PRODUCTIVITY COMMISSION

THE AUSTRALIAN NATIONAL UNIVERSITY

1998 Industry Economics Conference

Conference Proceedings

Canberra, 6–-7 July 1998

ISBN 1 74037 036 8

Publications Inquiries:
Media and Publications
Productivity Commission
Locked Bag 2
Collins Street East
Melbourne VIC 8003

Tel: (03) 9653 2244
Fax: (03) 9653 2303
Email: maps@pc.gov.au

General Inquiries:
Tel: (03) 9653 2100 or (02) 6240 3200

An appropriate citation for this paper is:
Productivity Commission and Australian National University (PC and ANU) 1999, *1998 Industry Economics Conference*, Conference Proceedings, Ausinfo, Canberra, September.

The Productivity Commission

The Productivity Commission, an independent Commonwealth agency, is the Government's principal review and advisory body on microeconomic policy and regulation. It conducts public inquiries and research into a broad range of economic and social issues affecting the welfare of Australians.

The Commission's independence is underpinned by an Act of Parliament. Its processes and outputs are open to public scrutiny and are driven by concern for the wellbeing of the community as a whole.

Information on the Productivity Commission, its publications and its current work program can be found on the World Wide Web at www.pc.gov.au or by contacting Media and Publications on (03) 9653 2244.

PREFACE

The Australian National University hosted the seventh Industry Economics Conference on 6–7 July 1998. The theme of the conference was *Privatisation, Regulation and Reform*. These conference proceedings publish those invited and contributed papers presented at the conference that are not subject to copyright elsewhere.

The aim of the conference is to bring together leading researchers and policy makers in the field of industry economics to discuss their current work, to examine emerging ideas and methodologies, to establish and extend communication channels and to encourage further research. To this end, the 1998 conference featured papers from invited speakers and contributed by other speakers. The conference organisers were particularly pleased that Professor Joskow of the Massachusetts Institute of Technology was the keynote speaker. A version of his presentation on *Electricity sector privatisation, competition and regulatory reform* was subsequently published in the *Energy Journal* (Joskow, P.L. 1998, 'Electricity sectors in transition', *Energy Journal*, vol. 19, no. 2, pp. 25–52).

Dr John Logan of the faculty of Economic and Commerce at the Australian National University organised the conference. Owen Gabbitas at the Productivity Commission prepared these proceedings.

September 1999

CONTENTS

ABBREVIATIONS AND EXPLANATIONS

Abbreviations

AA	Australian Archives
AAA	Australian Automobile Association
ABS	Australian Bureau of Statistics
ACCI	Australian Chamber of Commerce and Industry
ACIS	Automotive Competitiveness and Investment Scheme
ACT	Australian Capital Territory
ACT	Advanced communications technologies
ACTEW	Australian Capital Territory Electricity and Water
AGPS	Australian Government Publishing Service (now AusInfo)
ANU	Australian National University
ANZSIC	Australian and New Zealand Standard Industrial Classification
APEC	Asia-Pacific Economic Cooperation
ASX	Australian Stock Exchange
BARA	Board of Airline Representatives of Australia
BT	British Telecom
BTCA	British Tobacco Company Australia
DFA	Department of Finance and Administration
DIST	Department of Industry, Science and Tourism
DTF	Department of Treasury and Finance (Tasmania)
EDI	Electronic data interchange
EFS	Export Facilitation Scheme
FCAI	Federal Chamber of Automotive Industries
FMCA	Ford Motor Company of Australia

GDP	Gross Domestic Product
GM	General Motors
GTE	Government Trading Enterprise
IAC	Industries Assistance Commission
IBRD	International Bank of Reconstruction and Development
IC	Industry Commission
IMF	International Monetary Fund
IP	Intellectual property
IR&D	Industry Research and Development
km	kilometres
LAN	Local area network
LDC	Less developed country
LSDV	Least square dummy variable
ML	Maximum log likelihood
NOP	Net open position
NSW	New South Wales
OBM	Output-based management
PC	Productivity Commission
PMV	Passenger Motor Vehicles
R&D	Research and development
RACV	Royal Automobile Club of Victoria
RAG	Roading Advisory Group (New Zealand)
TAB	Totalisator Agency Board
TMCA	Toyota Motor Corporation Australia
UK	United Kingdom
USA	United States of America
Vic.	Victoria
WA	Western Australia
WWW	World wide web

SECTION 1

PRIVATISATION AND REGULATION

1 The equity premium puzzle and the privatisation paradox

*Simon Grant and John Quiggin**

The most promising resolution of the equity premium puzzle observed by Mehra and Prescott (1985) is the suggestion by Mankiw (1986) that capital markets do not spread risk perfectly, and, in particular, that systematic risk is concentrated ex post on a small number of people. There is a close link between this and what may be called the privatisation paradox. That is, the fact that, although privatisation is widely seen as increasing technical efficiency, the savings in public debt are frequently smaller than the forgone earnings of government enterprises. In order to explore this connection, a simple general equilibrium framework is developed in which capital markets operate to spread risk associated with physical capital, but owing to an adverse selection problem, risk associated with human capital cannot be insured. With such imperfect risk-sharing, public investment financed by government bonds can provide indirect human capital insurance benefits. This is because in the recession, in which the public investment fails to generate a return, the revenue needed to pay bondholders can be raised by levying a labour income tax. Hence, the optimal size of the public sector is non-zero. Moreover, the appropriate discount rate for public investments not only lies above the bond rate, but is bounded above by the rate of return to private equity that would be obtained if there were no market imperfections.

1.1 Introduction

Since its discovery by Mehra and Prescott (1985), the equity premium puzzle, that is, the fact that the premium between rates of return to equity and debt is much greater than can be explained on the basis of standard models of life-cycle

* Financial support for this project has been provided by the Australian Research Council's Large Grant A79800678. Quiggin gratefully acknowledges income support from an ARC senior research fellowship. The authors thank Vladimir Pavlov for excellent research assistance.

optimisation, has generated a large literature. Although many candidate resolutions have been offered, the most promising is the suggestion by Mankiw (1986) that capital markets do not spread risk perfectly, and in particular that systematic risk is concentrated ex post on a small number of people. When this idea is incorporated in a model with heterogeneous individuals (Constantinides and Duffie 1996), the results are consistent with the emergence of an equity premium.

Rather less attention has been paid to what may be called the privatisation paradox. That is, the fact that, although privatisation is widely seen as increasing efficiency, the savings in public debt interest obtained through programs of privatisation and debt reduction are frequently smaller than the forgone earnings of government business enterprises, at least for developed countries such as Australia, New Zealand and the United Kingdom (Quiggin 1995).

There is a close link between the puzzle and the paradox. Because the rate of return expected by holders of private equity is significantly greater than the rate of return to good quality public or private debt, the market value of an asset is significantly less than the present value of its expected future earnings, capitalised at the bond rate. Conversely, the annual saving in public debt interest associated with the sale price is less than the earnings forgone.

The privatisation paradox, in turn, may be linked with the debate over the appropriate rate of discount for risky public projects (Arrow and Lind 1970, Hirshleifer 1989). Reasoning similar to that of Mehra and Prescott may be used to support the view that only a small risk premium should be charged for public projects, and therefore that the appropriate rate of discount for public projects is close to the bond rate. On the other hand, in the absence of market imperfections, the discount rate for public projects should be the same as the rate of return on comparably risky private investments.

For representative investments, the private rate of return is well above the bond rate. If the equity premium is the result of imperfections in the private capital market, there is a prima facie case to suggest that the appropriate rate of discount for public sector investments is that which would be generated by a perfect capital market, rather than the observed rate incorporating the anomalous equity premium.

On the same basis, it may be argued that privatisation of a given enterprise increases welfare, if and only if the valuation of the enterprise generated by private capital markets exceeds the expected value of future earnings discounted at the socially optimal rate. This in turn will be true, if and only if gains in efficiency arising from privatisation outweigh the excessive cost of capital associated with the equity premium. However, it may be argued that acceptance of this view would imply

support for obviously inappropriate policies, such as comprehensive public ownership.

In view of the privatisation paradox, it is natural to ask whether public sector net wealth and social welfare are reduced as a result of privatisation. Defenders of privatisation, relying on the (implicit or explicit) assumptions of perfect capital markets and Ricardian equivalence, have argued that the apparent reduction in public sector net wealth arising from privatisation is illusory (Domberger 1995, Forsyth 1995). The argument is that risk borne by governments must ultimately translate into individual risk concerning tax liabilities. This is a strong form of Ricardian equivalence, since individuals must take account of the impact of government decisions, not merely on the expected value of future tax liabilities, but of the state-contingent distribution of those liabilities.[1]

The Ricardian equivalence argument would be convincing if the equity premium could be explained in a manner consistent with the assumptions of capital market perfection. If, however, a substantial portion of the equity premium is explained by capital market imperfections, it is necessary to reassess the implications of public sector holdings of risky assets for individual welfare.

The purpose of this paper is to undertake such a reassessment.

1.2 The equity premium puzzle

Long data series generally show that the rate of return to buying and holding the market portfolio of stocks is considerably greater than the rate of return to government bonds. For example, Mehra and Prescott (1985) present data showing that, over the period 1889–1978, the average annual yield on the Standard and Poor 500 Index was seven per cent, while the average yield on short-term debt was less than one per cent. Using a simple model of intertemporal optimisation of consumption, and evidence on the growth and variability of aggregate consumption, Mehra and Prescott compute equilibrium asset prices for debt and equity under a wide range of parameter values. They show that the equity premium should be no more than half a per cent.

The Mehra-Prescott argument may be expressed more simply in terms of the analysis of Grossman and Shiller (1982) and Grossman, Melino and Shiller (1987). Suppose r denotes the return of a riskless asset. Then, either by taking a log-linear

[1] This argument is concerned with pure risk and should be distinguished from the observation that the average profits of government business enterprises may be overstated as a result of failure to make an actuarially fair allowance for the cost of contingent guarantees.

approximation or assuming asset returns and per capita consumption are jointly log-normally distributed, it may be shown that in an efficient capital market, the equity premium, denoted by ρ is (approximately) given by:

$$E[\rho] = \sigma \operatorname{cov}(\rho, \Delta \log C) \qquad [1.1]$$

where $\Delta \log C$ is the rate of growth of aggregate consumption. The term $\operatorname{cov}(\rho, \Delta \log C)$ plays essentially the same role as the beta coefficient in the capital asset pricing model, measuring the systematic risk associated with the asset in question, while σ may be interpreted as the coefficient of relative risk aversion. Observe that no premium is associated with idiosyncratic risk, that is with risk that is uncorrelated with aggregate consumption.

The coefficient of variation of $\Delta \log C$ is around 0.03 in most OECD countries, including Australia and the United States. Estimates of σ based on direct elicitation of risk preferences are typically around 1.[2]

To approximate the expected rate of return to any given asset only requires knowledge of the standard deviation of the rate of return for that asset and the correlation between returns and aggregate consumption. For example, the standard deviation of the rate of return to the market portfolio of equities in the United States is about 20 per cent, and the correlation with aggregate consumption is about 0.33. This implies that:

$$\operatorname{cov}(\rho, \Delta \log C) = 0.33 \times 0.20 \times 0.03 = 0.002$$

so that for $\sigma = 1$, the implied premium over a riskless asset is about 0.2 per cent.

Mehra and Prescott coined the term ‘equity premium puzzle’ to describe the discrepancy between the observed equity premium and predictions derived from a standard model of intertemporal optimisation. The observed data constitutes a ‘puzzle’ because it seems to suggest that individual investors are not rationally optimising and also that there are unexploited opportunities for arbitrage. Risk aversion in a complete markets setting does not seem an adequate explanation — although individual shares are risky, diversification should reduce risk greatly.

Moreover, if investors were sufficiently risk averse to account for the observed equity premium, their desire to ‘smooth consumption’ across states (because of their aversion to risk) would also imply a strong desire to smooth consumption over time

[2] Estimates based on observations of labour supply tend to be smaller. Some larger estimates have been derived from financial market data, but these are derived from solving for σ on the assumption that a relation like [1.1] holds. They cannot be used to test whether [1.1] does in fact hold.

(an ‘aversion’ to non-constant, including increasing consumption profiles). But this is then inconsistent with low government bond returns and savings rate sufficient to generate the observed per capita consumption growth of around two per cent over this period. This is what Weil (1989) dubs the ‘risk free rate puzzle’.

Attempts to resolve the equity premium puzzle have fallen into three main classes. First, there have been claims that the economy is riskier than the Mehra-Prescott model would suggest, even though the data period for the model includes the Great Depression. Second, there have been arguments that the structure of preferences may be different from that assumed by Mehra and Prescott. Finally, there have been explanations based on imperfections in capital markets.

The first approach is developed by Rietz (1988), who argues that the equity premium may be explained by consideration of low-probability economic catastrophes. This explanation is dismissed by Mehra and Prescott (1988) who observe that, among other points, such catastrophic events frequently involve the expropriation of the wealth of bondholders, either through repudiation or through unanticipated inflation. The same point may be made about the observation that the data presented by Mehra and Prescott ignores stock markets that have disappeared completely, such as the Russian stock market in 1917. Once again, bondholders fare no better than stockholders in cases of this kind. Whatever the significance of risks of this kind in assessing the desirability of financial assets, as opposed to, say, gold, they are irrelevant in considering the relative prices of equity and bonds. Attempts to explain the equity premium in terms of the risk characteristics of the economy appear to have little promise.

The second approach is to consider different preference structures. Epstein and Zin (1990) observe that, whereas in expected utility models, aversion to risk and aversion to intertemporal variations in consumption are both determined by the curvature of the utility function, this link is broken in more general models, such as rank-dependent expected utility (Quiggin 1982). Hence, the sufficiently large within-period degree of risk aversion sufficient to accommodate the equity premium need not entail a low degree of intertemporal elasticity of substitution. However, Kocherlakota (1996) notes that the Epstein-Zin model still can only accommodate the equity premium by requiring levels of within-period risk aversion that economists generally deem implausibly high. Similar arguments apply to the analysis of Constantinides (1990), who relaxes the assumption of intertemporal separability, proposing instead that consumption levels at nearby points in time are complements.

The third approach is based on the observation that capital markets are imperfect in two major respects. First, because of moral hazard and adverse selection problems, individuals and non-corporate firms are unable to fully diversify systematic risks,

such as the possibility of suffering unemployment or bankruptcy during recessions. Second, whereas the perfect capital market is one of costless transactions, individuals face substantial transaction costs, particularly when borrowing to finance consumption.

Mankiw (1986) observes that, faced with undiversifiable background risk, individuals will demand a higher risk premium to bear additional systematic risk than would be the case in a perfect capital market. Kocherlakota argues that this explanation of the equity premium is inadequate because individuals could use intertemporal consumption-smoothing as a substitute for diversification of systematic risk. However, as Quiggin (1998) observes, the existence of transaction costs for borrowers undermines critique of Kocherlakota.

As Kocherlakota concludes, the equity premium remains a puzzle. No single approach has been fully successful. Nevertheless, it appears reasonable to suppose that market imperfections, notably including the unavailability of insurance against income risk, transactions costs of borrowing and other capital market transactions, play a major role in generating the equity premium.

1.3 The Arrow-Lind debate

Arrow and Lind (1970) argued that, in the absence of tax distortions, a 'small' public sector project yielding benefits uncorrelated with aggregate consumption is beneficial, if and only if the present value of expected benefits, evaluated at the real bond rate, is positive. This proposition, which will be referred to as the Arrow-Lind theorem, may be summarised by saying that for projects meeting the stated conditions, no risk premium should be charged. This proposition sparked an extensive debate, which remains unresolved, although both sides have long since declared victory and pulled out.

The central difficulty in the debate was that the two sides argued at cross-purposes. The opponents of Arrow and Lind were not primarily concerned to refute the Arrow-Lind theorem, but to defend the proposition that, in the absence of distortions and capital market imperfections, the risk premium for public projects should be the same as for private projects.[3] Central to the argument is the observation that, in the presence of perfect capital markets, rational individuals should be indifferent

[3] Strictly speaking a simple adjustment to the discount rate is an adequate method of adjustment for risk only under special conditions, see Little and Mirrlees (1991).

between bearing risk directly through financial markets or indirectly as taxpayers.[4] The equivalence proposition is put most clearly by Hirshleifer (1989, p. 111):

> The market rate of interest is generated by an equilibrium between marginal time preferences of consumers and the marginal time productivity of resources. … It is true that in a risky world there are many 'impure' time-plus-risk interest rates rather than one pure time-rate, but the way to take this into account is to use in the public sphere the rate employed for 'comparable' investments in the private sphere.

There is no logical conflict between the Arrow-Lind and Hirshleifer propositions. Indeed, application of the capital asset pricing model yields the conclusion that a private enterprise meeting the Arrow-Lind conditions has a beta of zero and its expected future earnings should therefore be discounted at the real bond rate. Why then, was the debate so heated?

The difficulty is that, while the propositions themselves are logically consistent, policy implications derived from them depend upon the assumption that their conditions are approximately satisfied. But these conditions are inconsistent with the presence of a large equity premium. If most public projects satisfy the Arrow-Lind conditions approximately, in that their systematic risk is small relative to their expected benefits, then the risk premium for such projects should be small also. Computation of a risk premium for a project with returns having a coefficient of variation of 0.03 and plausible levels of risk aversion yields results similar to those derived by Mehra and Prescott, namely, a risk premium well below 1 per cent. On the other hand, if capital markets are nearly perfect, the risk premium for a public project should be the same as that for a private project with similar risk characteristics and, for a typical project, this premium will be large. It can easily be seen, in retrospect, that the conflict between these claims arises from the fact that the observed equity premium is much larger than the premium that would be expected on the basis of standard assumptions about preferences.

Although Arrow and Lind did not address the perfect capital market hypothesis in detail, their remarks clearly indicate that they did not consider it an appropriate basis for analysis. By contrast, both Hirshleifer (1989) and Bailey and Jensen (1972) took the view that prima facie any result inconsistent with the perfect capital markets hypothesis was not an appropriate guide to policy. Hathaway (1997) summarises the position of Bailey and Jensen as follows:

> The argument that governments have access to opportunities for risk diversification that are unavailable to private investors suggest that there is some impediment in risk diversification in the private sector. But there is no logical reason why this is the case, nor is there any evidence that it is.

[4] In an intertemporal setting, note that this requires Ricardian equivalence, where Ricardian equivalence is implied by the joint hypotheses of rationality and perfect capital markets.

The debate remained confused, partly because the problem of the equity premium was not recognised and partly because, in the absence of a well-developed analysis of agency problems, the issue of capital market imperfections could only be addressed in vague and general terms. With the aid of modern agency theory, and in the light of the equity premium debate, it is now possible to provide a framework for analysis of criteria for public investment criteria, based on two key propositions. We begin by saying that the capital market is 'nearly perfect' in risk-spreading if the equity premium it generates is close to that which would arise from a perfect capital market.[5]

With this terminology, the first proposition is that, if the capital market is nearly perfect, the risk premium for public projects should be the same as that for private projects. The converse proposition is weaker. If the capital market is not nearly perfect, the optimal risk premium for public projects will not, in general, be that observed in the private capital market.

A natural conjecture is that whether or not the private capital market is perfect, public investments should be evaluated on the basis of the prices, including risk premiums, that would prevail in a perfect capital market. There are two reasons to doubt this conjecture. The first is a standard second-best argument. If the private sector allocation of capital is distorted, the adoption of the first-best set of public sector projects is unlikely to be desirable. The second objection is that any agency problems that prevent optimal risk-spreading through the capital market may also prevent optimal risk-spreading through the tax system.

These objections are likely to work in opposite directions. Assuming that private and public projects are substitutes, the second-best argument implies that the public sector should take on more risky projects than in the first best. The agency argument implies that the public sector risk premium should be larger than in the first-best and therefore that fewer risky projects should be undertaken. Of course, the fact that the objections work in opposite directions does not mean that they cancel each other out.

It is useful at this point to consider the distinction between moral hazard and adverse selection problems. A common, but not entirely satisfactory way of drawing this distinction is to say that moral hazard problems involve 'hidden action' while adverse selection problems involve 'hidden information'. From a state-contingent choice perspective however, both classes of problems involve hidden information. The crucial distinction is that, in the case of adverse selection, private information is observed before contracting takes place, whereas in the case of moral hazard,

[5] This definition allows for the possibility that deviations from the perfect capital market may be important in other contexts, provided they do not affect the equity premium.

information is observed after contracting takes place. The adverse selection problem arises because individuals whose private information indicates that they are unlikely to benefit from, say an insurance contract, will decline the contract, leaving the insurer with all the bad risks. The moral hazard problem arises because an insured party has an incentive to make a misleading report about the state of nature, and to supply less effort than is required under the contract.

Because of their coercive powers, governments can overcome adverse selection problems. Whereas private unemployment insurance schemes are likely to fail because of adverse selection problems, governments can require everyone to participate. By contrast, governments have no particular advantage in dealing with moral hazard problems. The moral hazard problems that would undermine a private unemployment insurance schemes reappear as adverse incentive effects of tax and welfare payments in the case of government insurance schemes.

1.4 Privatisation and nationalisation

Privatisation is the process of converting a government business enterprise to private ownership. Although some government business enterprises were created within the public sector, most privatisations represent the reversal of previous decisions to nationalise private enterprises. In a project evaluation framework, decisions on privatisation or nationalisation may be seen as a choice between incompatible projects, the project represented by the public enterprise and that represented by the private enterprise. Hence, any general criterion for the evaluation of public and private projects also gives rise to a decision criterion for privatisation and nationalisation.

Private and public enterprises differ both in the streams of benefits to which they give rise and to the way in which those benefits are distributed among members of the community. Although there is little agreement in the literature, there is a majority view that private enterprises will, on average, achieve greater operating efficiency and be more responsive to changes in consumer preferences than will public enterprises in the same industry. This case is strongest for owner-managed firms, where the residual income recipients bear the consequences. Conversely, there has been, at least until recently, a majority view that it is less costly to deal with problems of monopoly and externality though direct control over public enterprises than through regulation of private enterprises.

If both propositions are accepted, it follows that there will exist a spectrum of enterprises. At one end of the spectrum (competitive industries with small-scale firms and no externality problems) the aggregate stream of benefits from private

enterprises will be greater, on average, than that from public enterprises. At the other end of the spectrum (monopoly providers of pure public goods) the reverse will be true.

There remains the issue of how benefit streams from public enterprises should be evaluated, particularly with regard to risk. The analysis here is the same as that presented above for public projects. If the perfect capital market hypothesis is valid, public enterprises should be evaluated in the same way as private enterprises. If the perfect capital market hypothesis is invalid, it is necessary to determine an appropriate risk premium for public enterprises on the basis of second-best social optimality.

