption of the M-Form Organizational Structure: A Test of Imitation Hypothesis.” *Management Science* 34(10):1188-1201.

Martimort, D (2006) “An Agency Perspective on the Costs and Benefits of Privatization.” *Journal of Regulatory Economics* 30: 5-44.

Maskin, E, Y Qian, and C Xu (2000) “Incentives, Information, and Organizational Form.” *Review of Economic Studies* 67: 359-378.

Martimort, D, and J Pouyet (2008) “To Build or Not to Build: Normative and Positive Theories of Public-Private Partnerships.” *International Journal of Industrial Organisation* 26: 393-411.

Milgrom, Paul, and John Roberts (1992) *Economics Organization and Management*. Prentice Hall: New York, NY.

O’Regan, Katherine, and Sharon Oster (2005), “Does the Structure and Composition of the Board Matter? The Case of Nonprofit Organizations.” *The Journal of Law, Economics and Organization* 21(1): 205-227.

Pennings, Enrico (2004) “Optimal Pricing and Quality Choice When Investment in Quality is Irreversible.” *Journal of Industrial Economics* 52(4): 569-589.

Qian, Y, G Roland, and C Xu (2006) “Coordination and Experimentation in M-form and U-form Organizations.” *Journal of Political Economy* 114(2): 336-402.

Rajan, R, and L Zingales (1998) “Power in a Theory of the Firm.” *Quarterly Journal of Economics* 113(2): 387-432.

Rebitzer, James B, and Lowell J Taylor (2007) “When Knowledge is an Asset: Explaining the Organizational Structure of Large Law Firms.” *Journal of Labour Economics* 25(2): 201-229.

Rotemberg, J (1999) “Process- Versus Function-Based Hierarchies.” *Journal of Economics and Management Strategy* 8: 453-487.

Schleifer, A (1998) “State Versus Private Ownership.” *Journal of Economic Perspectives* 12: 133-150.

Schmidt, Klaus (1996) “The Costs and Benefits of Privatization: An Incomplete Contracts Approach.” *Journal of Law, Economics and Organisation* 12(1): 1-24.

Schwartz, A (1998) “Incomplete Contracts.” in P Newman (ed) *The New Palgrave Dictionary of Economics and the Law* (Macmillan) Volume 2: 277-283.

Scott, Graham (2001) *Public Management in New Zealand: Lessons and Challenges.* Wellington: New Zealand Business Roundtable.

Scott, Graham, Pat Duignan, and Patricia Faulkner (2010) *Review of Expenditure on Policy Advice* (Findings of the Committee Appointed by the Government to Review Expenditure on Policy Advice) Wellington, December.

Scott, Graham, and Peter Gorringe (1989) “Reform of the Core Public Sector: The New Zealand Experience.” Australian Journal of Public Administration 48: 81-92.

Shibata, Takashi (2009) “Investment Timing, Asymmetric Information, and Audit Structure: A Real Options Framework.” *Journal of Economic Dynamics and Control* 33: 903-921.

The Treasury (1987) *Government Management: Brief to the Incoming Government 1987*. Wellington.

Tirole, Jean (2006) *The Theory of Corporate Finance.* Princeton University Press.

Triantis, Alexander J, and George G Triantis (1998) “Timing Problems in Contract Breach Decisions.” *Journal of Law and Economics* 41(1): 163-207.

Turnbull, Geoffrey K (2010) “Irreversible development and eminent domain: Compensation rules, land use and efficiency.” *Journal of Housing Economics* 19: 243-254.

Williamson, O (1979) “Transaction Cost Economics: The Governance of Contractual Relations.” *Journal of Law and Economics* 22: 233-261.

Williamson, O (1999) “Public and Private Bureaucracies: A Transaction Cost Approach.” *Journal of Law, Economics and Organisation* 15: 306-342.

Winter, Eyal (2004) “Incentives and Discrimination.” *American Economic Review* 94(3): 764- 773.

Wulf, J (2007) “Authority, Risk and Performance Incentives: Evidence From Division Manager Positions Inside Firms.” *Journal of Industrial Economics* 55(1): 169-196.

Zingales, L (1998) ‘Corporate Governance.” in P Newman (ed) *The New Palgrave Dictionary of Economics and the Law* (Macmillan) Volume 1: 497-503.