The issue can be made more concrete by considering an example of an enterprise where considerations of monopoly and externality are not relevant. If the perfect capital markets hypothesis is valid, a necessary and sufficient condition for privatisation to be desirable is that it should lead to a net improvement in operating efficiency. If the perfect capital market hypothesis is invalid, it is necessary to weigh improvements in operating efficiency against any exacerbation of the consequences of capital market inefficiencies arising from privatisation. If a constrained optimal risk-adjusted discount rate for public projects is known, the necessary and sufficient condition for privatisation to be desirable is that the market value of the firm on privatisation should exceed the present value of expected earnings under public ownership discounted at the optimal public rate.

The consequences for the privatisation debate are significant. As has been shown above, the risk premium that would arise from a perfect capital market is very close to zero, so that the risk-adjusted discount rate is approximately equal to the riskless bond rate. As has been shown in Quiggin (1995, 1996 and 1998) very few privatisations in OECD countries have yielded market prices greater than the present value of expected earnings under public ownership discounted at the optimal public rate. Hence, if the optimal discount rate for public projects is close to the rate that would arise from a perfect capital market, the case for privatisation would be gravely weakened. Conversely, the case for a mixed economy, and possibly for an extension of public ownership, would be strengthened.

1.5 Developing a framework for analysis

To determine the appropriate treatment of risk in public investment, it is necessary to develop a modelling framework within which the private capital market equilibrium is characterised by an equity premium comparable to that observed in reality. Since the problem is trivial if the equity premium arises in a perfect capital

market, assume that the model is characterised by market imperfections. Assuming the existence of a set of possible public projects, there exists a subset of projects consistent with a (constrained) social welfare optimum. The problem is then to determine an evaluation criterion, preferably taking the form of a risk premium, under which only members of the optimal set are approved. Such a criterion will also provide a basis for the assessment of proposals for privatisation and nationalisation.

This section sketches an approach to the analysis, which is an elaboration of that considered by Mankiw (1986). As in Mankiw (1986), there are two global events, recession and boom. The boom is the same for all individuals, but only a subset of the population is affected by recession. Thus, a full specification of the set of states of nature contains a description of the effect of recession on each individual. The critical feature of the model is that individuals cannot fully spread the risk associated with recession and are therefore less willing to hold equity than they would be in a world of perfect capital markets. More formally, the payoff for securities can vary according to the global event (boom or recession), but must be independent of the state experienced by particular individuals. Thus, two securities are sufficient to span the set of possible securities. For simplicity, consider a bond paying 1 unit in each event and a pure equity paying 1 in the boom event and 0 in the recession event.

Consider a model of an economy with two inputs to production, physical capital and human capital (or labour capacity). The production technology and the determination of returns to human and physical capital are not modelled explicitly. There are two types of firms, ‘risky’ and ‘safe’, but no-one in the economy can determine the identity of a firm before uncertainty is resolved in period 1. Hence, the risk associated with being employed by a risky firm cannot be diversified through insurance or other market mechanisms.

Risky-type firms generate higher revenue in booms, but go ‘bust’ and generate no revenue in recessions. Safe-type firms generate the same revenue in booms and recessions and pay a constant amount to their employees and non-employee claimants. In recessions, therefore, returns to physical capital decline by more, on average, than returns to human capital. However, whereas portfolio diversification ensures that the reduction in returns to physical capital is the same for all owners of physical capital, the payoff to human capital is unchanged except for the subset of individuals who become unemployed.[6]

[6] It would be straightforward to allow for differences in risk attitudes at the cost of introducing distributional complications.

With this setup, it is possible to derive a securities market equilibrium and compare it to the first-best state-claims equilibrium arising when the risk associated with returns to human capital can be fully diversified. The difference between the expected rate of return to equity and the bond rate is higher in the second-best securities market equilibrium than in the first-best state-claims equilibrium. With plausible parameters, the equilibrium set of state-claim and security prices incorporates an equity premium comparable to that observed by Mehra and Prescott. Assuming all individuals have identical homothetic preferences yields the result that in the first-best, with the ability to pool idiosyncratic human capital risk, the equity premium is just under one per cent. But without the ability to pool idiosyncratic risks, the equity premium is almost seven per cent.

The argument is most simply presented on the basis of prices for event-contingent claims. For any given individual there are three possible events:

1. boom;
2. recession without job loss; and
3. recession with job loss.

A risk-averse individual will pay more than a risk-neutral individual for claims that yield income only in the third event and less than a risk-neutral individual for claims, such as pure equities, that yield income only in the first event. Provided risk preferences display the standard property of prudence (see Kimball 1990), the value of a security yielding a payoff in states (2) and (3) is greater when income differs between these states than when it can be pooled across the two states.

1.6 Introducing government enterprises

Under the perfect capital market hypothesis, shareholders will face a complete set of state-contingent markets and will therefore be unanimous in desiring value-maximisation. Hence, only the aggregate stream of benefits is of interest. Similarly, any stream of benefits flowing from public projects has a unique market value, independent of its distribution across the community. However, in the presence of capital market failure, the distribution of benefits across the community is relevant. For example, if it is impossible to diversify the risk of becoming unemployed, then the marginal value of consumption in a state of nature where individual *A* becomes unemployed and individual *B* does not will be greater for *A* than for *B*.

In these circumstances, public ownership of enterprises can increase their value by changing the distribution of returns. Consider a government with a balanced budget constraint and a single taxation instrument — a proportional income tax which may

be levied at either positive or negative rates. Assume that in period 0, the government issues bonds and purchases equity at the competitively determined price with the proceeds. If period 1 turns out to be a recession event (that is, the event in which the equity the government holds has a zero payoff), then payment by the government for the bonds it issued in period 0 is achieved by levying a proportional labour income tax on the private sector. Conversely, any additional payoff from the equity the government holds in a booming period 1, over and above that needed to meet its bond payments, is remitted to the private sector by means of a negative proportional labour income tax.

Restricting the government to only levying proportional income tax rates greatly simplifies the analysis in an economy with proportional endowments and homothetic preferences. Another motivation for the restriction to proportional tax rates is to abstract away any explicit redistributive role for taxation.

The use of a proportional tax to distribute the profits and losses arising from publicly owned government business enterprises provides insurance against the losses that would arise from unemployment relative to the alternative of direct private ownership. Hence, other things being equal, public ownership will raise economic welfare.

As usual, other things are not equal. For large classes of enterprises, particularly those in which owner-operation is feasible, government ownership is likely to be associated with a loss of operating efficiency. On the other hand, government ownership may improve efficiency if production externalities or monopoly problems require extensive intervention. The fact that similar industries are found in public ownership in the majority of mixed economies is evidence that there exists a relatively stable ordering of industries, from those in which the costs of public ownership relative to private ownership are greatest (eg agriculture) to those in which the costs are smallest (eg police services).

Therefore, consider a menu of public investment opportunities, with declining marginal returns. If the size of the public capital stock is denoted by G, the marginal investment is assumed to have returns that differ from those of the private sector by a proportional factor $(1-\phi(G))$ in each state of the world, where $\phi'(G) \geq 0$. That is, it is assumed that the projects with the highest rates of return are implemented first. The problem of determining the optimal set of public investments is therefore reduced that of determining the optimal choice of G.

An equivalent, and probably more useful interpretation of the solution arises if it is expressed in terms of the public sector rate of discount. Observing that the expected rate of return to private sector capital r^* is a weighted average of the return to equity and the return to bonds, any solution for G may be expressed in terms of the

rule that public investments should be undertaken, if and only if the rate of return exceeds $(1-\phi(G))r^*$.

From the analysis of the case when public and private investments are equally efficient, it is obvious that the optimum must have $\phi(G)>0$, that is, that the appropriate discount rate for the public sector must be lower than the average rate of return to private capital. On the other hand, a straightforward stochastic dominance argument shows that public investments must have an expected rate of return at least as high as the riskless bond rate.

In terms of the state-contingent analysis presented above, a public investment financed by the issue of debt has a positive payoff in event 1, a negative payoff in event 2 and a zero payoff in event 3 from the viewpoint of a taxpayer-owner. In the absence of differences in the expected rate of return, this is more attractive than ownership of private equity financed by debt which yields an equal negative return in events 2 and 3. On the other hand, if the expected return on the investment were only equal to the bond rate, an investment financed by debt would yield zero expected return, but would reduce income in the unfavourable events 2 and 3 and would therefore reduce welfare.

Moreover, it follows from the state-contingent payoffs that the public sector rate of discount should lie between the bond rate and the first-best rate of return to private equity. This would imply a real rate of discount for public projects no more than 1 percentage point over the real bond rate.

1.7 Extensions

The model described above is designed to be as simple as possible, while showing how an equity premium can arise from the failure of capital markets to spread risk perfectly and how the risk associated with public ownership of capital may be spread more effectively through the tax system. The model could be elaborated in a number of ways.

First, the inclusion of agency problems would also have important implications for the analysis of government policies based on state-contingent taxes. The incentive effects of taxes on individual effort may be seen as analogous to the moral hazard problems associated with insurance against income losses. Such effects would need to be taken into account in analysis of the welfare effects of public ownership of equity. It should be noted, however, that public ownership of equity implies an increase in the state-contingent variability of tax rates rather than an increase in the average rate of taxation. Small increases in the state-contingent variability of tax rates will, in general, have second-order welfare effects.

More generally, there is no explicit modelling of the agency problems assumed to account for differences in the efficiency of public enterprises and private corporations. While it is reasonable to suppose that these agency problems are in the class normally analysed in terms of moral hazard, there is no generally accepted way of modelling the agency relationship between taxpayers and the managers of government business enterprises or between shareholders and the CEOs of private corporations.[7]

Extension of the analysis to take account of moral hazard would also yield a more complete account of the private sector equity premium. Moral hazard problems for individuals could be modelled, as in Grossman and Hart (1982), by assuming that γ, the probability of loss in the global recession event by an individual is not exogenous, but depends on unobservable effort.[8] Kahn (1990), using the Grossman-Hart approach, shows that moral hazard problems alone are not sufficient to explain the observed equity premium. However, the interaction between adverse selection and moral hazard problems might produce a richer set of results. It would also be desirable to allow for agency problems on the part of the managers of firms issuing securities. There is, however, no general agreement on the best way of incorporating such problems in a model of security market equilibrium.

Second, the analysis is based on a single period of consumption and does not permit consideration of the possibility of smoothing consumption over time through borrowing, lending and the liquidation of assets. Kocherlakota (1996) argues that the solution of Mankiw to the equity premium problem is unsatisfactory because intertemporal consumption smoothing would overcome any difficulties associated with the absence of insurance markets. However, Kocherlakota's argument is based on the assumption that individuals can borrow and lend freely at the bond rate. In practice, some individuals are credit constrained, and all face borrowing interest rates significantly higher than the bond rate, particularly where the purpose of borrowing is to fund current consumption. The existence of credit constraints and borrowing costs implies that Ricardian equivalence does not hold.

Since governments can borrow and lend freely at the bond rate, a dynamic analog to this model, where individuals face idiosyncratic and uninsurable human capital risks

---

[7] It is straightforward to show that either class of enterprise will face agency problems that do not arise in the case of an owner-managed firm (Gans and Quiggin 1997, King and Pitchford 1998) and to observe that these agency problems must be offset by scale economies if large enterprises are to survive in competition with small owner-managed firms. However, since, by definition, owner-managed firms do not issue equity, this point is not relevant to analysis of the equity premium.

[8] Alternatively, as in Quiggin and Chambers (1998), unobservable effort could determine state-contingent payoffs.

that are correlated with the systemic risk to equity that unfolds through time, would yield results at least as favourable to public investment as those derived above. As well as spreading consumption across states of nature through the tax system, governments could use borrowing and lending transactions to spread consumption over time.

1.8 Policy implications

Most privatisations in OECD countries have been undertaken primarily because of the resulting cosmetic improvements to budget aggregates. Governments have used the proceeds of asset sales to 'finance' increases in public expenditure or reductions in taxes. It is now generally recognised that this is inappropriate and for this reason 'underlying' measures of the budget balance, excluding the impact of asset sales, have become popular. These measures are an improvement on the previous cash balance, but are misleading because they treat government business enterprises solely as a source of dividends, with retained earnings being ignored. A number of recent privatisation proposals have been advocated on the basis that the interest savings from using sale proceeds to repay debt exceed the dividends forgone as a result of privatisation. In effect, this analysis values retained earnings at zero. For private enterprises, the Modigliani-Miller theorem shows that dividends and retained earnings are equally valid. Investor preferences for dividends are normally explained in terms of differential tax treatment or the idea that dividends are a signal that profit reports are accurate. Until recently, Australian advocates of privatisation, such as the Department of Finance (1996), claimed that the Modigliani-Miller theorem did not apply to public enterprises and that the retained earnings of such enterprises were 'locked up forever and never used'.

The fallacious nature of this argument has now been recognised, as least in the Federal bureaucracy. The Office of Asset Sales, quoted in the majority report of the committee of inquiry into the proposal for the sale of Telstra (p. 13), correctly states the position in the absence of differences in operating efficiency:

> If perfect capital markets with full information exist the proceeds the government receives will be equal to the stream of dividends plus the retained earnings in Telstra. Therefore the net effect would be neutral.

As has been shown in this paper, the existence of the equity premium is evidence of capital market imperfections which raise the rate of return demanded by private holders of equity. It follows that, in the absence of efficiency differences, the proceeds the government receives will be less than the present value of the expected stream of dividends plus the retained earnings, discounted at the appropriate risk-adjusted rate derived above.

Note, however, that even if the public sector discount rate is lower than the rate for private enterprises with similar risk characteristics, a policy of complete nationalisation will not, in general, be optimal. Differences in operating efficiency must be weighed against differences in the risk-adjusted discount rate. In particular, there are some sectors of the economy, such as agriculture, where the efficiency advantages of private ownership and particularly those of owner-operated firms, are so great that public enterprises in those sectors have consistently failed to cover their variable costs. Obviously, no advantage with respect to the cost of capital can convert a stream of losses to a positive present value. Conversely, even under the perfect capital market hypothesis, externality and monopoly problems imply that public provision will be superior in some areas of the economy. Hence, the issue is one of drawing the boundaries between the public and private sectors and not a choice between pure communism and pure laissez-faire.

The analysis presented in this paper provides a market test for the benefits of privatisation (assuming that any environmental or other externalities have been appropriately internalised). Suppose that the expected profits and risk characteristics of a government business enterprise, assuming continued public ownership, are known. Then, using an estimate of the risk-adjusted cost of capital to government derived from a model of the kind developed here, it is possible to value the enterprise in public ownership. This value may be compared to the sale price realisable through privatisation. Other things equal, privatisation is desirable, if and only if the sale price exceeds the value in public ownership.

In making calculations of this kind, it is necessary to emphasise that the expected value calculation should take account of the possibility of adverse or favourable shocks and should not be a 'surprise-free' projection. The expected value estimated, discounted at the bond rate, would therefore be actuarially fair. The fact that taxpayers are risk-averse is taken into account through the use of a public sector discount rate higher than the real bond rate, but lower than the private sector cost of capital.

1.9 Concluding comments

The Arrow-Lind proposition on the public sector discount rate, the fiscal impacts of privatisation and the equity premium puzzle have all been the subject of lengthy, and often confused, debate. In this paper, it has been shown that the central issue in all of these debates is the same. If the observed equity premium is larger than that which would be generated by a perfect capital market, the optimal public sector discount rate will be lower than the private sector cost of capital and, in most cases, close to the real bond rate, as claimed by Arrow and Lind. Similarly, in the absence

of differences in operating efficiency, privatisation will be welfare-reducing rather than neutral.

More significantly, in the case of privatisation, the analysis presented here shows that differences in operating efficiency associated with privatisation can be balanced against differences in the cost of capital to yield a straightforward test for the desirability of privatisation. Other things such as externalities being equal, privatisation is desirable, if and only if the sale price exceeds the value in public ownership, calculated as the presented value of expected future earnings discount at the optimal public sector discount rate.

References

Arrow, K. and Lind, R. 1970, 'Uncertainty and the evaluation of public investment decisions', *American Economic Review*, vol. 60, no. 2, pp. 364–78.

Bailey, M. and Jensen, M. 1972, 'Risk and the discount rate for public investment', in Jensen, M. (ed.), *Studies in the Theory of Capital Markets*, Praeger, New York.

Constantinides, G.M. 1990, 'Habit formation: A resolution of the equity premium puzzle', *Journal of Political Economy*, vol. 98, no. 3, pp. 519–43.

Constantinides, G.M. and Duffie, D. 1996, 'Asset pricing with heterogeneous consumers', *Journal of Political Economy*, vol. 104, no. 2, pp. 219–40.

Department of Finance 1996, Submission to Senate Environment, Recreation, Communications and the Arts Reference Committee inquiry into the Telstra Dilution of Public Ownership Bill 1996, submission 188, vol. 7.

Domberger, S. 1995, 'What does privatisation achieve? — A comment on Quiggin', *Australian Economic Review*, no. 110, 2nd quarter, pp. 43–7.

Epstein, L.G. and Zin, S.E. 1990, 'First-order risk aversion and the equity premium puzzle', *Journal of Monetary Economics*, vol. 26, no. 3, pp. 387–407.

Forsyth, P. 1995, *The states, microeconomic reform and the revenue problem*, Discussion Paper no. 27, Federalism Research Centre, Research School of Social Sciences, Australian National University, Canberra.

—— 1997, 'Great expectations or hard times: Review of Great Expectations: Microeconomic Reform and Australia by John Quiggin', *Policy*, Autumn, pp. 40–3.

Gans, J.S. and Quiggin, J. 1997, A technological and organisationalexplanation for the size distribution of firms, University of Melbourne, unpublished.

Grossman, S. and Hart, O. 1983, 'An analysis of principal-agent problems', *Econometrica*, vol. 51, no. 1, pp. 7–46.

Grossman, S., Melino, A. and Shiller, R. 1987, 'Estimating the continuous time consumption-based asset-pricing model', *Journal of Business and Economic Statistics*, vol. 5, no. 3, pp. 315–27.

Grossman, S.J. and Shiller, R.J. 1982, 'Consumption correlatedness and risk measurement in economies with non-traded assets and heterogeneous information', *Journal of Financial Economics*, vol. 10, no. 2, pp. 195–210.

Hansen, S. and Singleton, K. 1983, 'Stochastic consumption, risk aversion and the temporal behavior of asset returns', *Journal of Political Economy*, vol. 91, no. 2, pp. 249–65.

Hathaway, N. 1997, 'Privatisation and the cost of capital', *Agenda*, Autumn, pp. 1–10.

Hirshleifer, J. 1989, *Time, uncertainty, and information*, Blackwell, New York.

Kahn, J.A. 1990, 'Moral hazard, imperfect risk-sharing, and the behavior of asset returns', *Journal of Monetary Economics*, vol. 26, no. 1, pp. 27–44.

Kimball, M.S. 1990, 'Precautionary saving in the small and in the large', *Econometrica*, vol. 58, no. 1, pp. 53–73.

King S.P. and Pitchford R. 1998, 'Understanding the incentives in public and private firms', *Australian Economic Review*, vol. 31, no. 4, pp. 313–28.

Kocherlakota, N. 1996, 'The equity premium: It's still a puzzle', *Journal of Economic Literature*, vol. 34, no. 1, pp. 42–71.

Little, I. and Mirrlees, J. 1991, 'Project appraisal and planning twenty years on', in Fischer, S., de Tray, D. and Shekhar, S. (eds), *Proceedings of the World Bank Annual Conference on Development Economics 1990*, World Bank, Washington DC, pp. 351–82.

Machina, M.J. 1989, 'Dynamic consistency and non-expected utility models of choice under uncertainty', *Journal of Economic Literature*, vol. 27, no. 4, pp. 1622–68.

Mankiw, N.G. 1986, 'The equity premium and the concentration of aggregate shocks', *Journal of Financial Economics*, vol. 17, no. 1, pp. 211–9.

Mehra, R. and Prescott, E.C. 1985, 'The equity premium: a puzzle', *Journal of Monetary Economics*, vol. 15, no. 2, pp. 145–61.

—— 1988, 'The equity risk premium: A solution?', *Journal of Monetary Economics*, vol. 22, no. 1, pp. 133–6.

Office of Asset Sales (Department of Finance) 1988, Submission to Senate Environment, Recreation, Communications and the Arts Legislation Committee inquiry into Telstra Transition to Full Private Ownership Bill 1988.

Quiggin, J. 1982, 'A theory of anticipated utility', *Journal of Economic Behavior and Organization*, vol. 3, no. 4, pp. 323–43.

—— 1995, 'Does privatisation pay?', *Australian Economic Review*, no. 110, 2nd quarter, pp. 23–42.

—— 1996, The partial privatisation of Telstra: An assessment, Submission to Senate Environment, Recreation, Communications and the Arts Reference Committee inquiry into the Telstra Dilution of Public Ownership Bill 1996.

—— 1998, The proposed privatisation of Telstra: An assessment, Submission to Senate Environment, Recreation, Communications and the Arts Legislation Committee inquiry into Telstra Transition to Full Private Ownership Bill 1988.

Quiggin, J. and Chambers, R.G. 1998, 'A state-contingent production approach to principal-agent problems with an application to point-source pollution control', *Journal of Public Economics*, vol. 70, pp. 441–72.

Rietz, T.A. 1988, 'The equity risk premium: A solution', *Journal of Monetary Economics*, vol. 22, no. 1, pp. 117–31.

2 Privatisation: Does reality match the rhetoric?

Stephen King and Rohan Pitchford

2.1 Introduction

Privatisation — the movement of assets from public to private ownership — is an important part of economic policy in the 1990s. Governments in a range of countries and of all political persuasions have privatised state-owned assets. The reasons for privatisation have varied. Privatisation, it has been argued, improves productive efficiency, reduces claims on tax revenues and limits the need for government intervention and control of the economy. Asset sales also provide a 'quick fix' to government budget deficits.

This paper presents some of the lessons from formal research on ownership and privatisation. In many cases, these lessons differ significantly from the rhetoric of privatisation. Further, experience with privatisation, particularly in developed countries such as Australia, tends to corroborate these research findings.

Formal research into privatisation provides important lessons for government policy makers. In particular, the choice of either public or private ownership is part of a wider regulatory decision. Ownership and regulation are not distinct. Rather, they are part of the same process of aligning private and public incentives. If privatisation policies are pursued without considering the broader regulatory framework then the result may be less efficiency rather than more.

Privatisation may increase government intervention. This is likely to be the case where governments privatise assets without considering the wider regulatory implications. The owners of newly privatised assets may be subject to a variety of constraints enforced by different regulatory authorities. These regulations, while in the public interest, may impede the 'efficiency gains' supposed to underlie the

benefits from privatisation. In some cases, public ownership may provide a more desirable outcome than privatisation with layers of intrusive regulation.[1]

2.2 Overview

Privatisation in Australia

Privatisation is a key part of Australian public policy. Since the 1989-90 financial year, government revenues from asset sales have exceeded $61 billion.[2] This includes just over $30 billion for the federal government and almost $24 billion for the Victorian government.

This process of selling public-sector assets to the private sector is set to continue over the next decade. The State governments in New South Wales, South Australia and Tasmania are moving to sell some of their public electricity utilities. The New South Wales government is selling its TAB. The ACT government has announced a review of its electricity and water utility, ACTEW, which will consider the potential for privatisation.[3] The Victorian government is in the process of selling its gas distribution and retailing companies and has called for expressions of interest for the sale of the Austin and Repatriation Medical Centre. The Western Australian government recently announced the sale of the Dampier-to-Bunbury natural gas pipeline for $2.4 billion.

While asset sales, particularly large public floats such as the Commonwealth Bank, Qantas and Telstra, dominate the newspaper headlines, there has been a broader move from public to private sector ownership and control. At the local government level, competitive tendering and the contracting out of services, such as office cleaning, park maintenance and garbage collection, has become the norm.[4] A variety of new infrastructure facilities, such as the Melbourne City-Link project, are being built, owned and operated by the private sector. In some cases these facilities will revert to public ownership at a set future date. State governments regularly

1 Formal research underlying the results presented in this paper is found in King and Pitchford (1998a) and (1998b). Other international researchers are also considering the underlying rationale behind privatisation. For example, see Hart, Shleifer and Vishny (1997) and Bolton and Xu (1997).

2 Throughout this paper, the term 'billion' refers to one thousand million. For a useful summary of government asset sales since 1990, see 'A country going, going, going … private' by Michael Bachelard, *The Australian*, 27 April 1998, p. 4.

3 See 'ACT orders review of power plans' by Mark Skulley, *The Australian Financial Review*, 11 May 1998, p. 5.

4 See Domberger and Hall (1995) for case studies on local government contracting.

'outsource' prison services so that over a quarter of Australia's prison population will be held in private institutions by the turn of the next century.[5]

Privatisation overseas

Privatisation is not just an Australian phenomenon. The UK conservative government under Margaret Thatcher led the recent international trend for privatisation with the sale of British Telecom, BP and electricity and gas utilities in the 1980s.[6] Others have followed including New Zealand and countries in South America and continental Europe. In telecommunications, Telecom New Zealand, the Chilean telephone companies CTC and Entel, and France Telecom have all been privatised.

Since the collapse of the Soviet Union, former eastern Bloc nations such as the Czech Republic, Hungry, Poland and Russia have sold a variety of state assets ranging from local shops to oil companies.[7] Different countries pursued different modes of privatisation. The Czech Republic privatised through a voucher scheme. Russia has used a mixture of vouchers and transfers to employees. Germany established a central agency, the Treuhandanstalt, to oversee the sale of thousands of state enterprises owned by the former German Democratic Republic.

Privatisation is not a new phenomenon. The West German government privatised Volkswagen shortly after the Second World War. French governments have used franchise contracts with private operators to manage water and sewerage operations for more than 100 years. The Roman emperors Caligula and Aurelius sold royal furniture and heirlooms to pay debts.[8] The current worldwide trend to privatisation, however, is certainly unprecedented in recent times.

2.3 The case for and against privatisation

A policy in search of a rationale

Despite its importance, there has been little research on the economics of privatisation or, more generally, on the difference between public and private

5 See 'A country going, going, going … private' by Michael Bachelard, *The Australian*, 27 April 1998, p. 4.

6 For a discussion and critique of the UK privatisation process, see Vickers and Yarrow (1988) and Armstrong, Cowan and Vickers (1994).

7 See Giersch (1997) for papers on the Eastern European privatisation experience.

8 See Smith (1987).

ownership. Much of the debate on privatisation involves rhetoric rather than research. It is simply stated that public ownership is less efficient that private ownership. But there is little analysis of the source of this inefficiency or any potential benefits that public ownership may create which could offset this inefficiency.

Much of the economic research has focused on the potential monetary gains from privatisation or the alternative modes of privatisation that may be chosen by the government.[9] The critical academic debate in Australia on privatisation has often centred on whether the government is 'getting value for money'.[10]

This lack of research led Kay and Thompson (1986) to declare privatisation a 'policy in search of a rationale'.

The arguments for privatisation

Many of the arguments presented for privatisation appear superficially attractive. However, they often lack intellectual rigour.

Efficiency: It is often claimed that the public sector is intrinsically less efficient than the private sector. While there is some empirical evidence to support this claim, it is unclear what these studies really show.[11] If public sector managers are provided with different incentives and objectives than their private sector counterparts, then measuring public sector performance by private sector benchmarks will provide little useful information. In fact, where both private and public sector managers have similar incentives and objectives, performance differences are less obvious.[12] Claims of efficiency gaps simply obscure the real issue — what are the underlying differences that distinguish public sector management and private ownership.

Government interference: Privatisation, it is said, may improve performance by limiting government interference in company operations.[13] While such interference

---

[9] On the potential benefits from contracting out, see Domberger, Hall and Li (1995) and Domberger and Li (1995). King (1995) considers whether or not a government should reform an enterprise prior to privatisation. Perotti (1995) considers the benefits of partial privatisation as a step towards full privatisation.

[10] See Quiggin (1995) and (1996), Forsyth (1994) and Domberger (1995).

[11] See Kay and Thompson (1986) and Vickers and Yarrow (1988) for a review of the relevant empirical literature.

[12] For example, see Forsyth and Hocking (1980).

[13] For a formalisation of this argument, see Boycko, Shleifer and Vishny (1996). Lopez-de-Silanes, Shleifer and Vishny (1997) provide evidence for the political nature of the privatisation process in the United States.

may be viewed as objectionable or problematic (although the people doing the interfering are our elected representatives) it is not limited to the public sector. Private firms are controlled by a variety of local, State and federal laws. Nowhere is this more in evidence than in the 'deregulated' industries such as telecommunications. There are currently (at least) five separate bodies involved in the regulation of telecommunications. These include the Australian Communications Authority, the Australian Competition and Consumer Commission, the Telecommunications Industry Ombudsman, the Australian Communications Industry Forum and the Australian Communications Access Forum.[14] In the United States, where telephone services have always been privately owned, telephone companies answer to both State and federal regulators. The degree of government interference in the United States, if anything, exceeds that in Australia. For example, the forced break-up of the private company AT&T, ordered by the courts in 1984, could not occur under current Australian competition laws.[15]

Lack of competition: Unlike private firms, it is argued, public firms are not subject to the rigours of a competitive market and this lack of competition leads to inefficiency. Unfortunately this argument is simply false. A number of public enterprises in Australia competed with private firms prior to their privatisation. These included Australian Airlines, the Commonwealth Bank and the NSW Government Insurance Office. It is not obvious that these firms were less efficient than were at least some of their privately owned competitors. More importantly, it is usually not public ownership that leads to any lack of competition. The privatisation of Melbourne, Perth, Brisbane and other Australian airports has not suddenly created increased competition. Rather, these are now privately-owned, regulated monopolies instead of publicly owned monopolies.[16]

Protection of public managers: An argument that is less often heard today after the significant restructuring of federal and State public services over the last decade relates to public manager incentives. Because of job security, public managers are not forced to operate as efficiently as private sector managers. Public managers are also immune from 'take-overs' that can result in them being sacked. While this may have been true at one stage, it is far from obvious today. Changes to the public sector and experience with private sector schemes, such as 'golden parachutes', that protect incumbent managers mean that public managers may have less security of

14 See 'We've got the watchdogs, but who's feeding them' by Paul Best, *The Age*, 28 April 1998, p. D9, for a brief description of each of these bodies.

15 For a discussion of the break-up of AT&T, see Brennan (1987) and Noll and Owen (1989).

16 That Australia's privatisation process is creating a group of regulated private firms, rather than increasing competition, is starting to be understood in the Australian debate. See, for example, 'Benign monopolies in private hands?' by Alan Kohler, *The Australian Financial Review*, 12 May 1998, p. 21.

tenure than their private sector counterparts. Even if the original argument remained true, it is an argument for restructuring public sector incentives, not changing ownership.

The case against privatisation

While the arguments for privatisation are less than compelling, so too are the arguments opposing privatisation. Many of the opponents of privatisation rely on claims that public managers will be more prepared to act in the public interest than will be private managers.[17] Despite the convenient juxtaposition of the word 'public', this defence lacks rigour. Why should a public manager behave in this way? If a private manager is intent on acting against the public interest, why can't this be prevented?

Other arguments against privatisation focus on the difference between claimed and actual cost savings. Quiggin (1994) for example notes that the cost savings claimed from 'contracting out' often include a transfer from workers. While the purchaser of the service gains, the workers lose. From an economic perspective, this is not a social gain. Quiggin (1995 and 1996) also argues that the sales revenue from privatisation may be less than the present value of the flow of future profits that the government could earn under public ownership. If the higher return accorded to private equity compared with government debt reflects an economic cost, then moving from debt to equity financing with privatisation will lead to an economic cost.

These arguments miss the point. Rather than defending public ownership, the arguments either support alternative modes of funding, reflect potential inequities associated with privatisation or are unsupported claims.[18]

2.4 Understanding privatisation

The key issues that need to be answered to understand privatisation are as follows:

1. What are the different incentives that face public sector managers and private sector owners?

[17] See, for example, the reaction by the AMA federal president, Keith Woollard, to the privatisation of the Austin and Repatriation Medical Centre, in 'Kennett opens hospital sell-off' by Michael Bachelard, *The Australian*, 16 May 1998.

[18] For example, Quiggin's argument on debt versus equity is an argument for government funding of private enterprise, not government ownership. The government could issue 'cheap' debt and then either buy private sector debt or buy non-voting equity in private firms.

2. Why can't the government establish incentive schemes that eliminate the differences between ownership regimes?
3. Why does the government want to use ownership as a tool of economic policy?
4. How should the government choose between public and private sector ownership?

Incentives and ownership

Public and private ownership will only matter if the incentives that face public sector managers and private sector owners differ. If public managers and private owners face the same incentives, then they would be expected to undertake the same actions. So before the differences between public and private ownership can be understood, the different incentives that face managers and owners under the different regimes need to be explored.

Incentives alter with ownership in two basic ways. First, a private owner, unlike a public manager, has direct incentives to increase asset value. A key difference between ownership regimes is the beneficiary of increases in the value of the assets that underlie the business. A public sector manager, at the end of their tenure, has no claim on these assets. The assets belong to the government. In contrast, a private owner retains the assets and has the right to sell them and receive the value of the assets through this sale.

Second, a private owner, unlike a public manager, directly bears any cost increases associated with production. Both a public manager and a private owner can waste resources. Both can 'cost pad' and divert resources for their own personal benefit. But a public sector manager is not personally liable to pay the costs of operating a public firm. In contrast, the private owner pays the bills.

In brief, compared to private owners, public managers have less incentive to undertake personal activities and investments that will increase asset value. Public managers also have less incentive to undertake personal activities and investments that will reduce current costs and, in fact, have an incentive to artificially raise current costs to the degree that the manager can turn these costs into private benefits.

Ownership and regulation

The government need not sit quietly and let either private owners or public managers act as they see fit. Public managers can be provided with incentives to undertake or avoid various activities through their employment contract. For

example, if the government can perfectly measure the value of the public assets under a manager's control then they can reward the manager for increasing the value of these assets. Similarly, if the government can perfectly measure the degree of cost-padding by a public manager then they can make the manager bear any unjustified cost increases through their contract.

Corporatisation — where the government requires a public firm to operate like a private firm — is one example of a public sector incentive scheme. The manager of a corporatised public firm is likely to be replaced if certain cost, output and dividend targets are not achieved.

Private firms can also be provided with incentives to alter their owner's behaviour. The government sets regulatory rules for private firms. These may specify certain activities that the firm must undertake, such as community service obligations. Alternatively regulations may limit private firm activities. For example, price-cap regulation limits the prices that a private firm can charge for certain products.

If the government could perfectly control public and private firms then it could remove any incentive differences associated with ownership and ownership would be irrelevant. Ownership only matters because incentive contracts and regulations are imperfect.

To see this, suppose the government has perfect measures of both asset value and actual costs for both public and private firms.[19] Denote these measures by a and c respectively. Assume that these measures are verifiable by a court so that they can be used in either incentive contracts or in regulatory rules. As these are perfect measures, the actual asset value and costs are denoted by a and c.

The government can set an incentive contract for a public manager based on a and c, $T(a,c)$. It can also set regulatory rules for a private firm denoted by $R(a,c)$. In the absence of an incentive contract a public manager will simply receive a flat wage w. In the absence of regulatory rules, a private owner receives the true asset value but pays the true costs, $a - c$.

Now the government can get the public manager to face exactly the same payoffs as a private owner. It simply sets $T(a,c) = a - c - w$. The public manager's payoff is then given by $a - c$, the same as the unregulated private owner. The government just rewards the public manager according to increases in asset value and 'punishes' the manager for increases in costs.

[19] To make the analysis slightly easier, cost padding is ignored in the example presented in this paper. A more complete treatment of this example is found in King and Pitchford (1998b).

The government can also make a private owner behave like a public manager by taking away the incentives to raise asset value and limit costs. The government can set a regulatory contract $R(a,c) = c - a + w$. In other words, the government can indemnify the private owner from bearing any increased costs but 'fine' the private owner whenever asset value is increased. The private manager's payoff is now identical to a traditional public manager.

Do incentive and regulatory contracts like these exist in practice? Standard rate-of-return regulation is very similar to the regulatory contract given above. Rate-of-return regulation establishes a profit level for the private firm. The firm is unable to exceed this profit target but in many cases is protected from profits falling below this target. Rate-of-return regulation is popular in the United States and is being used in Australia. Most notably, many newly privatised utilities, such as the electricity distribution systems in Victoria, are being controlled by rate-of-return style regulations.

Corporatised public firms involve incentive contracts like the one given above. In fact, the incentive contract shown above is a case of perfect corporatisation — the public manager will behave exactly like a private owner.

For ownership to matter, incentive and regulatory contracts must be imperfect. This is the case in the real world. The government cannot perfectly measure asset value or firm costs. If it tries to have a manager or owner report these variables then the manager or owner will have an incentive to distort the reports that they send to the government.

An immediate corollary is that ownership is a regulatory tool. With perfect regulation, ownership is irrelevant. Ownership matters because regulation is imperfect so that changing ownership alters incentives and changes behaviour.

When does ownership matter?

Ownership affects incentives. But why does the government care? Why does the government ever want to regulate a private firm?

The standard answer provided by economics is that private ownership may not provide socially optimal outcomes. If private owners have market power then they may set monopoly prices that maximise profits but reduce general social welfare. A private firm with market power may also choose a level of service quality that is not socially desirable.

The production process may lead to social costs, and activities that generate private profits may involve a social loss. Pollution is one obvious example. There may also

be spillovers between private firms. The activities of one firm may raise or lower the profits of another firm.

Economics refers to these spillovers between privately profitable activities and the welfare of other members of society as externalities. In the absence of either market power or spillovers there is no need for regulatory intervention. A private owner will seek to produce output in a way that minimises costs and will choose to invest in projects that maximise private gain. When private and social gain are aligned, unregulated private ownership is optimal.

Ownership is simply a tool of regulation. In the absence of external spillovers between private actions and social welfare, there is no reason for public ownership.

Options for the government

Because ownership is a regulatory tool, the optimal form of ownership depends on the type of spillover between private activities and social welfare and the other regulatory tools at the disposal of the government. As noted above, if regulation is perfect then ownership is irrelevant. If regulation is imperfect, incentives will depend on whether the firm is in public or private hands. When the behaviour of a public sector manager and a private sector owner differ then the government must consider which type of behaviour will be in society's interest.

The regulatory and ownership issues facing a government are likely to be complex. Should certain assets stay together or should they be separated? If assets are separated, which assets should be placed in private hands and which are best in public ownership? What forms of regulation should be placed on privately owned firms? What type of incentive contracts should be set for public managers?

To see the complexity of the problems, consider the case of telecommunications. Should Telstra be retained as a single company or should it be divided into a number of separate companies? If Telstra were vertically separated into a 'wires' company and a retail telephone company then this separation would alter the behaviour of Telstra under either public or private ownership. There may be loss of economies of scope between the upstream wires operation and the retail operations. There may also be a change in interactions with competitors. A number of other telecommunications companies in Australia have complained that Telstra, as an integrated company, can distort competition by charging different external and internal prices. While this can harm competitors it may provide positive benefits to

consumers. Separating Telstra may reduce anticompetitive behaviour but may also reduce social welfare.[20]

Gas reform in Victoria provides another example of the complex issues facing policy makers. The government needs to consider structure, regulation and ownership. The Victorian Government has already vertically separated gas transmission from distribution and retailing. It has also horizontally separated distribution into regional monopolies. Gas distribution and transmission prices are controlled by rate-of-return style regulations. Recent draft decisions by the Victorian Office of the Regulator General and the Australian Competition and Consumer Commission suggest that the allowed rate of return to these businesses will be approximately seven per cent. The State government is in the process of privatising these gas businesses. It is also likely to create a 'spot-market' to facilitate gas trading similar to the spot market used in the electricity industry.

It is far from obvious that these reforms are the most suitable for the Victorian gas industry. At present, almost all the natural gas sold in Victoria comes from a single source — the ESSO/BHP joint venture that controls the Bass Strait gas fields. With only a single upstream producer, downstream reforms may offer only cosmetic improvement. Alternatively, downstream reforms may actually lead to higher costs, less efficient transmission and distribution and higher prices for consumers. If vertical and horizontal economies of scope are lost by restructuring the gas industry then reforms can raise production costs. If separation leads each part of the vertical chain of production to add its own margin then final gas prices may rise unless there is rigorous regulation.[21] With only a single gas supplier, it is far from clear what will be achieved by creating a spot market in the gas industry.

Are these complex reforms necessary? In Victoria, where most gas is for residential rather than industrial consumption, competition may not be a problem. There are ready substitutes for many residential uses of gas. Electricity can be used for cooking, hot water and heating, as is the case in other Australian States. Solar power, coal, oil and wood are also substitute household fuels. If the electricity market is competitive, does society need to worry about the gas industry?

The Victorian gas industry may be a case where there is no conflict between private and social objectives. A private gas utility may be unable to exercise market power without losing market share to electricity. It is far from obvious that there are any spillovers from private gas consumption and production to wider public welfare. If

[20] King (1994) considers the ability for an integrated telecommunications firm like Telstra to discriminate between internal and external pricing.

[21] That vertical disaggregation may lead to higher prices is a well known result in economics called 'double marginalisation'.

private incentives and public welfare are aligned then unregulated private ownership is optimal. The Victorian Government may simply be wasting tax-payer's money by restructuring and regulating the gas industry. Rather, it could have simply privatised the former Gas and Fuel Corporation and allowed interfuel competition to provide the correct social incentives.[22]

Another issue that the government needs to tackle is the question of who should manage what? For example, say that there is a government owned national park next to a private resort. The private resort benefits from the national park's proximity and is interested in the maintenance of the park. However, handing ownership of the park to the resort may lead to undesirable spillovers. For example, it may be difficult to prevent the resort owners from exploiting market power in pricing the entrance fee to the park or excluding non-resort users from the park. Private ownership of the park may also make it difficult for the government to prevent the resort owners from extending their resort further into the park.

A useful compromise might be to retain the park in public ownership but to have the resort owners manage the park. This could provide the correct balance of incentives.

The optimal regulation and ownership package is likely to be complex and depend on a wide variety of industry specific factors. Analysing optimal ownership cannot be separated from regulatory decisions and will require careful case-by-case decision making.

Summary

Five simple lessons follow from the discussion in this section.

1. Ownership matters because it effects incentives.
2. If regulation is perfect there is no difference between public and private ownership.[23]
3. From (2), ownership is simply a regulatory tool.
4. Ownership and regulation are only issues when private and social incentives do not coincide.
5. Ownership, regulation and structure must be considered as an integrated package. They are not separate issues.

---

[22] For a slightly longer discussion of gas reform and a general discussion of regulatory issues in vertical production chains, see King and Maddock (1996).

[23] Note that this conclusion does not follow to the integration or separation of assets. Under perfect regulation, the government may still prefer one or other structure. See King and Pitchford (1998a).

2.5 Privatisation in reality

To see how ownership decisions and regulation interact, consider some simple examples.

Will a corporatised public firm really behave like an unregulated private firm?

Many public enterprises in Australia have been corporatised. Public managers answer to a board of directors and are required to pursue maximum profits. But how do these 'corporate' incentives affect a public manager's behaviour?

If the signals of asset value and costs used by the government to provide corporate-style incentives for the public manager are imperfect, a corporatised public firm is unlikely to behave exactly like a private firm. In particular, the manager of a corporatised public firm is rewarded for *reported* increases in asset value and decreases in cost. The manager will have strong incentives to manipulate these reports rather than change actual costs and asset value. As the manager of a corporatised public firm is being rewarded for improving appearances, the manager will precisely take actions to improve appearances.

A corporatised public firm may be more aggressive to both competitors and customers than an unregulated private sector firm. Suppose that the government has a good measure of current costs and profits but only a poor idea of future value. The public sector manager will be rewarded for increasing short-term profits but will face little sanction when undermining long-term value. The manager may act aggressively against current competitors — more aggressively than would a private firm — in order to increase current profit. The manager may have little interest in firm reputation and long-term customer relations. The manager may also have an incentive to neglect long-term maintenance. This will reduce current costs, albeit at the expense of long-term value. But the manager is rewarded for short-term cost reductions and, unlike a private owner, does not bear the burden of lower long-term value.

A corporatised public firm can appear to be highly successful. In fact, by the measures used by the government, it will often appear more successful than will an equivalent private firm. But this reflects inadequate performance measures, not actual performance. If the corporatised firm engages in behaviour that improves measured (but not actual) performance but that reduces social welfare, it may be better to privatise the firm completely.

When is a 'poor' public performance better than a 'good' private performance?

Suppose the government wishes to dispose of some highly toxic waste. It can have a public or a private firm dispose of the waste. Further, assume that the government has few tools available to check on the adequate disposal of the waste. The government is also unable to check indirectly on disposal by observing either firm costs or asset value. Is public or private disposal preferred?

Because the government cannot observe costs, the only regulatory contract available to a private firm is a flat payment for disposal. Similarly the only incentive contract available to a public manager is a flat wage contract. In this situation, the public manager has little if any incentives to minimise costs. However, a private owner has strong incentives to reduce costs.

While the private firm may be more 'cost efficient' than the public firm, this does not necessarily imply that private firm should dispose of the waste. In particular, cost reductions by the private firm may have strong negative social spillovers. This was shown in New Jersey in 1987 and Los Angeles in 1996. Private firms were employed to dispose of medical waste. This usually involves high temperature incineration. But to minimise costs, the private firms found that it was cheaper to simply dump the waste at sea. When the waste began to wash up on local beaches, questions were raised about the desirability of cost minimisation through dumping waste near the shore.[24]

Similar issues arise in the disposal of toxic gases. It is difficult to check if the gases have been disposed of properly. Public disposal may be more costly but this is desirable. Society does not want the manager to cut corners and higher costs may reflect that the job is being done properly.

The costs of separating assets

When the airports in the United Kingdom were privatised, they were sold as an integrated operation. In Australia, the Federal Government decided to sell the airports separately.

There are important operational spillovers between airports. If a plane is delayed leaving Melbourne airport then this can lead to significant rescheduling at the destination airport — say Brisbane. Further, it is hard to regulate to avoid delays that benefit the originating airport but are otherwise unjustified. No regulator would

24 See 'Medical-waste mess leads to fees for health professionals' by Shelby Grad, *The Los Angeles Times*, 29 December 1996, Metro, p. 1.

force a plane to take off if the airport claimed that it was currently unsafe to do so due to weather or congestion.

If a privatised Melbourne airport acts to lower its own costs by not properly scheduling air movements or by downgrading maintenance which leads to delayed departures, this will raise the costs to the privatised Brisbane airport. It will also place an extra burden on passengers. To avoid these spillovers it may be desirable to have joint ownership of the airports. In this way, when the owner of Melbourne airport imposes costs on Brisbane airport, the same owner who creates the costs bears the costs. Following the British example may increase efficiency.

The water industry also illustrates the dangers of arbitrarily separating assets. In Melbourne, for example, the body that controls the headwaters and water transmission, Melbourne Water, has been vertically separated from water distribution. The water distributors have also been horizontally separated into local monopolies. Does this separation offer scope for 'game-playing' between different parts of the production chain?

Suppose a consumer notices an increase in water turbidity. To whom do they complain? The water distributor is likely to blame the transmission company, claiming that they receive 'dirty' water. The transmission company will blame the distributor, arguing that the turbidity is due to a broken pipe or high leakage levels. It will be difficult for the consumer or a regulator to correctly allocate liability for a reduction in water quality.

Because of the spillover between the vertical stages of production in the water industry, separation is likely to result in lower levels of maintenance and a reduction in water quality over time. Each party only bears part of the cost of a reduction in quality due to lower maintenance and only receives part of the credit when they increase quality by improving maintenance. As a result, separated companies will neglect maintenance compared with an integrated company. This will hold under either private or corporatised public ownership. Whichever ownership policy the government adopts, it is better to keep the assets together.

Privatisation can increase government intervention in industry

If it is difficult to regulate a privatised firm and this firm can engage in undesirable practices then public ownership, possibly with an alternative level of private sector involvement, may be preferred.

Again, the airports provide a useful example. The airports in Melbourne, Brisbane, Perth and other capital cities are local monopolies. It is not helpful when travelling

to Brisbane for a business meeting to be told that the airline believes landing charges at Brisbane are too high, so the plane will be diverted to the cheaper Perth airport.

Because of their monopoly status, the airports are regulated by a complex set of rules including price-cap regulations and access rules. However, it is difficult to set regulations perfectly. Despite the best efforts of the relevant regulators, 'loopholes' will exist. The airports may be able to exploit these loopholes to their own benefit but to the loss of wider society.

For example, the Board of Airline Representatives of Australia (BARA) has recently complained about a charge being applied by Brisbane airport to fuel pumped through the airport facilities. BARA claimed that this charge was simply a way around the government's price cap regulation.[25] If this is correct, this is not surprising. The new private owners of Brisbane airport have every incentive to discover ways around the regulatory regime and to increase profits.

As firms find ways around regulations, these regulations are tightened. Regulatory control tends to increase rather than decrease over time. A good example is provided by the privatised British Telecom (BT). When BT was first privatised, price-cap regulations were placed on a number of its services. These regulations were supposed to 'fall away' as competition developed. Unfortunately, the opposite occurred. The price caps have been both reviewed and extended over time as the expected level of competition has not emerged.[26]

Privatisation can create a new range of government bureaucracy and control that may exceed any burden under public ownership. Are the benefits of private ownership worth the cost? When weighing up the regulatory cost against any benefits, the result is far from obvious.

It may be possible to have private benefits with reduced regulatory cost in some cases. For example, rather than selling the airports, the government could have tendered for the operation and maintenance of the facilities. To improve efficiency in the allocation of landing times, the government could have moved to an auctioned landing slot procedure.

Similarly Telstra could have been split into a 'wires' company and a retail telecommunications company. The government could tender for construction of new cable networks and for the operation and maintenance services of the wires

25 See 'Airlines head for airport charges showdown' by Ian Thomas, *The Australian Financial Review*, 19 May 1998. See also the *Courier Mail*, 5 June 1998.

26 See Armstrong, Cowan and Vickers (1994).

company. Spectrum in the cable could be auctioned much as the mobile phone spectrum is auctioned in Australia. While the wires company would remain in public ownership, the net result would be less government intervention compared with the current path of telecommunications reform.

There is often a variety of ownership and regulatory alternatives open to the government. Choosing 'standard' privatisation may result in heavy regulation and high costs. Privatisation may lead to an effective increase in government intervention because of the need for highly intrusive regulation. When this occurs, rather than improving performance, privatisation may increase bureaucracy and destroy any beneficial change in incentives.

2.6 Conclusion

This paper presents a brief overview of privatisation research. A series of clear lessons for policy makers follow from this research.

- Many standard arguments for and against privatisation are superficial or misleading.
- Public performance cannot simply be compared to private performance without considering performance objectives.
- Ownership alters incentives for public managers and private owners and these incentives lie at the heart of the costs and benefits of privatisation.
- Ownership decisions are interlinked with regulatory decisions and are part of the broader regulatory framework.
- Ownership only matters because of regulatory imperfections and understanding these imperfections is necessary to understand ownership.
- Ownership and standard regulatory tools must be considered together with issues of structure and management. Each of these factors will interact with the others.
- Ownership and regulatory decisions should align private and social incentives. Regulation and public ownership are only issues when private and public objectives differ.
- Corporatised public firms will not behave like private firms and may be more aggressive and act less in the public interest than a private firm.
- Privatisation may increase government intervention rather than decrease government intervention.

Governments that make ownership decisions, either public or private, without considering these factors are likely to make public policy that lowers rather than

raises general welfare. Unfortunately, it is far from clear that governments take account of the relevant trade-offs and complexities in their privatisation decisions. In particular, governments at both the State and federal level in Australia appear to pay little attention to the reality of privatisation, preferring to follow their own rhetoric.

References

Armstrong, M., Cowan, S. and Vickers, J. 1994, *Regulatory Reform: Economic Analysis and the British Experience*, MIT Press, Cambridge, Massachusetts.

Bolton, P. and Xu, C. 1997, Ownership and competition: An application to schools, unpublished.

Boycko, M., Shleifer, A. and Vishny, R. 1996, 'A theory of privatization', *Economic Journal*, vol. 106, no. 435, pp. 309–19.

Brennan, T. 1987, 'Why regulated firms should be kept out of unregulated markets: Understanding the divestiture in United States v. AT&T', *Antitrust Bulletin*, vol. 32, no. 3, pp. 741–93.

Domberger, S. 1995, 'What does privatisation achieve? A comment on Quiggin', *Australian Economic Review*, no. 110, 2nd quarter, pp. 43–7.

Domberger, S. and Hall, C. (eds) 1995, *The Contracting Casebook: Competitive Tendering in Action*, AGPS, Canberra.

Domberger, S., Hall, C. and Li, E. 1995, 'The determinants of price and quality in competitively tendered contracts', *Economic Journal*, vol. 105, no. 433, pp. 1454–70.

Domberger, S. and Li, E. 1995, 'An ordered-probit analysis of price and performance in competitively tendered contracts', *Australian Economic Review*, no. 110, 2nd quarter, pp. 14–22.

Forsyth, P. 1994, The states, microeconomic reform and the revenue problem, Paper presented at the conference of economists, Gold Coast.

Forsyth, P. and Hocking, R. 1980, 'Property rights and efficiency in a regulated environment: The case of Australian Airlines', *Economic Record*, vol. 56, no. 153, pp. 182–5.

Giersch, H. (ed.) 1997, *Privatization at the End of the Century*, Springer, Berlin.

Hart, O., Shleifer, A. and Vishny, R. 1997, 'The proper scope of government: Theory and application to prisons', *Quarterly Journal of Economics*, vol. 112, no. 4, pp. 1127–61.

Kay, J. and Thompson, D. 1986, 'Privatisation: A policy in search of a rationale', *Economic Journal*, vol. 96, pp. 18–32.

King, S. 1994, 'Competition and vertical separation', Centre for Economic Policy Research Discussion Paper no. 317, Australian National University, Canberra.

—— 1995, 'Privatization and reformation', *Economic Record*, vol. 71, no. 212, pp. 18–26.

King, S. and Maddock, R. 1996, *Unlocking the Infrastructure: The Reform of Public Utilities in Australia*, Allen and Unwin, Sydney.

King, S. and Pitchford, R. 1998a, 'Private or public? A taxonomy of optimal ownership and management regimes', Department of Economics Research Paper no. 625, University of Melbourne.

—— 1998b, Ownership, regulation and incentives, unpublished.

Lopez-de-Silanes, F., Shleifer, A. and Vishny, R. 1997, 'Privatization in the United States', *Rand Journal of Economics*, vol. 28, no. 3, pp. 447–71.

Noll, R. and Owen, B. 1989, 'The anticompetitive uses of regulation: United States v. AT&T (1982)', in Kwoka, J. and White, L. (eds), *The Antitrust Revolution*, Scott Foresman, Glenview, Illinois.

Perotti, E. 1995, 'Credible privatization', *American Economic Review*, vol. 85, no. 4, pp. 847–59.

Quiggin, J. 1994, 'The fiscal gains from contracting out: Transfers or efficiency improvements', *Australian Economic Review*, no. 107, 3rd quarter, pp. 97–102.

—— 1995, 'Does privatisation pay?', *Australian Economic Review*, no. 110, 2nd quarter, pp. 23–42.

—— 1996, *Great Expectations: Microeconomic Reform and Australia*, Allen and Unwin, Sydney.

Smith, V. 1987, 'Auctions', in Eatwell, J., Milgate, M. and Newman, P. (eds), *The New Palgrave: Allocation, Information and Markets*, W.W. Norton, New York.

Vickers, J. and Yarrow, G. 1988, *Privatization: An Economic Analysis*, MIT Press, Cambridge, Massachusetts.

3 Privatisation: A practitioner's perspective

David Greig

Privatisation is relevant for government activities that can be defined contractually — those that might loosely be called 'business activities'. If the activity can be represented by a contract or set of contracts covering quantity, quality (in various dimensions) and price, and if there is actual or potential competition, experience and theory suggest that private control is likely to produce better performance than government control.

There are some core government activities that do not lend themselves to contracts because the output cannot be clearly defined — for example, the core activities of foreign ministries, of defence forces and of establishing and maintaining the legal system. In such areas, private involvement is typically confined to outsourcing at the fringe (eg honorary consuls, Ghurkhas and legal aid practitioners). Other core government activities are generally considered to include much of the police and some areas of central policy advice and the welfare system. There is lively debate about the best role of government in the health and education sectors. In a previous year, Oliver Hart considered public versus private provision of prison services at this conference.

In much of the world the rest of the public sector — that with 'business' characteristics — has been privatised or is likely to be. The reasons are well documented elsewhere. There is now enough experience to justify the conclusion that well-conducted privatisations, in a competitive environment, improve performance as measured by technical efficiency, innovation, financial results and growth (World Bank 1992, 1995). Supporting theoretical insights can be gained from principal-agent theory, the theory of the firm and public choice theory.

An alternative to privatisation is corporatisation. However, State Owned Enterprises are likely to be less effective than private firms for a number of well-known reasons, the most important being the lack of a market for corporate control: a falling share price tells directors and managers to lift their game or face dismissal following a takeover.

The criticism that privatisation may be financially unattractive to government often rests on a misunderstanding of the reason why interest rates on government debt are lower than those on private debt — government business activities are not inherently less risky, but governments can both diversify risk and cover it through the tax system.

At the practical level there are two basic steps in preparing for a successful privatisation:

- to unbundle government activities — typically economic regulation, non-economic regulation, policy advice, purchase of outputs and production of outputs; and
- to allow competition or at least the threat of it (contestability).

Unbundling puts at arm's length those activities that would involve a conflict of interest if included in a privatisation. For example, postal services once regulated (prohibited) competition for their own activity. Old railway authorities used to help determine which subsidised services they should provide. Of the unbundled elements only the business of producing outputs should be privatised. Other activities should stay in the government or be stopped. Often economic regulation is not necessary once the merits of competition are understood. The government or an agency for which it provides statutory cover may continue non-economic regulation if necessary (eg safety standards). Policy advice is likely to remain a core government activity although parts of it may be outsourced. The government will continue to 'purchase' or subsidise certain activities (eg hospital care or public transport).

The errors of early UK privatisations have confirmed the importance of establishing competitive markets before privatising the previously government-owned monopoly producer. Privatising a monopoly creates a new constituency opposed to liberalisation and creates a need for regulation with its well-known problems.

The unbundling task is usually difficult because it forces the government to define what its regulatory policies are and what its subsidy policies are. In the past these have often been buried or poorly defined in a large public sector entity financed over many years from taxes or monopoly rents. The government will, for example, need to form a view on externalities — whether they exist, are measurable, and whether they justify subsidies or government ownership (aspects of this are addressed in chapter 4).

Victoria's Transport Reform Unit has been preparing for privatisation of the train and tram services by unbundling the activities previously buried in the Public

Transport Corporation and by creating competitive pressures through a franchise/concession model whereby:

- the government is defining minimum service levels and maximum fares based on social and externality considerations. Punctuality and reliability are being encouraged through penalties and bonuses, and bidders are asked what subsidies they require to run the businesses;
- competition is being created by periodic re-tendering of the businesses;
- safety standards will be independently regulated; and
- the Department of Infrastructure will monitor performance and provide policy advice.

Other privatisations differ mainly in the degree of government 'purchase', ranging from 100 per cent with prisons to near zero in the energy sector, compared with about half in public transport.

Experience with transport privatisation in the United Kingdom, Argentina, New Zealand and now in Victoria suggests that it will lead to reduced subsidies, greater innovation and improved services.

References

Megginson, W.L. Nash, R.C. and van Randenborgh, M. 1994, 'The financial and operating performance of newly privatised firms: An international empirical analysis', *Journal of Finance*, vol. 49, no. 2, pp. 403–52.

Vining, A. and Boardman, A. 1992, 'Ownership vs. competition: Efficiency in public enterprise', *Public Choice*, vol. 73, no. 2, pp. 205–39.

World Bank 1992 *Privatisation The Lessons of Experience*, Country Economics Department, World Bank, Washington DC.

—— 1995, *Bureaucrats in Business*, World Bank, Washington DC.

SECTION 2

INDUSTRIAL ECONOMICS AND THEORY

4 A comparison of institutional arrangements for road provision

Barry Abrams, Peter Cribbett and Don Gunasekera[*]

The way in which road provision is undertaken differs considerably in four generic institutional arrangements compared in this paper — the traditional departmental approach, output-based management (OBM), effective road funds and the public utility model. The differences relate to the:

- assignment of responsibility for undertaking specific tasks;
- matching of responsibility to accountably for outcomes achieved; and
- methods used to monitor performance.

The structure of the paper is as follows. The next section describes a broad framework that can be used to compare institutional arrangements. This is followed by a description and discussion of the four institutional arrangement and their relative merits. The final section identifies areas for further work required to help meet the challenges and impediments associated with improving road provision in Australia.

4.1 A framework for comparison

The framework for comparison used in this paper corresponds broadly to the 'principal-agent' model. This model provides guidance on the ways in which accountability, responsibility and autonomy can be appropriately allocated between a 'principal' and 'agent' (Doern 1993 and IC 1996). Transparency is also considered because it is important to allow interest groups to evaluate the performance of government and the road agency.

[*] This paper is a shortened version of Abrams, Cribbett and Gunasekera (1998). The views expressed in this paper are those of the staff involved and do not necessarily reflect those of the Productivity Commission.

Accountability

Accountability applies to both government and the road agency. Nevertheless, the government is ultimately accountable to the public — devolution of responsibility to other parties for undertaking road-related tasks does not reduce the government's overarching accountability for outcomes achieved.

For a road agency, accountability means that the managers of the road agency are held responsible for decisions, with rewards and sanctions for results as evaluated by a satisfactory performance monitoring system. This arrangement is most effective where the road agency has sufficient autonomy to carry out its tasks. Without corresponding autonomy, rewards and sanctions lose effectiveness because a road agency cannot sensibly be held accountable for decisions imposed on it.

Responsibility

In the case of road provision, responsibility is placed with either government or the road agency. Performance monitoring should appropriately reflect the assignment of responsibility. Allocation of responsibility is linked to the tasks associated with road provision. Responsibility is appropriately placed with the road agency where the agency also has sufficient autonomy to significantly influence the success of its assigned tasks.

In all institutional approaches, the government decides the road-related outcomes it desires to achieve and creates the environment within which each institutional approach operates. This has important implications when considering the merits of each institutional approach. In particular, it emphasises that, while the government may transfer responsibility for decision-making to another party, the government is still ultimately responsible to the public for the road-related outcomes achieved.

Autonomy

Autonomy means that a road agency has full internal control to undertake assigned tasks without undue external interference. Autonomy is important because it strengthens accountability and performance monitoring. This is because, if a road agency has sufficient autonomy to achieve specified objectives, external interference cannot be used as an excuse for poor performance.

In the absence of concurrent reforms, such as those establishing accountability and transparency, autonomy in a public sector institution could result in a loss of control. Therefore, in considering the devolution of responsibility for decision-making, the

government needs to establish clear and consistent objectives for the agency supported by effective performance monitoring.

Transparency

In the case of road provision, transparency relates to the ease with which interest groups can observe how well the government and road agency are operating. Transparency is particularly important because it means that the parliament and the public are able to make assessments about the effectiveness of the road agency's operations.

An appropriate level of transparency starts with publication of the road agency's performance monitoring results as established under accountability. However, in the case of road provision, transparency should also extend to many of the decision-making processes used in the planning and delivery of road projects. This could include full publication of investment plans including benefit-cost analyses. Transparency may also be served by including expert assessment of the road agency's operation. This may involve auditing of the agency's activities by the auditor-general or a regulatory overseer.

4.2 Comparison of institutional arrangements

This section applies the framework for comparison developed in the previous section to four generic institutional arrangements — the traditional departmental approach, OBM, effective road funds and the public utility model.

Assignment of responsibilities varies considerably between the four approaches (table 4.1 and figure 4.1). For example, as described later, under the traditional departmental and OBM approaches, the government retains the greatest control over road provision tasks. In contrast, the effective road fund approach and public utility model involve a devolution in responsibility over decision-making away from the government towards the road fund and public utility.

Traditional departmental approach

The traditional departmental approach for road provision outlined is the normal allocation of responsibility and accountability historically associated with most government entities. First, the electorate delegates responsibility to Members of Parliament. Parliament then delegates responsibility to Cabinet/Ministers, who in turn delegate responsibility to a Chief Executive Officer to manage the road agency.

Table 4.1 **Responsibility for undertaking key road provision tasks**

	Institutional arrangement			
Task	*Traditional department*	*Output-based management*	*Effective road fund*	*Public utility*
Setting of objectives	Government	Government	Government	Government
Operating environment	Government	Government	Government	Government
Aggregate expenditure	Government	Government	Road fund/ Government[a]	Public utility/ Regulator[b]
Expenditure allocation	Government	Government/ Road agency[c]	Road fund	Public utility
Project appraisal	Road agency	Road agency	Road fund/ Road agency[d]	Public utility
Project delivery	Road agency	Road agency	Road agency	Public utility
Charging for road use	Government	Government	Road fund/ Government	Public utility/ Regulator

[a] The government and road fund are jointly responsible for setting road user charges and the revenue collected will influence aggregate expenditure levels over the long run. [b] The prices charged for road use by the utility are subject to regulatory oversight which will influence aggregate expenditure over the long run. [c] The road agency is paid to produce a range of outputs (projects) but has autonomy in how the revenue is allocated subject to the constraint that it must produce the agreed outputs. [d] The road controlling agency is primarily responsible for project appraisal but is subject to audit of its appraisals by the road fund.

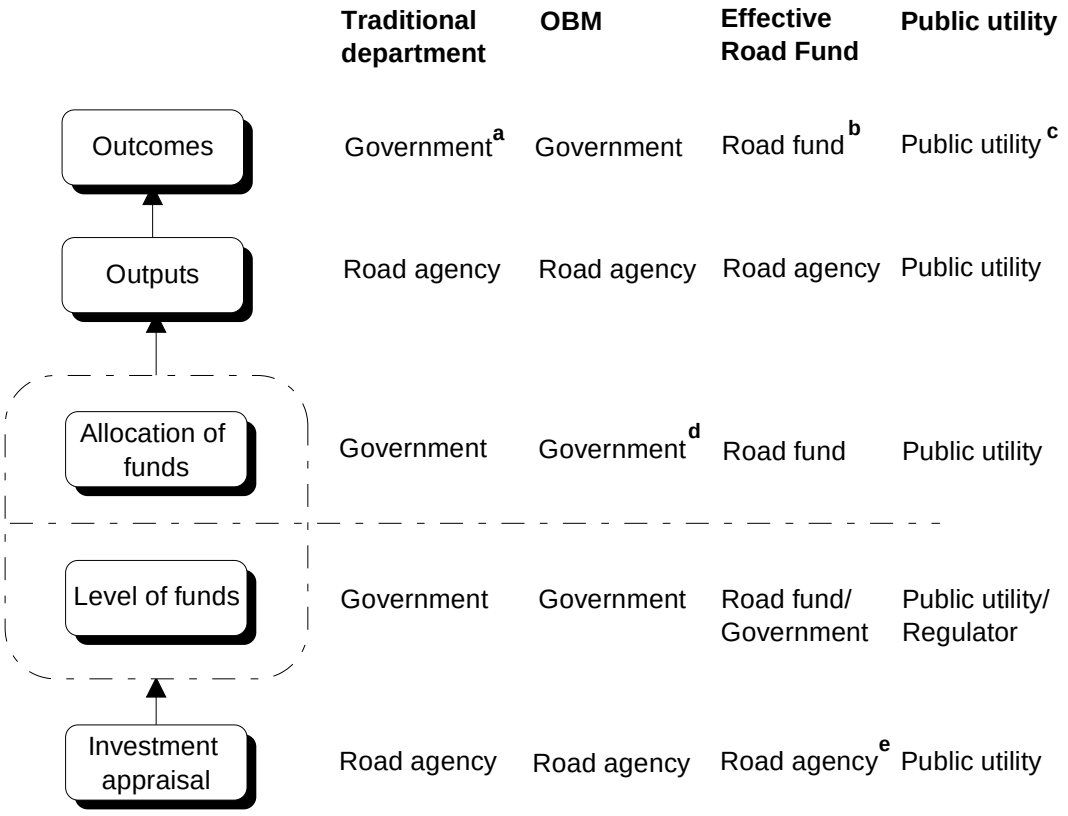
Within the departmental approach, the government ultimately has the authority to determine the aggregate level of expenditure on road infrastructure and where these funds will be spent. Experience suggests that road-related investment decisions under this approach are often influenced and constrained by various political and institutional considerations as well as community expectations.

Within this allocation of responsibility, the government is accountable directly to the electorate for road-related outcomes achieved, with the road agency primarily accountable to the government. The main mechanism through which interested parties can influence road-related outcomes is through the voting process during polling time.

Legislative arrangements

Current legislative arrangements may, in some instances, create confusion over the delineation of responsibility and accountability in many State road agencies. However, regardless of their legislative arrangements, most State road agencies are characterised by the departmental and OBM approaches. For example, VicRoads states that:

Figure 4.1 **Allocation of responsibility for the four generic institutional arrangements**



[a] There is an element of joint accountability between the government and road agency in achieving road-related outcomes. [b] The road fund is responsible for achieving outcomes specified by government. [c] Public utility is responsible for achieving specific targets set by government. [d] While the government is responsible for deciding which projects will be undertaken, the road agency has greater freedom in allocating expenditure to undertake specified projects. [e] The road fund oversees the investment appraisal process of the road agency.

Sources: Based on road agency annual reports, IC (1996), WA Treasury (1996), Parliament of Tasmania (1997), Heggie (1996), Gwiliam and Shalizi (1996), AAA (1997) and Roading Advisory Group (1997).

> The Roads Corporation is a Victorian statutory authority operating under the registered business name of 'VicRoads' (VicRoads 1996, p. 5).

This statement may give the impression that VicRoads is governed in a similar manner to government trading enterprises (GTEs) that also provide economic infrastructure, such as electricity and gas. Yet, the functions and governance of State road agencies are markedly different from those of GTEs. State road agencies are not GTEs because they receive virtually no income from their services. Also, many GTEs have been 'corporatised' — an initiative aimed at replicating many of the commercial incentives that apply to private firms.

Matching accountability to responsibility

A key issue in the departmental approach (and all other institutional arrangements) is the extent to which the government and road agency are held accountable for the decisions they make. An important feature of the departmental approach is that the government and road agency are jointly accountable for road-related outcomes. Nevertheless, despite this joint accountability, the government is still the one primarily responsible for road-related outcomes. This occurs for two main reasons. First, as noted earlier, regardless of the allocation of responsibility for undertaking road provision tasks, it is the government that is directly accountable to the electorate for the road-related outcomes achieved. Second, Ministers retain the ultimate authority to direct the activities of road agencies. Typically, directions do not need to be in writing, made publicly available or tabled before parliament.

Under the traditional approach, the primary focus of the road agency is to determine the benefits and costs of alternative road projects through investment appraisal. The agency also supervises the delivery of approved projects. Improved accountability and performance monitoring regimes may be best served by concentrating on how well road agencies have undertaken these assigned tasks. The areas for performance monitoring may include:

- how the preferences of affected parties have been obtained and incorporated formally into investment appraisal;
- the extent to which the estimated benefits and costs of completed projects have been achieved; and
- the proportion of road construction and maintenance projects that has been completed on schedule and within budget.

Indicators used to assess performance in the areas listed above can be found in most State road agency annual reports and associated documents. Continual improvement of performance indicators, combined with their appropriate use in applying rewards and sanctions to road agency management, can improve the accountability of road agencies to the government and the wider community.

An important mechanism to allow a robust assessment of the performance of road agencies is the public availability of investment analysis. This would allow interest groups to comment directly to the government, road agency or other designated body on the extent to which they believe their preferences have been incorporated adequately into the investment appraisal process.

However, experience suggests that only some elements of investment appraisal documentation are available for public scrutiny. Currently, it is the results of investment appraisal and intended investment programs that are usually available for

public scrutiny and not the actual appraisal documentation itself. For example, VicRoads publishes regional and route plans and major road projects in New South Wales are subject to Environmental Impact Statements or Reviews of Environmental Factors, which include only the results of any benefit–cost analyses.

Output-based management (OBM)

OBM involves the same allocation of responsibility and accountability as the traditional departmental approach. The primary purpose of OBM is to clarify the respective roles of the government and road agency in providing road infrastructure. In doing this, accountability can be strengthened through improved performance monitoring regimes.

Under OBM, the government specifies the desired road-related outcomes. By doing this, it clarifies that government, not the road agency, is accountable directly for road-related outcomes.

The government then specifies the outputs (road construction and maintenance projects) it will purchase to achieve the desired outcomes.[1] As government develops a list of specified outputs, providers (both government and non-government) receive clear directions on what is expected of them. In particular, OBM clarifies that the provider is responsible for providing specified outputs and the government for achieving overall road-related outcomes.

Matching accountability to responsibility

As described above, a key purpose of OBM is to clarify the respective roles of the government and road agency. Specifying and publishing desired road-related outcomes — along with associated performance indicators and targets — allows the public to assess whether the government achieves its stated objectives.

However, the potential advantages of OBM can be limited by various implementation issues. They include:

- the role of the road agency in undertaking investment appraisal;
- the ability to develop and weight robust performance indicators; and
- the extent to which an arm's length relationship between the Minister and road agency can be maintained.

[1] In the case of road provision, the government is technically purchasing 'inputs' and not 'outputs' by purchasing road construction and maintenance projects. This is because each road project is an input into producing road services and not an actual service or output itself.

In common with the traditional departmental approach, under OBM, the government relies on the road agency to estimate the benefits and costs of proposed projects. This information is used by government to decide which outputs to purchase to achieve desired outcomes. This, however, blurs the distinction OBM attempts to overcome — of separating the role of the government and road agency as purchaser and provider.

A crucial requirement of OBM is whether sufficient information can be generated to allow the public to make informed decisions about whether the government has achieved stated objectives. This information relies on development of robust performance indicators for each of the specified road-related outcomes. There are a number of problems in achieving this. In particular, because governments specify multiple road-related outcomes, assessing overall performance requires that each outcome be assigned a weighting — to determine their relative importance and to make tradeoffs between (sometimes conflicting) outcomes transparent. In addition, each road-related outcome may be influenced by various outputs, only some of which provided by the road agency.

Effective performance monitoring for road provision is also hampered by the supply and demand characteristics of roads. Annual evaluations of road-related outcomes may have limited value because roads provide services to users over a long lifetime. For example, increased expenditure on roads to achieve specified economic and regional development goals may not produce tangible results for many years.

Successful implementation of OBM requires that the role of the government and road agency as purchaser and provider be maintained over time. However, problems may develop in maintaining this separation:

> Agencies [a term used to describe a variant of the OBM framework] are not an appropriate solution when they cannot be at arms length from the Minister and involve very politically sensitive issues (DFA 1998, p. 6).

This issue is worth noting in light of the effect individual road projects can have on both the government's budget and road-related outcomes. For example, in New South Wales, the Roads and Traffic Authority's approved capital investment program for 1997-98 is around $1.7 billion. This represents over 40 per cent of the total capital expenditure of New South Wales budget sector agencies (NSW Budget Paper no. 4 1997). Some of the Road and Traffic Authority's major projects have an estimated total cost of more than $500 million.

Effective road funds

The effective road fund ('road fund') approach represents a significant departure from the traditional departmental and OBM approaches in the allocation of responsibility. With the road fund approach, the government's main role is to develop and monitor the institutional environment in which the road fund operates. The government also uses its taxation powers to levy road user charges and to distribute these revenues to the road fund.

Under the approach proposed by Heggie (1996), the representative board of the road fund undertakes the key road provision tasks of deciding the level of expenditure on roads and how funds are to be allocated between different projects. However, the road fund is still reliant upon the government to levy and distribute road user charges to the road fund. This means that, in practice, the level of road user charges payable to the road fund would be determined jointly by government and the road fund (as is the case in New Zealand).

As described in Abrams, Cribbett and Gunasekera (1998), the advantage of the road fund approach is that the representative board has a vested interest to provide an optimal supply of road services from an economic, social and environmental perspective.

Another element of the road fund approach is the introduction of a purchaser-provider split. A purchaser-provider split occurs because the road fund does not undertake investment appraisal or project delivery. Instead, the road fund purchases these outputs from road controlling agencies and local governments.

Information flows and issues surrounding the representative board

With the road fund approach, it is the fund — and not the Minister — that determines the balance between the benefits of increased road expenditure and the cost of provision. However, the road fund is still a government entity and the approach does not preclude avenues for the Minister to influence the activities of the road fund. The government has mechanisms such as a terms of reference, statement of intent or through the road fund's enabling legislation to influence the board. In addition, the Minister may make written directions to the road fund. Such directions should be tabled in parliament and be made available for public scrutiny.

Gwilliam and Shalizi (1996) and Heggie (1996) identify three issues arising from the use of a representative board. They are:

- the ability of the representative board to represent the interest of all parties affected by the provision of road infrastructure;

- the ability of the board to make correct decisions regarding the tradeoff between maintenance and new investment; and
- the lack of evidence currently available to demonstrate that road funds result in improved road-related outcomes.

The effectiveness of the representative board is highly dependent on whether board members are selected in a way that increases the likelihood they will act in the interests of the community as a whole. Gwilliam and Shalizi caution that a board representing only road users, for example, may not adequately address valid social and environmental concerns:

> Beyond the narrow confines of road maintenance, Road Boards dominated by 'user representatives' may not allocate resources optimally (Gwilliam and Shalizi 1996, p. 6).

To avoid this, Heggie suggests that the board should have representatives from all parties affected by the provision and use of road infrastructure.

The benefits of the road fund approach (in terms of improved road-related outcomes) in supplying roads in a modern economy are yet to be established. This is because the effective road funds is a relatively new approach. In particular, this variant of the road fund has only been applied in New Zealand and developing countries.

Accountability and scope for redress against the representative board

The success of the road fund approach is dependent on the mechanisms available to hold the representative board accountable for outcomes achieved. Simply shifting the responsibility for making road-related investment decisions to a representative board will not necessarily result in improved road-related outcomes. The representative board must also face appropriate rewards and sanctions to encourage improved performance.

Under the road fund approach, there are two main mechanisms to hold the representative board accountable for outcomes achieved. First, each representative board member should be nominated by, and hence answerable to, their constituency. Second, the Minister should have scope to replace board members. However, a crucial and still largely unresolved issue concerns the circumstances and conditions under which the Minister should be able to remove board members for poor performance.

The public utility model

The public utility model aims to achieve an optimum supply of road services by treating the provision of roads in a manner similar to other forms of economic infrastructure. However, the public utility model for road provision is still only an 'in principle' proposal and is yet to be implemented on a broad scale in any developed country.

A road utility should have the autonomy to charge directly for the use of roads (subject to regulatory oversight) and to supply road services on a commercial basis. (However, adoption of a public utility model for road provision need not preclude the government from pursing clearly specified social, safety and environmental objectives.)

There are important features of road provision that must be taken into account when considering application of a public utility model. These features include:

- some aspects of road provision having natural monopoly characteristics;
- significant externalities associated with the provision and use of roads, such as environmental externalities; and
- avenues for the government to pursue both economic and non-economic objectives.

Regulatory oversight

The Roading Advisory Group (1997) suggests that, because of the pervasive nature of some road provision decisions, a road utility should be subject to strong mandatory consultation and disclosure requirements. Also, a public utility providing road infrastructure may have the potential to exploit 'monopoly' power by overcharging for road use which may also require regulation.

That said, there has been criticism of past attempts to regulate the activities of public utilities. Criticisms have focused on the appropriateness of incentives created by regulators and the effects that these incentives have upon resource allocation.

Externalities and non-economic objectives

A public utility providing road infrastructure and pursuing only commercial objectives may fail to provide an optimum level of road services. This is because the provision and use of roads may result in various externalities (eg noise and air pollution) which should be reflected in the pricing and provision of roads. In addition, the government may wish to pursue social objectives. For example, it may

wish to provide roads to a higher standard in rural areas than justified purely on an economic and commercial basis.

The government has two main mechanisms to pursue these objectives: through legislation, or the direct purchasing of outputs from the utility. In evaluating which mechanism to use, the government should consider the extent to which legislation can be an effective tool in promoting optimal environmental outcomes and the ability of government to clearly specify and cost road outputs to achieve social objectives.

Accountability and scope for redress

The public utility model represents the greatest devolution of responsibility away from the Minister towards the infrastructure provider. However, this does not mean that the Minister is no longer accountable for road-related outcomes achieved. Rather, it strengthens accountability because the Minister must specify a combination of financial and non-financial outcomes that balances the profit motive of the road utility against the expectations of the community in regard to the level and quality of road services provided.

With a public utility model, the Minister could bring redress against the board of the utility the way it would with any GTE. In particular, the Minister could remove board members for poor performance when agreed performance targets have not been achieved.

As noted earlier, the provision and use of roads affects all members of the community, so members of the community need to be able to seek redress directly from the utility. The ways they may do so include:

- by using alternative transport modes, such as rail, water and air transport (also available under other institutional approaches);
- through commercial laws governing aspects of the utility's operations; and
- through the regulator directly.

The effectiveness of these mechanisms will depend on various factors which may require more detailed examination. These include:

- the availably of alternative transport modes;
- the financial and legal resources of affected parties; and
- the authority and mechanisms to bring redress established as part of the regulatory regime.

4.3 Issues for consideration

Two recent Parliamentary Committees of Inquiry into road provision (and work by other authors) highlight concerns about the way in which Australian governments provide road infrastructure. In particular, submissions to the two Parliamentary Committees of Inquiry argue that:

- governments have not set clear road-related objectives;
- current investment patterns do not maximise the quality of service to road users; and
- political imperatives and funding uncertainty hamper long-term planning in relation to road provision.

These concerns do not necessarily imply unacceptable road-related outcomes. However, there does appear to be room for further improvement in the way governments in Australia provide road infrastructure.

As a starting point, an issue to be considered is whether road-related outcomes can be enhanced by improving existing Australian institutional arrangements or whether a new set of arrangements is required.

Improving current Australian institutional arrangements

Each alternative institutional approach described contains elements that would appear to provide scope to improve the accountability, transparency and efficiency of road agencies and government within the current framework. However, implementing these elements may prove challenging.

Most State governments have or are implementing some version of the OBM approach as part of general management reforms. Application of OBM provides one avenue for governments to improve accountability and transparency with minimal change to current operating environments. However, in the case of road provision, there are important implementation issues which may limit the potential gains of OBM. For example, OBM requires the development of robust performance indicators and maintenance of the arm's length relationship between the government and road agency.

The effective road fund represents a significant departure from current Australian institutional arrangements. The closest practical version of the effective road fund may be found in the operations of Transfund (a road fund) in New Zealand. The effective road fund approach uses a representative board of management to promote a more optimal supply of road services. However, the merits of the road fund

approach for a modern economy are yet to be demonstrated and some important accountability issues are yet to be resolved.

The public utility model represents the greatest devolution of responsibility to the provider of road infrastructure. Versions of the public utility model have been proposed by a Roading Advisory Group in New Zealand and in the Australian Automobile Association's submission to the Neville Committee. As discussed earlier, road provision exhibits a number of important features that must be taken into account in considering the application of the public utility model.

Areas for further research

There are still unresolved issues requiring further research surrounding the effective road fund approach and the public utility model. For example, the effective road fund approach has unresolved issues concerning the composition and accountability of the representative board. A road utility may misuse monopoly power and require effective regulation. These are issues which require further research.

Apart from the possibility of moving to a new set of institutional arrangements, there may also be value in exploring further the options for improving current Australian institutional arrangements. For example, there may be elements of the effective road fund approach and public utility model which could be readily applied to current Australian arrangements. These would act to strengthen accountability and transparency. Examples include the requirement that directions from the Minister to the road agency must be made in writing and tabled in parliament. The road agency could be made subject to the same strong mandatory disclosure laws proposed for the public utility model.

Change must also be considered within the context of Australia's three tier system of government. All three tiers provide road infrastructure with the tiers having overlapping responsibilities but varying abilities to raise revenue to finance spending programs. Australia's three tier system of government raises the following issues which must be addressed before changing current institutional arrangements:

- the allocation (also termed classification) of roads between different levels of government;
- the ability of each level of government to collect road user charges;
- the pricing of road use and the allocation of road revenues;
- the coordination of road infrastructure with other forms of transport infrastructure, such as rail and sea ports; and

- the cost of implementing institutional change, especially for the smaller jurisdictions in Australia.

References

AAA (Australian Automobile Association) 1997, *Submission to the Inquiry into Federal Road Funding*, Australian Automobile Association, Canberra, February.

Abrams, B., Cribbett, P. and Gunasekera, D. 1998, *A Comparison of Institutional Arrangements for Road Provision*, Staff Research Paper, Productivity Commission, AusInfo, Canberra, June.

DFA (Department of Finance and Administration) 1998, *Lessons Learned from Others: The Establishment of Agencies*, Discussion Paper 1, March.

Doern, G. 1993, *Political Accountability and Efficiency*, Ottawa Government and Competitiveness School of Policy Studies, Queens University, Discussion Paper no. 93-20, Ottawa.

Gwilliam, K.M. and Zmarak M.S. 1996, *Road funds, User Charges and Taxes*, TWU papers — TWU 26, World Bank, Washington DC, October.

Heggie, I. 1995, *Management and Financing of Roads. An Agenda for Reform*, Technical Paper no. 275, World Bank, Washington DC, January.

—— 1996, *Management and Financing of Roads — Lessons from Experience*, World Bank, Washington DC, December.

IC (Industry Commission) 1996, *Competitive Tendering and Contracting by Public Sector Agencies*, Report no. 48, AGPS, Melbourne, January.

NSW Budget Paper 1997, *State Capital Program*, Budget Paper no. 4, NSW Treasury, Sydney.

Parliament of Tasmania 1997, *Budget Paper no. 2, Operations of Government Department 1997-98*, vol. 2, G.C. Priestley, Government Printer, Hobart.

RAG (Roading Advisory Group) 1997, *Road Reform The Way Forward*, Final Report, Wellington, New Zealand, December.

VicRoads 1996, *VicRoads' Annual Report 1995–96*, VicRoads, Melbourne.

WA Treasury (Treasury Department Western Australia) 1996, *Output Based Management (OBM): An Overview*, WA Treasury, Government of Western Australia, Perth, July.

5 Firm ownership structure and adoption of multiple communication technologies

*Heli Koski**

This paper analyses the adoption of three communication technologies by Finnish firms. It does so by developing an econometric model of the simultaneous diffusion of multiple technologies that allows for correlation between the time-to-adoption variables via unobserved heterogeneity. The results indicate significant interdependence in the adoption of different communication technologies. The ownership structure of the firm has a substantial impact on the timing of adoption. Further, the results indicate the presence of non-negligible agency problems in adopting new technologies where ownership and management are separate.

5.1 Introduction

The new economic theory of the firm has extended the traditional economic approach of the behaviour of the firm towards a more profound understanding of the organisational structure and its relationship to firm behaviour (see Holmström and Tirole 1989 and Milgrom and Roberts 1992 for an extensive discussion of the new theory of the firm). The literature stresses that agency problems may arise within an organisation where there is separation of ownership and control, conflicting interests of various self-interested parties within the firm, imperfect and asymmetric information, and costly contracts. These agency problems have non-negligible implications for the investment behaviour of firms (Jensen and Meckling 1976, Fama 1980, Shleifer and Vishny 1986). This theoretical framework motivates the following question explored in this paper: What are the implications of different

* The work reported in this paper draws substantially on research undertaken during a visit to Stanford University. The author is grateful to Frank Wolak whose helpful suggestions and comments have greatly benefited this version of the paper.

ownership structures — in particular, ownership concentration[1] and shareholder diversification — for the adoption of new technologies by firms?

This paper also contributes to the previous empirical studies that econometrically explore the adoption of *multiple* technologies (Colombo and Mosconi 1995, Stoneman and Kwon 1994). In doing so, it develops a sound econometric model of the simultaneous diffusion of multiple technologies by jointly modelling the interrelated time-to-adoption decisions faced by firms for three technologies. It is assumed that the time-to-adoption variables are Weibull distributed. Therefore, the corresponding logarithms have the extreme value distribution. The econometric model takes the interdependence of adoption decisions of the different technologies into account by incorporating the unobservable heterogeneity into the model. Also, the econometric model provides sound estimates of the degree of technological complementarity between the technologies.

This paper considers the adoption of three types of advanced communication technologies (ACT) — intra-firm electronic mail (e-mail) technology, inter-firm e-mail technology and electronic data interchange (EDI) — in the Finnish metal industry between 1981–95. The intra-firm and inter-firm e-mail technologies provide a means for transferring information from brief notes to extensive files between the computers of users, respectively, within the firm and between the firms. EDI involves exchanging trade data and business documents — such as purchase orders, invoices and quotes — in a standardised format between the computers of the users.[2]

The data set allows the examination of whether the adoption of new communication technologies differs between:

1. manager-owned and professionally managed firms;
2. concentrated and diffusely held firms; and
3. firms whose owners have well-diversified and poorly diversified portfolios.

The empirical findings indicate that the ownership structure of a firm plays a critical role in the timing of adoption of e-mail technologies. Firms run by a manager-owner

[1] The previous economic literature on the investment behaviour of firms suggests that ownership concentration may have notable implications for entrepreneurial investments (see Ihamuotila 1994 for an overview of the literature). The study of McEachern and Romeo (1978), for instance, implies that firms that are controlled by the dominant outside shareholder are more innovative — measured by their R&D spending — than others.

[2] EDI requires trading partners to agree upon the information they exchange and upon the standard format for exchange. However, firms are free to use different document processing software, as EDI automatically converts the standard data format into the company or industry specific data format and vice versa.

tend to adopt e-mail technologies later than firms where professional managers[3] make the adoption decisions. There is also some evidence that owners who have a well-diversified portfolio are more risk averse in terms of their technology investments and adopt e-mail technologies later than poorly diversified owners do. Moreover, the data set indicates the presence of significant complementarity among the technologies. Finally, the data suggest positive time dependence for all of the technologies concerned — the probability of adopting all of the technologies increases with time.

Section 5.2 discusses potentially relevant factors affecting the adoption of new communication technologies. Section 5.3 presents the econometric model and introduces the data set used in the empirical exploration. Section 5.4 presents the empirical results. Some concluding remarks are presented in section 5.5.

5.2 Adoption of multiple communication technologies

This paper covers three branches of the Finnish metal industry: the basic metal industry, the mechanical engineering industry and the electronics and electrical industry.[4] The firms of the Finnish metal industry typically use high technology, are internationalised[5] and their R&D expenditures are higher than average Finnish firms. Modern manufacturing also involves flexible production techniques that firms can quickly adapt to changes in demand and the business environment (Milgrom and Roberts 1990 and Wellenius, Miller, and Dahlman 1993).[6] These characteristics stress the importance of information and its efficient, fast processing and delivery to firms. New communication technologies enable firms to achieve the required efficiency, flexibility and speed.

One of the essential factors affecting the adoption of new communication technologies may be the ownership structure of the firm. Previous economic literature provides scarce empirical evidence of the importance of ownership structures.[7] None of the previous studies appears to have explored the effects of

[3] Professional managers are the managers of firms whose ownership and management are separate.

[4] Production in the metal industry comprises a variety of different products, for example, telecommunications equipment, consumer electronics, electrical motors and metal products.

[5] The Finnish metal industry is the most important export sector in Finland, covering almost half of all export revenues in Finland.

[6] This is also noted in the world wide web (WWW) pages of the Federation of the Finnish Metal Industry (see http:\\www.met.fi).

[7] Rose and Joskow (1990) suggest that the investor-owned utilities in the electrical industry tend to adopt new technologies earlier than publicly-owned or cooperatively-owned utilities. Loh and Venkatraman (1993) find that the order of magnitude of investments in information technology

concentration of ownership and shareholder diversification on the adoption of new technologies by firms. The data set used here distinguishes between ownership structures of firms. The decision-making environment differs between ownership structures through potential information asymmetries, agency problems between owners and managers, and through the portfolio diversification of the owners. The current model distinguishes between ownership structure through the following dummy variables:

- **own1** = 1, if the firm is managed by its owner (ie family-owned and with concentrated ownership[8]) and 0 otherwise;
- **own2** = 1, if the ownership of the firm is concentrated (ie family-owned or the two biggest shareholders own over 30 per cent of the stock) and 0 otherwise; and
- **own3** = 1, if the owner has a poorly diversified portfolio (ie a firm is a main investment object of its owner or two biggest shareholders) and 0 otherwise.

The above variables were formed on the basis of the questionnaire incorporating three questions about the ownership structure of the firm. First, respondents were asked whether the firm was family-owned, owned by another industrial company or owned by a financial company or investment fund. The respondents were then asked to answer yes/no to questions regarding whether (1) the firm was family-owned or the biggest shareholders owned over 30 per cent of the stock in the firm, and (2) the firm is a main investment object of its owner or two biggest shareholders. This paper briefly discusses the theoretical background behind the use of these variables in explaining the adoption of new communication technologies.

The variable 'own1' distinguishes manager-owned firms from those managed by professional managers.[9] The adoption of new communication technologies may differ between manager-owners and professional managers. Professional managers do not bear the entire costs or risk of an investment, nor do they receive the entire monetary benefit arising from an investment. This means that the expected costs and

by a firm is negatively related both to the concentration of ownership and to the stock ownership of the inside shareholders. However, the latter study does not distinguish the types of monitoring or controlling shareholders by their portfolio diversification.

8 All family-owned firms with the exception of three had a concentrated ownership structure. Thus, the variable 'own1' generally represents family-owned firms in the sample.

9 It is believed that the family-owned firms with concentrated ownership structures represent manager-owned firms. The basis of this argument is that the family-owned firms in the Finnish metal industry are typically smaller than other firms in the industry — in the sample, the family-owned firms are, on average, about 28 per cent of the size of the other firms in terms of their turnover and about 40 per cent of the size of the other firms in terms of their number of employees — and that top managers of family-owned firms are typically the owners.

benefits of adopting new communication technology may differ between the manager-owner and the professional manager.

The variable ‘own2’ distinguishes firms with dispersed ownership from ones that have one or two owners who own a notable share of the firm. In the diffusely held companies, managerial incentive problems arise when the following two conditions hold simultaneously: (1) the shareholders do not have the same information as the manager has, and (2) the interests or preferences of the shareholders conflict with those of the managers (Greenwald and Stiglitz 1990). The small individual shareholders often find it too costly to monitor managers (Shleifer and Vishny 1986). The manager may then reach a different conclusion regarding the benefits of adopting a certain technology than the shareholders would if they were making the same decision. This may happen, for instance, if a manager expects future career opportunities to increase with the ability to use modern communication technologies.

Another important characteristic of the ownership structure of a firm — captured by the variable ‘own3’ — which may critically affect the adoption of new technologies is shareholder diversification (Ihamuotila 1994). This variable roughly distinguishes between owners with a poorly or non-diversified portfolio (ie the firm is a main investment object of the owner) from the ones with a well-diversified portfolio (ie owners who have diversified their portfolio into several different investment objects). It seems intuitively ambiguous whether firms with owners who have a well-diversified portfolio are likely to adopt new technologies earlier or later than those who have concentrated their wealth into one particular firm. On one hand, an investor can reduce the risk related to a single investment object by diversifying their portfolio to include several investment objects. Consequently, well-diversified investors may represent a more risk averse owner type who is less likely to commit to uncertain investments in new technologies than the owners who have concentrated their wealth into a single investment object. On the other hand, owners who have tied their personal wealth to one firm may value controlling managers more than the well-diversified owners and may also be more risk averse in their investment decisions, since their personal wealth is closely tied to that single investment object.

Based on the above reasoning, it is unclear whether the considered three ownership characteristics should be positively or negatively related to the time taken to adopt new communication technologies.

Control variables

This paper also explores the effects of various factors suggested by the previous theoretical studies of the adoption of new communication technologies.

Firm size and market structure

Based on the previous economic literature, it seems reasonable to assume that the duration times with respect to the adoption of different communication technologies are negatively related to the size of a firm. Large firms may adopt new communication technologies earlier than others, not only because of their better financial resources or the availability of skilled labour and technical expertise, but also because of their potentially larger communications need compared with the smaller firms. The game-theoretic literature (Reinganum 1981 and Quirmbach 1986) suggests that the competitive environment or market structure also affects the innovative behaviour of firms. Reinganum (1981) shows that a decrease in the number of firms in the market accelerates the diffusion of new technologies when firms behave noncooperatively. Quirmbach (1986) models the timing of adoption when firms form a joint venture and choose adoption dates such that they maximise joint industry profit. In this case, the diffusion of new technologies is instead hindered by market concentration.[10]

Since the three branches of the Finnish metal industry covered in this paper each incorporate thousands of firms, it does not seem credible that such a high number of firms could coordinate their adoption dates in order to maximise their joint profits, even if they had some interest in doing so. Thus, the duration times of different technologies are hypothesised to be negatively related to the degree of competition, such that a higher concentration in the industry and the higher market power of a firm facilitate the adoption of communication technologies.[11] Firm size and market concentration are measured, respectively, by the following variables:

- **llab** = the (log) number of employees of the firm in 1994; and

[10] Several previous empirical investigations indicate that the adoption of new technologies is positively related to firm size (Mansfield 1968, Hannan and McDowell 1984, Rose and Joskow 1990, Colombo and Mosconi 1995), but they do not provide any clear evidence of either positive or negative relationships between the timing of adoption of new technology by a firm and the degree of competition in its industry.

[11] The game-theoretic literature also suggests that, as the number of rivals adopting a certain technology increases, the returns to the marginal adopter from new technology decreases and that the order in adopting a new technology affects the returns to a firm from using the technology (see Karshenas and Stoneman (1993) for a discussion of this topic). Due to the limited data set, it is not possible to investigate the effect of the adoption behaviour of rivals on that of the firm.

- **conc** = the market share of the ten largest firms in the industry in 1994 (expressed as a percentage).

The variable 'llab' came directly from the questionnaire indicating the number of employees in the firm in 1994. Data on concentration ratios are based on Statistics Finland (1995). The variables 'llab' and 'conc' are negatively related to the duration variables.

Epidemic effects

Empirical models of the diffusion of new technologies have for a long time suggested that the diffusion of technologies follows an 'epidemic' path. Since information regarding a new technology increases with the number of its users, the probability of the adoption of technologies increases over time (Mansfield 1968 and Stoneman 1983). It is assumed that a variable describing elapsed time after the introduction of different communication technologies captures the epidemic effect.

The propensity score

The propensity score is used to correct for selection bias in the estimated econometric model (section 5.3):

- **ps** = the propensity score, with the propensity towards variable 'own1' receiving the value 1 given the observed explanatory variables. The propensity score is estimated by using a logit model (Rosenbaum and Rubin 1983, 1985).

The next section presents the econometric model used to explore the adoption of multiple technologies and introduces the data set used in this study.

5.3 Econometric framework and data

Econometric model

This section builds an econometric model to explore the relationship between ownership structure and the adoption of multiple communication technologies. The first concern in building the model relates to the fact that it is impossible to collect data such that each firm owned by a manager-owner has an equal counterpart — in terms of other observable explanatory variables — managed by a professional manager. The same concern applies to the other ownership variables. In other words, the estimation results may be biased due to missing data or the estimations

may lead to false conclusions due to selection bias. This can be tested by comparing the distribution of the observable explanatory variables for groups with different ownership characteristics, before evaluating the need to correct this bias in the estimated econometric model.

The calculations of the two-sample *t* tests[12] indicate that the sample means of the group of manager-owners (own1=1) and professional managers (own1=0) differ statistically significantly (t>2) in firm size — the sample means were 4.97 and 3.73, respectively, in the two groups — but not with respect to other explanatory variables. These differences in sample means also apply to the ownership variables 'own2' and 'own3'. These substantial differences in the sample means indicate that there is a need to correct the difference in the means of the observed variables in the sample prior to the estimation of the econometric model. This paper uses the propensity score method of Rosenbaum and Rubin (1983, 1985) — that assumes that the source of bias is selection on the observable variables — to control this bias.

This paper estimates propensity scores for the three ownership variables using a logit model (ie the propensities towards variables 'own1', 'own2' and 'own3' having the value 1 given the observed explanatory variables). The estimated propensity scores were highly correlated with one another[13] and thus only one of the estimated propensity scores — the one relating to the variable 'own1' — was used to correct for selection bias in the estimated econometric model. Also, the use of three propensity scores would have led to the problem of multicollinearity and these additional variables would have further complicated estimation of the fairly complicated econometric model.

There are various reasons why the different communication technologies are likely to be characterised by technological complementarity. That is, the net benefits from jointly using all of the technologies exceeds the sum of the net benefits from using the technology separately. Technological complementarity may take place since all communication systems require certain common components (eg personal computers) in which a firm has to invest when adopting its first communication system, but not necessarily after that. Another prominent source of technological complementary relates to the fact that the use of communication technologies typically involves remarkable learning costs. Once learnt, the knowledge and skills achieved can be readily transferred to other advanced communication systems.

---

12 The two-sample *t* statistics can be rewritten as: $t = (\bar{x}_1 - \bar{x}_0) / \sqrt{(\sigma_1^2 / N_1) - (\sigma_0^2 / N_0)}$, where $\bar{x}_i$ is the sample mean of group *i*, σ_i^2 is its variance and N_i is the size of group *i*.

13 The correlation between the propensity scores of variables 'own1 ' and 'own2' was 0.77 and between variables 'own1' and 'own3' was 0.96, both approaching collinearity.

Therefore, it follows that a firm that has adopted one technology is more likely to adopt its complement than it would had it not adopted that technology.

The aim is to jointly model the adoption decision of different communication technologies and to explore how the several exogenous variables discussed above affect the interdependent timing of adoption of these technologies. The duration or failure times t_i — which are non-negative, random variables denoting the time elapsed from the introduction of communications technology *i* until its adoption — are used to describe the timing of adopting communication technologies. These variables are independent of the existence of the firm at the time of the introduction of the technologies. Therefore, the estimated models describe the probability of a firm adopting each technology in a given year.[14] A substantial number of the firms in the sample did not adopt any of the communication technologies concerned. Consequently, the task requires an econometric model that not only allows correlation among duration times, but also controls the right censored adoption dates of the non-adopters.[15]

The duration times are assumed to be Weibull distributed. The Weibull distribution provides flexibility regarding the relationship between the probabilities of the adoption and time — the Weibull model allows the hazard function of duration[16] to be increasing, decreasing or constant in time. A restrictive feature of the Weibull model is that it assumes a monotone relationship, which is not necessarily the case in reality. This constraint of the Weibull model stresses the importance of checking the appropriateness of the model when it is used for describing the underlying distribution of time-to-adoption variables.

This paper considers the logarithms of time-to-adoption variables in the model and denotes the log time-to-adoption variable of technology *i* by $y_i = \log(t_i)$, where *i*=1,2,3 corresponds to the three different communication technologies. The assumption of the Weibull distribution of time-to-adoption variables implies that the corresponding log variables have the extreme value distribution which is directly related to the Weibull distribution (Lawless 1982). It is assumed that y_is may be correlated, but that they are determined independently from one another conditional

14 The time-to-adoption variables do not differ remarkably from the duration variables that capture the probability of adoption after a given spell of non-adoption, since, according to the list of members of the Finnish metal industry, about 85 per cent of the firms (reporting their establishment year) were established before 1990 and about 70 per cent of the firms were established before 1985.

15 Right censoring means that the exact adoption dates of the non-users of technologies are not known, but these firms have not adopted a certain technology by the time the data were gathered.

16 The hazard function of duration variable *i* in the model describes the rate of adoption of technology *i* at time *t* conditional on the fact that a firm has not adopted technology *i* until *t*.

on unobservable heterogeneity random terms, θ_i. The econometric model allows for these unobservable heterogeneity random terms to be correlated one with another. The probability that a firm adopts technology *i* at time *t* conditional on θ_i and on the vector of exogenous variables, x_i, can be defined by the following density function:

$$f(y_i \mid x_i, \theta_i) = \frac{1}{\sigma_i} \exp[\frac{y_i - x_i\beta_i - \theta_i}{\sigma_i} - \exp\left(\frac{y_i - x_i\beta_i - \theta_i}{\sigma_i}\right)] \qquad i = 1,2,3 \qquad [5.1]$$

The order of magnitude of σ_i — which is a constant scale parameter — provides information on the shape of the hazard function of time-to-adoption variable *i*. The constancy of σ_i further implies that the variance of log time-to-adoption *i* is also constant. The mean value of the log of the time-to-adoption is determined by the observable heterogeneity components, x_i, and by the unobservable heterogeneity random term, θ_i. The extreme value distribution indicates that the hazard function of (log) time-to-adoption variable *i* is increasing, decreasing and constant, respectively, when σ_i<1, σ_i>1 and σ_i=1.

Correspondingly, the conditional survival function, or the probability that the adoption of technology *i* takes places beyond time *t* conditional on θ_i and x_i, is:

$$S(y_i \mid x_i, \theta_i) = \exp[-\exp\left(\frac{y_i - x_i\beta_i - \theta_i}{\sigma_i}\right)] \qquad i = 1,2,3 \qquad [5.2]$$

The assumption of independency of the y_is conditional on the unobservable heterogeneity random terms allows the joint density function of the adoption of intra-firm e-mail, inter-firm e-mail and EDI conditional on θ_i and x_i to be written as the product of separate technology-specific density and survival functions. For instance, when a firm has adopted all of the three communication technologies considered, the conditional joint probability of log duration times with respect to the adoption of technologies can be written is:

$$f(y_1, y_2, y_3 \mid \theta_i, x_i) = \prod_{i=1}^{3} f(y_i \mid \theta_i, x_i)$$

When a firm has not adopted any of the technologies, the contribution of these three joint observations to the likelihood function is the following conditional joint survival function of log duration times with respect to the adoption of technologies:

$$S(y_1, y_2, y_3 \mid \theta_i, x_i) = \prod_{i=1}^{3} S(y_i \mid \theta_i, x_i)$$

In addition, the data set incorporates six 'intermediate' cases of which either one or two of the technologies are adopted and the adoption dates of the other technologies are right censored. The conditional joint density or the likelihood function which embodies the contribution of these eight different joint observations of duration times in the sample of *n* firms can be written as follows:

$$pr(y_1, y_2, y_3 \mid \theta_i, x_i) = \prod_{i=1}^{n} f(y_1 \mid \theta_1, x_1)^{d_1} f(y_2 \mid \theta_2, x_2)^{d_2} f(y_3 \mid \theta_3, x_3)^{d_3}$$

$$S(y_1 \mid \theta_1, x_1)^{c_1} S(y_2 \mid \theta_2, x_2)^{c_2} S(y_3 \mid \theta_3, x_3)^{c_3} \qquad [5.3]$$

where d_j and c_k, j,k=1,2,3, denote indicator variables, respectively, for the adopters and non-adopters of technology *i*. The indicator variable d_j is set to 1 when a firm has adopted technology *i* and it takes the value 0 otherwise. Similarly, the indicator variable c_k takes the value 1 when a firm is the non-adopter of technology *i* and the value 0 otherwise. The unconditional joint density of the y_is can be derived by integrating the conditional joint density function — equation [5.3] — with respect to the vector of unknown densities of unobservable random heterogeneity terms $\theta = (\theta_1, \theta_2, \theta_3)$. Staiger and Wolak (1994) point out that a multitude of heterogeneity distributions can be captured by the discrete factor approximations. Applying this discrete factor approach to the duration model and assuming that θ has a discrete density (π_k, θ_k) where *k*=1,...,K denotes the number of points of support of the density of θ, π_k describes the corresponding probabilities of the points of support and $\theta_k = (\theta_{1k}, \theta_{2k}, \theta_{3k})$. Assume *k*=1,2, since the previous empirical studies suggest that satisfactory parameter estimates can be obtained by assuming a small number of points of support for the density of θ.[17] Models with *k*=1,2,3 were also estimated (not reported) and the inferences from the parameter estimates do not change much. The unconditional joint density or likelihood function of time-to-adoption variables is:

$$L(y_1, y_2, y_3) = \sum_{k=1}^{2} \pi_k \, pr(y_1, y_2, y_3 \mid \theta_k, x) \qquad [5.4]$$

The log-likelihood function of the estimated model can be obtained by taking the logarithm of equation [5.4].

[17] Staiger and Wolak (1994) find that using the number of points of support beyond K=2 does not result in a significant change in the parameter estimates and their standard deviations. They also refer to the similar evidence of previous Monte Carlo experiments using discrete factor approximations.

The econometric model, which allows correlation between the adoption decisions of different technologies by economic agents, has the attractive property that the technology-specific hazard functions are also easily attainable. The separate hazard functions regarding the adoption of different technologies, $h(y_i \mid \theta, x)$, are independent conditional on θ and thus can simply be obtained from $h(y_i \mid \theta, x) = f(y_i \mid \theta, x)$.

Before presenting the results of the estimation of model [5.4], the data set used for estimation will be briefly discussed in the following section.

Data

The data set was collected by randomly selecting 1020 firms from the membership list of the Finnish metal industry and sending a mail questionnaire to these firms in June 1995. Two separate questionnaires were designed — one for the users of advanced communication technologies (ACT) and the other one for the non-users of ACT — and the respondents were asked to choose the one they found appropriate. The response rate of the inquiry amounted to (about) 22 per cent or 220 responses in total. Of these, 124 were users of ACT and 96 were not.

Table 5.1 presents some descriptive statistics regarding the industry and ownership structure of firms in the data set. Clearly, metal engineering represents the largest group in the sample (61 per cent of all firms, 50 and 75.3 per cent of users and non-users of advanced communication systems, respectively). The groups of mechanical and electrical engineering are of equal size, but it seems that the firms in electrical engineering are more frequent adopters of ACT (30.5 per cent of the users of ACT and 5.4 per cent of the non-users of ACT) than the firms in mechanical engineering (19.5 per cent of the users of ACT and 19.3 per cent of the non-users of ACT). The adopters and non-adopters of communication technologies also seem to differ by ownership structure. Firms owned by another manufacturing company cover the largest portion of the users of ACT (58.3 per cent), whereas a high portion of the non-users of ACT are family-owned firms (72.6 per cent). A relatively small number of firms in the sample is owned by a financial company (ie by bank, insurance company or investing fund). A concentrated ownership structure is very common among the sampled firms, especially the non-adopters of ACT. The firms of which the owner has a well-diversified portfolio seem to be more often the users of ACT than the ones whose owners have poorly diversified their portfolios.

Table 5.2 describes the adoption of different communication technologies by firms in the Finnish metal industry. It shows that the diffusion speed of different communication systems varies highly when measured by the penetration rates or the number of adopters of technologies, but the differences in the average or median

Table 5.1 **Descriptive statistics on industry and ownership structure**

Variable	*Frequency (per cent of the observed values)*		
	Users	*Non-users*	*All*
Industry:			
metal engineering	59 (50.0)	70 (75.3)	129 (61.1)
mechanical engineering	23 (19.5)	18 (19.3)	41 (19.4)
electrical engineering	36 (30.5)	5 (5.4)	41 (19.4)
missing information	6	3	12
Firm is owned by:			
manager-owner	34 (31.5)	61 (72.6)	95 (49.7)
other manufacturing company	63 (58.3)	20 (23.8)	83 (43.5)
financial company	11 (9.2)	3 (3.6)	13 (4.8)
missing information	16	13	32
Ownership structure:[a]			
concentrated	88 (73.9)	86 (93.5)	174 (82.5)
dispersed	31 (26.1)	6 (6.5)	37 (17.5)
missing information	5	4	12
Diversification:[b]			
well-diversified portfolio	71 (62.8)	29 (32.2)	100 (49.3)
poorly diversified portfolio	42 (37.2)	61 (67.8)	103 (50.7)
missing information	11	6	20

a A concentrated ownership structure means that a firm is either family-owned or the two biggest shareholders own over 30 per cent of the stock. Otherwise, ownership structure is defined to be dispersed. **b** Well-diversified portfolio means that a firm is not the main investment object of its owner or two biggest shareholders. Otherwise, the portfolio is poorly diversified.

adoption years of different technologies seem less dramatic. The two most commonly adopted advanced communication technologies, inter-firm and intra-firm e-mail, were adopted by 44 per cent and 38 per cent of the firms, respectively. A smaller portion of the firms used EDI (14.5 per cent) or some other form of ACT that was not specified in the questionnaire (19 per cent).[18]

The section concludes by discussing the dependent variables and descriptive statistics of the covariates of the econometric model.

Dependent variables

This paper examines the adoption of intra-firm e-mail, inter-firm e-mail and EDI by firms, based on the time that has elapsed since the technologies were introduced until they were adopted. It is assumed that the years describing the availability of

18 These empirical estimations did not consider the adoption of communication technologies reported in the 'some other' category, since this category incorporates a number of different communication systems whose use differs between firms.

Table 5.2 **Adoption of communication technologies by Finnish firms**

Communication system	*Number of adopters* (percentage of all firms)	*Average year of adoption* (median)
E-mail (internal)	97 (44.1)	90.6 (91.0)
E-mail (external)	84 (38.2)	91.5 (92.0)
EDI	32 (14.5)	91.4 (91.5)
Some other	23 (10.5)	90.9 (93.0)

sufficient infrastructure for the utilisation of different communication systems provide a good proxy for the introduction dates of different communication systems.[19] The log durations regarding the adoption of intra-firm e-mail, inter-firm e-mail and EDI are denoted — respectively — by *iny*, *outy* and *ediy*. The explanatory variables are formulated in terms of the durations measured by the numbers of years between the introduction of a technology and the year a firm adopts that technology for the first time. These variables do not take into account whether the firm existed or not at the time of the introduction of the technology. Thus, the estimated model describes the probability that a firm adopts each of the new technologies in a given year, rather than the probability of adoption after a given spell of non-adoption.[20] The non-users of different communication technologies were treated as censored observations. These observations are coded as if the firms had adopted technologies — which they did not use at the time they responded to the questionnaire — in 1996.

Covariates

The explanatory variables in the model are discussed in section 5.2. Table 5.3 defines the covariates of the econometric model and presents the mean values and standard deviations of these variables.

[19] It is assumed that the time when local area networks (LANs) were introduced on the market well describes the year of introduction of advanced intra-firm e-mail technologies. LANs — which typically combine ACT technologies within a single firm and are used for intra-firm communications — were introduced in Finland at the beginning of the 1980s. The analog telecommunication networks were first used for inter-firm data transfers, but since they were initially designed for transmitting voice only, the reliability and speed of the network appeared to be insufficient. The first Finnish public network that was designed merely for transmitting data was introduced in 1981. This year is used to describe the introduction of advanced inter-firm communication technologies, EDI and inter-firm e-mail.

[20] As Colombo and Mosconi (1995) point out, the variables capturing calendar time are independent of the firm's existence, whereas the duration variables are not. Only when the majority of the sampled individuals/firms exist at the date of the introduction of the innovation, will calendar and duration times coincide.

Table 5.3 **Descriptive statistics of the covariates[a]**

Variable	*Definition*	*Mean*	*Standard deviation*
own1	1 if a firm is family-owned, 0 otherwise.	0.455	0.499
own2	1 if a firm is family-owned or the two biggest shareholders hold over 30 per cent of the stocks, 0 otherwise.	0.884	0.321
own3	1 if the firm is not the main investment object of its owner/two biggest shareholders, 0 otherwise.	0.530	0.501
llab	(Log) number of employees in the firm.	210.238	475.663
conc	Market share of the ten largest firms in the branch of industry in which the firm operates.	0.509	0.081
ps	The propensity score.	0.496	0.221

This completes the discussion of the potential factors determining the adoption of new communication technologies, the econometric model and data set. The next section presents the results of the empirical estimation.

5.4 Results

The log-likelihood function of the estimated model is determined by equation [5.4]. It assumes that the number of points of support of the density of unobservable heterogeneity terms is two (ie k=1,2), but the estimation results are discussed when k=1,2,3 (not presented here) when they deviate significantly from the results presented here. Table 5.4 presents the maximum-likelihood estimates of the parameters of the model.

The estimation results indicate a clear positive relationship between the variable 'own1' and the variables describing the time elapsed between the introduction of intra-firm and inter-firm communication technologies and their adoption. The manager-owned firms seem to adopt e-mail technologies later than other firms. This means that manager-owners differ from professional managers in their adoption of new communication technologies. These results suggest that agency costs are significant in the adoption of communication technologies by firms — professional managers neither bear the entire costs of the adoption of new technologies nor get the total returns of investment. Consequently, their evaluation of the expected net benefits of adopting new technologies differs from that of the manager-owner. In particular, professional managers seem to neglect the entire costs or risk associated with adoption of new communication technologies, which they can also use for their private use, and are more likely to adopt these technologies than the owners of the firm.

The estimated coefficients on the variable 'own2' suggest a negative relationship between ownership concentration and the adoption date of new communication technologies, but they are not statistically significant regarding the timing of adoption of new communication technologies. Thus, there is no evidence that ownership concentration affects the adoption of new technologies or that diffusely held companies have such managerial incentive problems that are reflected in their investment in new communication technologies. It seems also possible — in light of the evidence of the results regarding the variable 'own1' — that this dummy variable does not distinguish very clearly firms with controlling owners from those with non-controlling owners. If it does, these results may also suggest that, even if ownership structure is concentrated, notable information asymmetries remain between the owner and manager of the firm. Consequently, there is no statistically significant difference between the adoption of technologies of concentrated and dispersedly held firms.

The results concerning the variable 'own3' indicate that (portfolio) diversification does not impact in a statistically significant manner on the timing of adopting new communication technologies. The variable 'own3' — which describes the owner's concentration of wealth in one firm — is positively related to the time-to-adoption variables of e-mail technologies and negatively related to the time-to-adoption variable of EDI. None of these relationships are, however, statistically significant. The estimation results of the model with three points of support for θ (not presented here) indicate that 'own3' and the timing of adoption of e-mail technologies are positively related and statistically significant. This suggests that well-diversified investors are, to some extent, more risk averse in their investments in new

Table 5.4 **The maximum-likelihood estimates of duration models for intra-firm e-mail, inter-firm e-mail and EDI[a]**

Variable	*iny*		*outy*		*ediy*	
const	1.255	(1.189)	0.660	(0.954)	10.362	(2.701)
own1	0.291	(0.102)	0.224	(0.098)	0.035	(0.190)
own2	-0.093	(0.085)	-0.046	(0.082)	-0.216	(0.204)
own3	0.007	(0.085)	0.041	(0.089)	-0.059	(0.161)
llab	0.188	(0.131)	0.258	(0.110)	-0.784	(0.279)
conc	-0.172	(0.598)	0.477	(0.445)	-3.108	(1.330)
ps	2.042	(0.827)	2.146	(0.709)	-3.328	(1.775)
σ_i	0.230	(0.027)	0.199	(0.023)	0.312	(0.069)
No. of observations	165					
Log-Likelihood	-185.332					
Correlations	ρ_{12}=0.995					
	ρ_{13}=0.924					
	ρ_{23}=0.926					

a Standard errors in parentheses.

technologies than poorly diversified investors are.

It is clear from the results that different ownership structures involve different adoption behaviour of e-mail technologies. However, no statistically significant difference arises in the case of EDI. A reasonable explanation — which gives further support for the argument of the presence of agency problems — for this phenomenon is that EDI, unlike e-mail technologies, does not involve substantial private benefits for the managers. EDI, unlike e-mail technologies which can be used for private purposes, is used only for business transactions. Hence, EDI does not offer private consumption possibilities to professional managers. This effect may reduce the incentives of professional managers — which seem to be higher than the manager-owner or controlling shareholders in the case of e-mail technologies — to invest in EDI and their adoption behaviour does not then differ remarkably from that of firms that have a different ownership structure.

Firm size seems to accelerate the adoption of EDI, but smaller firms tend to adopt inter-firm e-mail technology earlier than bigger firms (though the estimation results of the model with k=1,2,3 do not find the relationship between the size of a firm and the adoption of inter-firm e-mail technology to be statistically significant). This result may be related to higher adoption costs — a longer period of learning and higher technical expertise — associated with EDI than e-mail. Consequently, large firms that have more financial resources or better availability of skilled labour tend to adopt EDI earlier than small firms.[21] The parameter estimates on the variable 'conc' are clearly statistically significant in the estimated model dealing with the adoption of EDI, but not for e-mail technologies. In other words, market concentration seems to accelerate the adoption of EDI. This finding is consistent with the behavioural implications of the model of Reinganum on the effect of market structure on the timing of adoption of technologies. The earlier adopters of EDI seem to be able to reap higher benefits from its use the more concentrated the industry is. It seems possible that e-mail technologies are not expected to involve as high benefits as EDI, since market concentration and the timing of adoption of e-mail technologies by firms are not statistically significantly related.

The model incorporates unobservable random heterogeneity terms that capture correlation among the time-to-adoption variables or technological complementarity between the different communication technologies. Table 5.4 presents the estimated correlations between the different technologies (see appendix 5A for details of the estimated correlations). The values for ρ_{12}, ρ_{13} and ρ_{23} in table 5.4 denote the

[21] However, it should be noted that these findings may also reflect the fact that the variable 'lab' describes the size of the firm in 1994. It does not then necessarily capture the size of the firms at the date of adoption of technologies accurately. The same caution should be applied to the interpretation of the coefficient estimate of the variable 'conc', the industry concentration ratio.

correlations between the time-to-adoption variables for intra-firm and inter-firm e-mail, intra-firm e-mail and EDI and inter-firm e-mail and EDI, respectively. The presence of technological complementarity is obvious. The estimated correlations suggest a very high degree of correlation between the adoption of different types of advanced communication technologies.

The estimated coefficients of the variable 'ps', the propensity score, are statistically significant explanatory variables of the time taken to adopt e-mail technologies. This suggests that they were necessary for correcting selection bias. The data also give support to the presence of epidemic effects in the adoption of new communication technologies by firms. The estimates of the shape parameters, σ_is, are all less than 1, indicating positive time dependence (ie that the hazard functions of durations are increasing in time).

Although the Weibull distribution allows flexibility regarding the hazard function of duration (by allowing it to be increasing, decreasing or constant in time), it is restrictive in the sense that it assumes a monotone relationship between the hazard function and time. Two graphical procedures are used to visualise the shape of the hazard functions and to check the appropriateness of the distributional assumption. Figure 5.1 presents the plots of the hazard functions of durations for intra-firm e-mail technology, inter-firm e-mail technology and EDI, respectively. The plots confirm the interpretation of the order of magnitude of the estimated shape parameters. These plots are based on $\hat{f}(y_i \mid \theta, x)$ and $\hat{S}(y_i \mid \theta, x)$. That is, the plots are based on the ML estimates of the technology-specific density functions and survival functions conditional on θ.

The assumption of the Weibull distribution for the time-to-adoption variables implies that $\ln[-\ln(S(y_i)] = (y_i - \beta_i' x_i)/\sigma_i$. This means that, if the assumption of the Weibull distribution is reasonable in the model, the plot of $\{\ln[-\ln S(y_i)], y_i\}$ should result in a roughly linear line. Figure 5.2 presents separate plots for each one of the time-to-adoption variables.

Figure 5.1 **Hazard functions for advanced communication technologies**

Figure 1 a. Hazard function for intra-firm e-mail

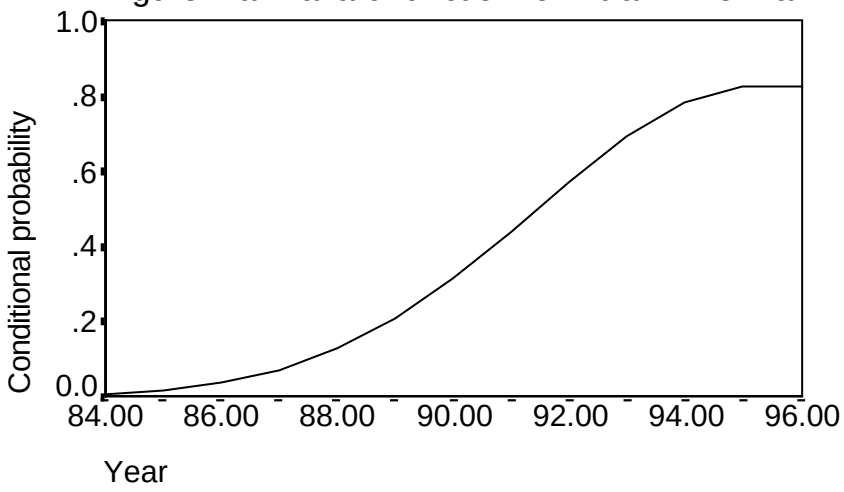


Figure 1 b. Hazard function for inter-firm e-mail

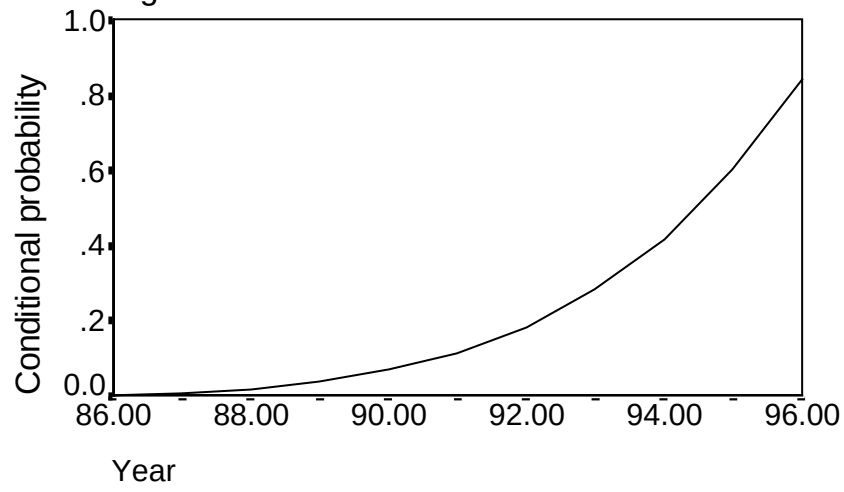


Figure 1 c. Hazard function for EDI

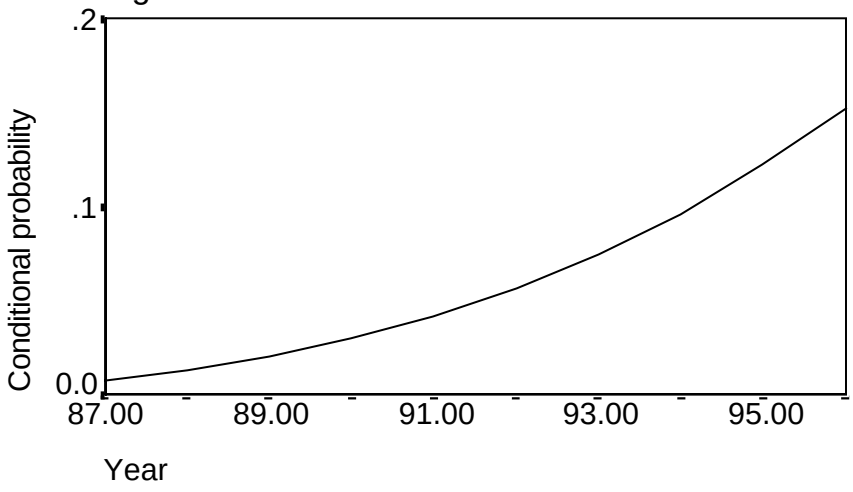
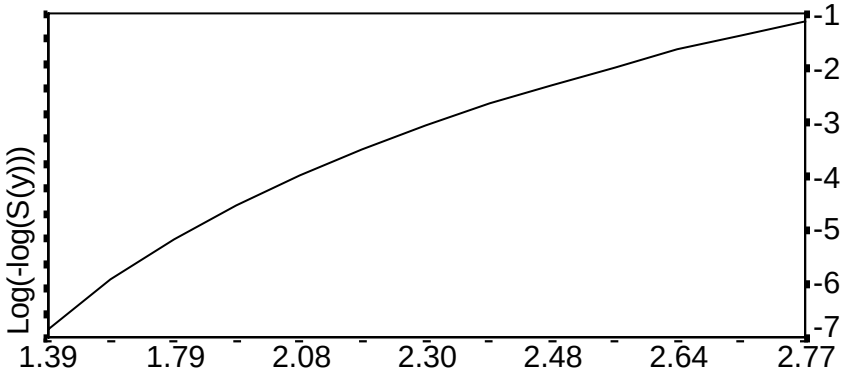


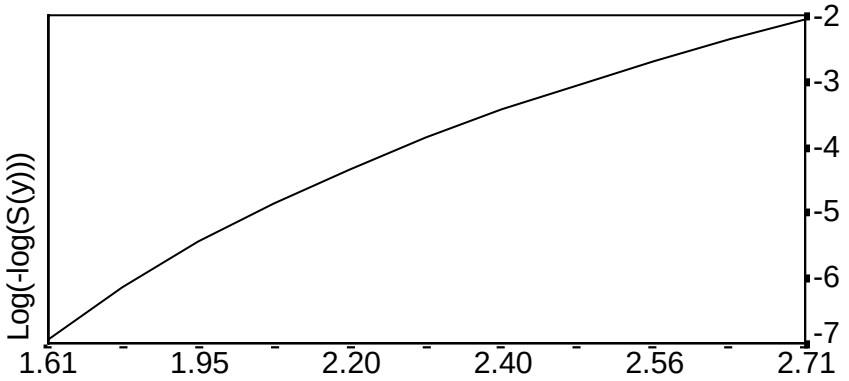
Figure 5.2 **Time-to-adoption for advanced communication technologies**

Figure 2 a. Plot of log(-log(S(y))) vs. log(y)
for intra-firm e-mail



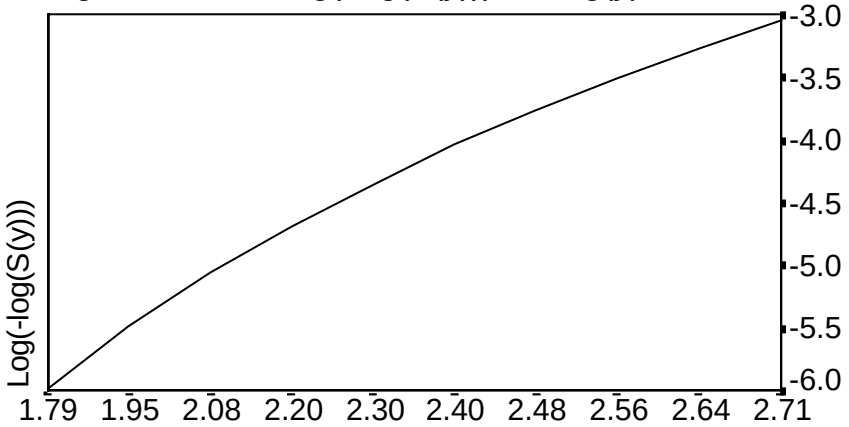
Log duration

Figure 2 b. Plot log(-log(S(y))) vs. log(y)
for inter-firm e-mail



Log duration

Figure 2 c. Plot log(-log(S(y))) vs. log(y) for EDI



Log duration

All of the plotted lines are approximately linear, even though the plot for inter-firm e-mail technology seems to be slightly concave. Thus, the assumption of the Weibull distribution of time-to-adoption variables is adequate and it is appropriate to use the corresponding extreme value distribution for the log time-to-adoption variables in the estimated models.

5.5 Conclusions

This paper has econometrically modelled and tested a decision on the timing of adopting multiple communication technologies by firms. The results suggest that the ownership structure of a firm has a substantial impact on the adoption of new communication technologies. Manager-owners differ with regard to their adoption behaviour of new communication technologies from the more diffusely held firms, where professional managers may decide about the adoption of technologies. This evidence supports the argument that agency problems may have noticeable implications for the adoption of new technologies. Since professional managers bear fewer risks and costs of adopting new technologies, and that they may derive private benefits from the use of the technology, they tend to adopt these technologies earlier than the manager-owners do.

There is significant interdependence between different communication technologies. The adoption decisions of firms regarding different communication technologies appears to be highly correlated with each another. These findings suggest that the different communication technologies embrace substantial complementarities. Consequently, it is important to incorporate technological complementarities into economic and econometric models of the adoption of new technologies.

In summary, this study illustrates that several internal and external factors identified in previous economic studies are relevant to the adoption of new technology by firms. The results find evidence for:

1. the epidemic models that suggest that the probability of the adoption of technologies increases with time;
2. the game-theoretic studies of adopting new technologies by firms that stress the importance of the competitive environment; and
3. the new economic theory of the firm that suggests that informational imperfections, separation of ownership and control, and self-interested behaviour of the individual actors have a decisive role in the behaviour of a firm.

The last finding suggests that this important topic, which is largely neglected by the literature on the adoption of new technologies by firms, deserves more profound

attention among the economists in this field. It would be particularly interesting to see how technology-specific and industry-specific the effect of ownership structure on the adoption of new technologies is.

Appendix 5A

The correlations between duration variables for intra-firm and inter-firm e-mail, intra-firm e-mail and EDI and inter-firm e-mail and EDI — ρ_{12}, ρ_{13} and ρ_{23} respectively — were estimated as follows. The error terms of the model consist of two components: $v_i = \varepsilon_i + \theta_i = \varepsilon_i + \rho_i \theta^*$, i=1,2,3. The estimated parameter ρ_i determines how the unobservable random heterogeneity term θ^* affects the mean value of duration of technology i. Assume that the ε_i s are i.i.d. with means of:

$$x_i \beta + \theta_i + \gamma\theta_i$$

where γ = 0.5722 (Euler's constant), and are mutually independent with variance σ^2 and that θ^* has zero mean[22] and a covariance matrix Σ. Correlation between the error terms v_i and v_j, $i \neq j$, can then be defined as:

$$\rho_{ij} = \mathrm{cov}(e_i + \rho_i\theta^*, e_j + \rho_j\theta^*) / \sqrt{\mathrm{var}(\varepsilon_i) + \mathrm{var}(\rho_i\theta^*)}\sqrt{\mathrm{var}(\varepsilon_j) + \mathrm{var}(\rho_j\theta^*)}$$

Since ε_i and ε_j are distributed with the extreme value distribution, where N denotes the number of observations and π=3.14159,. $\mathrm{var}(\varepsilon_i) = \frac{1}{6}\frac{\pi^2\sigma_i^2}{N}$ can be similarly defined. The discrete density assumption for θ^*, (π_k, θ_k^*), k=1,2, implies that $\mathrm{var}(\theta^*) = \pi_1(\theta_1^*)^2 + \pi_2(\theta_2^*)^2$. Thus, the estimated correlation between the error terms v_i and v_j, ρ_{ij}, is:

$$= \rho_i\rho_j E(\theta^{*2}) / \sqrt{\frac{1}{6}\frac{\pi^2\sigma_i^2}{N} + \rho_i^2(\pi_1(\theta_1^*)^2 + \pi_2(\theta_2^*)^2)}\sqrt{\frac{1}{6}\frac{\pi^2\sigma_j^2}{N} + \rho_j^2(\pi_1(\theta_1^*)^2 + \pi_2(\theta_2^*)^2)}$$

The estimated values for correlations and their standard errors are obtained by the delta method from the estimated parameter values.

[22] Note that the restriction that θ^* has zero mean imposes in the estimation that $\theta_2 = -(\pi_1\theta_1)/(1-\pi_1)$, since $\pi_1\theta_1 + \pi_2\theta_2 = 0$.

References

Colombo, M.G. and Mosconi, R. 1995, 'Complementarity and cumulative learning effects in the early diffusion of multiple technologies', *Journal of Industrial Economics*, vol. 43, no. 1, pp. 13–48.

Fama, E.F. 1980, 'Agency problems and the theory of the firm', *Journal of Political Economy*, vol. 88, no. 2, pp. 288–307.

Greenwald, B.C. and Stiglitz, J.E. 1990, 'Asymmetric information and the new theory of the firm: Financial constraints and risk behaviour', *American Economic Review*, vol. 80, no. 2, pp. 160–5.

Hannan, T.H. and McDowell, J.M. 1984, 'The determinants of technology adoption: The case of the banking firm', *Rand Journal of Economics*, vol. 15, no. 3, pp. 328–35.

Holmström, B.R. and Tirole, J. 1989, 'The theory of the firm', Chapter 2 in Schmalensee, R. and Willig, R.D. (eds), *Handbook of Industrial Organization*, vol. 1, North Holland, Amsterdam, pp. 61–133.

Ihamuotila, M. 1994, 'Corporate ownership, capital structure and investment', Helsinki School of Economics and Business Administration, A:97.

Jensen, M.C. and Meckling, W.H. 1976, 'Theory of the firm: Managerial behavior, agency costs and ownership structure', *Journal of Financial Economics*, vol. 3, no. 4, pp. 305–60.

Karshenas, M. and Stoneman, P.L. 1993, 'Rank, stock, order, and epidemic effects in the diffusion of new process technologies: An empirical model', *Rand Journal of Economics*, vol. 24, no. 4, pp. 503–28.

Lawless, J.F. 1982, *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons.

Loh, L. and Venkatraman, N. 1993, 'Corporate governance and strategic resource allocation: the case of information technology investments', *Accounting, Management & Information Technologies*, vol. 3, no. 4, pp. 213–28.

McEachern, W.A. and Romeo, A.A. 1978, 'Stockholder control, uncertainty and the allocation of resources to research and development', *Journal of Industrial Economics*, vol. 26, no. 4, pp. 349–61.

Mansfield, E. 1968, *Industrial Research and Technological Innovation, An Econometric Analysis*, W.W. Norton & Company, New York.

Milgrom, P. and Roberts, J. 1990, 'The economics of modern manufacturing: Technology, strategy, and organization', *American Economic Review*, vol. 80, no. 3, pp. 511–28.

—— 1992, *Economics, Organization and Management,* Prentice-Hall, Englewood Cliffs, New Jersey.

Quirmbach, H.C. 1986, 'The diffusion of new technology and the market for an innovation', *Rand Journal of Economics*, vol. 17, no. 1, pp. 33–47.

Reinganum, J. 1981, 'On the diffusion of new technology: A game theoretic approach', *Review of Economic Studies,* vol. 48, no. 3, pp. 395–405.

Rose, N.L. and Joskow, P.L. 1990, 'The diffusion of new technologies: Evidence from the electric utility industry', *Rand Journal of Economics*, vol. 21, no. 3, pp. 354–73.

Rosenbaum, P.R. and Rubin, D.B. 1983, 'The central role of the propensity score in observational studies for causal effects', *Biometrika,* vol. 70, no. 1, pp. 41–55.

—— 1985, 'Constructing a control group using multivariate matched sampling methods that incorporate the propensity score', *American Statistical Association,* vol. 39, pp. 33–42.

Shleifer, A. and Vishny, R.W. 1986, 'Large shareholders and corporate control', *Journal of Political Economy*, vol. 94, no. 3, pp. 461–88.

Staiger, R.W. and Wolak, F.A. 1994, 'Measuring industry-specific protection: Antidumping in the United States', *Brookings Papers*, Microeconomics, pp. 51–103.

Stoneman, P. 1983, *Economic Analysis of Technological Change,* Oxford University Press, Oxford.

Stoneman, P. and Kwon, M.J. 1994, 'The diffusion of multiple process technologies', *Economic Journal*, vol. 104, no. 423, pp. 420–31.

Wellenius, B., Miller, A. and Dahlman, C.J. (eds) 1993, *Developing The Electronics Industry*, A World Bank Symposium, World Bank, Washington DC.

6 Horizontal mergers and delegation

Steffen Ziss

This paper shows that the profitability of merger is significantly enhanced in the presence of delegation regimes. The source of the increased profitability is the softening of incentive competition brought about by merger. Two consequences of increased profitability are that the minimum market share that the merging parties require in order to merge profitably without efficiency gains, as well as the maximum market share that the merging parties can possess in order to guarantee that a profitable merger is welfare enhancing, are reduced.

6.1 Introduction

It is well established that anti-competitive mergers are generally unprofitable in Cournot oligopoly models in which marginal cost is constant. In particular, Levin (1990) has shown that, in the absence of cost efficiency gains, any output-reducing merger is profitable only if it involves more than 50 per cent of the firms in the industry.[1] This result emerges because any internalisation of market power gains brought about by a contraction in output of the merging parties is more than offset by an expansion in output of the non-merging parties. Several authors have pointed out that the profitability of merger in output setting regimes is improved if the loss of market share to non-merging parties is muted, as will be the case if marginal costs are increasing (Perry and Porter 1985), products are differentiated (Deneckere and Davidson 1983) or merger bestows upon the merged entity a first mover advantage (Daughety 1990 and d'Aspremont et al. 1983). Alternately, the profitability of merger is enhanced if the merger results in cost efficiency gains (Perry and Porter 1985, Levin 1990, Farrell and Shapiro 1990 and McAfee and

[1] This result also follows from the analysis of Kamien and Zang (1990). Gaudet and Salant (1991) and Cheung (1992) derive similar results for a slightly different set of demand assumptions. It should be pointed out that a 50 percent market share is a necessary, but not sufficient, condition for profitable merger. For example, if demand is linear then Salant, Switzer and Reynolds (1983) have shown that profitable merger requires the merging parties' market share to be at least 80 percent.

Williams 1992). Finally, the profitability of merger is substantially altered if products are differentiated and firms compete in prices instead of outputs. In particular, Deneckere and Davidson (1985) show that any merger is profitable under differentiated price competition regimes.[2]

A common feature of oligopoly markets which is ignored in most models dealing with horizontal merger is that firms operating in such markets are often large and thus find it advantageous to delegate the output decision to a manager or to divest themselves of either their downstream or upstream operations. Moreover, it is well known that delegation or divesture creates an agency setting which a firm can exploit in order to shift rents via the terms of the incentive contract offered to agent.[3] The purpose of this paper is to argue that the profitability of merger is significantly enhanced in the presence of delegation regimes. The source of the increased profitability is the softening of incentive competition brought about by merger. Two consequences of increased profitability are that the minimum market share required by the merging parties to merge profitably without efficiency gains, as well as the maximum market share that the merging parties can possess in order to guarantee that a profitable merger is welfare enhancing, are reduced.

An important assumption in the model in this paper is that the agent's contract specifies both a fixed and a variable component. The variable component is set so as to maximise surplus whereas the fixed component is used to transfer surplus. The importance of this assumption for merger is that the firm and the agent each receive a share of any increase in surplus created by merger.[4] The results of Horn and Wolinsky (1988) show that this will no longer be the case if the contract between firm and agent specifies only a linear variable component and no fixed component. In their model, the firm is an upstream supplier or union and the agent is a downstream firm. It is shown that if the terms of the linear price is determined via a Nash bargaining game then a merger of two upstream suppliers to form an upstream

2 In their review of the empirical literature Scherer and Ross (1990, pp. 172–4) conclude that there is at best only weak evidence that horizontal mergers are profitable. This evidence is difficult to reconcile with the differentiated price competition result that any merger is profitable. On the other hand, this evidence is consistent with the Cournot result that profitable merger either requires sufficiently large cost efficiency gains or sufficiently large market share. In particular, the weak evidence of profitability can be explained in a Cournot context either by firms fearing antitrust reprisals and therefore not going ahead with large mergers or by firms not realising the cost efficiency gains that they had anticipated prior to merger.

3 The owner-manager agency setting is discussed in Fershtman (1985), Vickers (1985), Sklivas (1987) and Fershtman and Judd (1987) whereas the manufacturer-retailer setting is discussed in Lin (1988) and Bonanno and Vickers (1988).

4 The firm's share of any increase in surplus arising from merger is determined by its exogenously given bargaining strength.

monopoly can result in adverse bargaining effects.[5] In other words, the gains from internalising market power are not necessarily evenly distributed when the contract between firm and agent specifies a linear price as opposed to a two part-tariff. As a result of the adverse distribution of the gains from merger, the upstream parties may be deterred from merging. Since merger for monopoly is always profitable under a non-delegation regime, then the results in Horn and Wolinsky (1988) testify to the fact that the result of this paper — that delegation enhances the profitability of merger — does not necessarily hold when the agent's contract specifies a linear price rather than a two-part tariff.

A second important assumption of the model in this paper is that firms choose incentive contracting as the rent-shifting tool. An alternate method of rent shifting, which has been suggested by a number of authors including Baye, Crocker and Ju (1996), involves the establishing of autonomous horizontal divisions within the firm. In this setting the terms of the incentive contract offered to each divisional manager is not set strategically. Instead each divisional manager is charged with maximising divisional profits and acts independently of the other divisional managers. The intra-firm competition created by divisionalisation allows the firm to commit to more output, which then deters rival output and shifts rents.[6] The effect of mergers under this so-called 'divisionalisation regime' is similar to what it is under the 'strategic contracting regime' considered in this paper. In particular, mergers serve to reduce the incentive to shift rents. Under a divisionalisation regime, this reduction in rent shifting results in a reduction in a number of divisions per firm[7] as opposed to a softening of incentive competition. The important difference between divisionalisation and strategic incentive contracting regimes is that divisionalisation is assumed to be costly. As a result, a reduction in rent shifting brought about by merger under a divisionalisation regime also implies a cost saving for both merging and non-merging firms. This endogenous cost saving improves the profitability of merger and enhances the likelihood that a profitable merger will be welfare enhancing.[8] Consequently, the result that delegation enhances the profitability of merger continues to hold if firms choose divisionalisation rather than incentive contracting as the rent shifting tool. On the other hand, the result that

[5] Whether or not merger is profitable for the upstream suppliers depends on whether the merged upstream supplier bargains sequentially or simultaneously with the downstream firms.

[6] Alternately, the effects of divisionalisation can be achieved via franchising. In order to establish intra-firm (or intra-brand) competition in this setting the franchise licences must specify non-exclusive territories and must be sold to independent owners.

[7] See Baye, Crocker and Ju (1996).

[8] Under fairly general assumptions Baye, Crocker and Ju (1996) show that duopoly is the market structure that maximises welfare in the absence of price regulation. As a result, any merger that does not result in monopoly will be welfare enhancing.

delegation reduces the set of profitable mergers that are welfare enhancing is overturned under a divisionalisation regime.

A final group of papers that consider mergers under a delegation regime are those in the 'common agency' literature.[9] The mergers considered in this literature differ from this paper in that they involve mergers between downstream agents rather than between upstream firms. These mergers can either be brought about by the downstream firms themselves or by the upstream firms choosing to employ a common agent rather than a set of independent agents. If the agent's contract specifies a two part-tariff then common agency can either serve to intensify or soften wholesale price competition, depending on the nature of the agent's break even constraint.[10] Furthermore, upstream firms will prefer common agency to independent agents depending on whether wholesale price competition softens or intensifies and on whether the countervailing power of the common agent increases or remains the same.[11] If the agent's contract specifies a linear price then Horn and Wolinsky (1988) show that a merger of agents will not be profitable for the downstream agents due to adverse bargaining effects. Both the linear and two-part pricing results obtained in the common agency literature imply that this result that delegation improves the profitability of merger does not necessarily hold when it is the agents, as opposed to the firms, that merge.

This paper is organised as follows. Section 6.2 derives the profitability results. Section 6.3 introduces a diagram to discuss the welfare consequences of mergers under delegation regimes. Section 6.4 offers some concluding remarks and an avenue for future research.

9 Papers in this literature include Bernheim and Whinston (1985), Lin (1990), Ziss (1995), O'Brien and Shaffer (1993) and Horn and Wolinsky (1988).

10 If upstream firm *i* sets the fixed component of the two part-tariff so as to make the common retailer break even on product *i* then Lin (1990) and Ziss (1995) show that common agency serves to intensify wholesale price competition. If upstream firm *i* sets the fixed fee so as to make the common retailer break even on the sale of products *i* and *j* overall then O'Brien and Shaffer (1993) and Bernheim and Whinston (1985) show that common agency softens wholesale price competition.

11 Upstream firms prefer independent agents instead a common agent either if: common agency softens wholesale price competition and the common agent acquires countervailing power by credibly refusing to sell one of the products (O'Brien and Shaffer 1993) or if common agency intensifies incentive competition and the common agent does not acquire countervailing power (Lin 1990, Ziss 1995). On the other hand, if common agency softens wholesale price competition and does not result in countervailing market power for the agent then Bernheim and Whinston (1985) show that upstream firms prefer a common agent to independent agents.

6.2 The model and the profitability results

A comparison of the profitability of merger under delegation and non-delegation regimes is analysed in a model consisting of n identical firms. Each firm produces a homogeneous product at constant marginal cost c and zero fixed cost. If Q and q_i represent industry and firm output, respectively, and if $P(Q)$ denotes the inverted demand function, then the profits of product i[12] are given by

$$\pi_i(q_i, Q) = (P(Q) - c)q_i \qquad i = 1,\ldots,n \qquad [6.1]$$

The inverted demand function $P(Q)$ is assumed to be strictly decreasing and twice continuously differentiable. Furthermore, it is assumed that

$$2P'(Q) + QP''(Q) < 0 \qquad [6.2]$$

which ensures that industry marginal revenue is downward sloping and thus that the second order conditions are satisfied for any market structure. Equation [6.2] is equivalent to $\beta(Q) > -2$ where $\beta(Q) \equiv P''(Q)Q/P'(Q)$ is defined as the degree of concavity of demand.[13]

Delegation regime

Under the delegation regime, each firm delegates the output decision to a downstream agent such as a manager or a separate downstream entity.[14] The firm's contract with the agent specifies: a variable, w_i,[15] that determines the marginal cost

---

12 Under a delegation regime in which the agent is a separate downstream entity, c will refer to the sum of upstream and downstream marginal cost and profit refers to revenues minus the sum of upstream and downstream costs.

13 In a symmetric model the condition that ensures that outputs are strategic substitutes is $\beta \geq -n$. If $n > 2$ then equation [6.2] implies, and is stronger than, the strategic substitutes condition. In the merger literature the strategic substitutes assumption has been used by Farrell and Shapiro (1990), Levin (1990) and Gaudet and Salant (1991), whereas the downward sloping industry marginal revenue assumption has been used by Kamien and Zang (1990), Cheung (1992) and Faulí-Oller (1997).

14 These results also hold if the agent is an upstream firm. In this case, the output decision is made by the firm rather than by the agent. The role played by the agent in this setting is to provide the output demanded by the downstream firm at the agreed upon transfer price.

15 If the agent is a separate downstream entity with marginal cost then w_i is determined by the wholesale price. For example, if the wholesale price is set below upstream marginal cost then the marginal cost faced by the downstream firm will be below the sum of upstream and downstream cost (ie $w_i < c$). If the agent is a manager and if the firm bases the manager's compensation on a convex combination of profits and revenues (see Fershtman and Judd 1987) then w_i is

faced by the agent, and one or more rent-sharing components that are used to divide the oligopoly rents between the firm and the agent. The rent-sharing components consist of a fixed payment T_i and a share parameter σ_i. The fixed component can either be interpreted as a salary paid *to* the agent, if the agent is a manager, or as a fixed licensing fee paid *by* the agent if the agent is a separate downstream entity. The agent's payoff, denoted A_i, is thus given by

$$A_i = T_i + \sigma_i[(P(Q) - w_i)q_i] \qquad i = 1,\ldots,n \qquad [6.3]$$

where $\sigma_i = 1$ if the agent is a separate downstream entity.[16] The firm's payoff, denoted F_i, is given by profits net of the agent's payoff

$$F_i = \pi_i(q_i, Q) - A_i \qquad i = 1,\ldots,n \qquad [6.4]$$

The n firms play a two-stage game that involves simultaneous choice in each stage. In the first stage, each firm chooses the terms of the contract (w_i, T_i, σ_i) and in the second stage, the agents choose output q_i. This game is solved using backward induction.

Output stage

Differentiating equation [6.3] with respect to q_i yields the following first order condition associated with the agent's choice of output

$$P(Q) + q_i P'(Q) - w_i = 0 \qquad i = 1,\ldots,n \qquad [6.5]$$

which yields solutions denoted $q_i(w)$ and $Q(w)$ where $w = (w_1,\ldots,w_n)$. In appendix 6A it is shown that

$$\frac{\partial Q(w)}{\partial w_i} = \frac{1}{(n+1)P'(Q) + QP''(Q)} \qquad i = 1,\ldots,n \qquad [6.6]$$

and

$$\frac{\partial q_i(w)}{\partial w_i} = \frac{n + \frac{n-1}{n}\beta(Q)}{(n+1)P'(Q) + QP''(Q)} \qquad i = 1,\ldots,n \qquad [6.7]$$

determined by the profit share θ_i (ie if $\theta_i = 1$ then $w_i = c$, if $\theta_i <$ (resp. >) 1 then $w_i <$ (resp. >) c).

16 Since there is no uncertainty to and thus no risk sharing or moral hazard in this model, then the only role of the σ_i parameter is avoid the problem of a manager being paid a negative salary.

Contract stage

Since the contract specifies a fixed fee then regardless of whether the terms of the contract are specified unilaterally by the firm or arise as the outcome of a bargaining game, it will be the case that w_i will be set so as to maximise profits.[17] Substituting $q_i(w)$ and $Q(w)$ into equation [6.1] and then differentiating with respect to w_i yields that the profit maximising choice of w_i satisfies

$$(P(Q) - c)\frac{\partial q_i(w)}{\partial w_i} + q_i P'(Q)\frac{\partial Q(w)}{\partial w_i} = 0 \qquad i = 1,\ldots,n$$

Divide by $\frac{\partial q_i(w)}{\partial w_i}$ and let $\frac{\partial Q}{\partial q_i}(Q,n) \equiv \frac{\partial Q(w)/\partial w_i}{\partial q_i(w)/\partial w_i}$ denote the change in industry output per change in own output as elicited from a change in w_i. Now substitute $q_i = Q/n$ to obtain that the symmetric equilibrium level of Q solves

$$P(Q) + \frac{QP'(Q)}{n}\frac{\partial Q}{\partial q_i}(Q,n) - c = 0 \qquad [6.8]$$

where equations [6.6] and [6.7] imply

$$\frac{\partial Q}{\partial q_i}(Q,n) \equiv \frac{\partial Q/\partial w_i}{\partial q_i/\partial w_i} = \frac{1}{n + \frac{n-1}{n}\beta(Q)} \qquad i = 1,\ldots,n \qquad [6.9]$$

---

17 The explanation is as follows. Re-write the payoffs as $A_i = x_i + T_i$ and $F_i = y_i - T_i$ where x_i and y_i represent the non-fixed component of the agent's and firm's payoff, respectively, and where $x_i + y_i = \pi_i$. Now let the agent's opportunity cost equal $\overline{A}$. If the firm unilaterally chooses the terms of the contract then it will set $T_i = \overline{A} - x_i$ which leaves the agent with a payoff equal to $\overline{A}$ and the firm with a payoff equal to $F_i^* = \pi_i - \overline{A}$. The firm then chooses w_i so as to maximise F_i^*. Since $\overline{A}$ is constant, then the w_i that maximises F_i^* also maximises π_i. Now suppose that the firm and agent bargain over the terms of the contract. Furthermore, suppose that the threat point payoffs for the agent and the firm are given by $\overline{A}$ and zero, respectively. A Nash bargaining solution is then a contract that maximises the Nash products given by $\phi_i = (x_i + T_i)^{\alpha_i}(y_i - T_i)^{1-\alpha_i}$ where $\alpha_i \in (0,1)$ is a measure of the agent's bargaining power. Differentiating ϕ_i with respect to T_i yields $T_i = \alpha_i y_i - (1-\alpha_i)(x_i - \overline{A})$ which implies payoffs of $\alpha_i(\pi_i - \overline{A})$ for the agent and $(1-\alpha_i)(\pi_i - \overline{A})$ for the firm. Substituting these values back into ϕ_i yields an equilibrium Nash product given by $\phi_i^* = \alpha_i^{\alpha_i}(1-\alpha_i)^{1-\alpha_i}(\pi_i - \overline{A})$. The equilibrium choice of w_i is the one that maximises ϕ_i^*. Since ϕ_i^* is a linear function of π_i then the w_i that maximises ϕ_i^* will also maximise π_i. As a result, the equilibrium choice of w_i maximises π_i regardless of whether w_i is decided unilaterally by the firm or in a bargaining context.

In order to establish the link between merger and incentive competition substitute $q_i = Q/n$ and equation [6.8] into equation [6.5] to obtain

$$w_i - c = \frac{Q}{n}\left(-P'(Q)\right)\left(\frac{\partial Q}{\partial q_i}(Q,n) - 1\right) \qquad i = 1,\ldots,n \qquad [6.10]$$

Since equations [6.2] and [6.9] imply that $\frac{\partial Q}{\partial q_i}(Q,n) < 1$ then equation [6.10] yields the result that $w_i < c$. This marginal cost distortion is beneficial as it allows the firm to deter rival output and commit itself to the Stackelberg outcome. Equation [6.10] also reveals that the size of firm *i*'s distortion depends on: the extent to which rival output is deterred (ie $\frac{\partial Q}{\partial q_i}(Q,n) - 1$),[18] the extent to which reductions in rival output affect price (ie $-P'(Q)$) and on firm *i*'s output (ie Q/n). Merger increases Q/n (*increased output effect*), has an ambiguous effect on $-P'(Q)$ (*ambiguous slope effect*) and reduces n. If the degree of concavity of demand is constant (ie $\beta(Q) = \beta$) then equation [6.9] implies that a reduction in n brought about by merger implies a reduction in the deterrence effect of aggressive incentive competition (ie if $\beta(Q) = \beta$ then a reduction in *n* reduces $\frac{\partial Q}{\partial q_i}(Q,n) - 1$) (*reduced deterrence effect*). Merger will thus soften incentive competition if the reduced deterrence effect dominates the slope and output effects.

Merger analysis

Now let the subscript $r \in \{D, ND\}$ denote whether the regime is a delegation (*D*) or non-delegation (*ND*) regime and let the equilibrium profits under regime r when there are n competing units be denoted $\pi_r(n)$.[19] Since marginal cost is constant, a merger involving $k+1$ firms under either regime will result in the merged entity operating only one of the firms involved in the merger and shutting down the remaining k firms. Under a delegation regime, this implies that the merged entity will use only a single agent to sell its output and will abandon the remaining k agents. If $N \geq 2$ represents the number of firms in the pre-merger setting then a

---

18 Since w_i only shifts the reaction function of firm *i*'s agent then $\partial Q/\partial q_i = 1 + \partial Q_{-i}/\partial q_i$ or $\partial Q/\partial q_i - 1 = \partial Q_{-i}/\partial q_i$, where $\partial Q_{-i}/\partial q_i$ is the slope of the collective reaction function of the agents of firm *i*'s rivals.

19 Under the delegation regime the equilibrium profits are those that arise from the firms' equilibrium choice of incentives and from the agents' equilibrium choice of outputs given the incentive structure. Given the aggressiveness of incentive competition, it will be the case that equilibrium profits will be lower under delegation regimes than under non-delegation regimes for any given *n*.

merger of $k+1$ firms is assumed to result in: no change in behaviour of the competing firms[20] or in the level of marginal cost of any firm; and a reduction in n from N to $N-k$. Under either regime, merger is profitable provided the collective pre-merger payoffs of the merging firms exceed the post-merger profit of the single merged entity. Under a non-delegation regime, the pre-merger payoffs of the merging firms are given by $(k+1)\pi_{ND}(N)$, whereas the post-merger payoff of the single merged entity is given by $\pi_{ND}(N-k)$. Under a delegation regime, the pre- and post-merger payoffs of the merging parties are $(k+1)(1-\alpha_i)\pi_D(N)$ and $(1-\alpha_i)\pi_D(N-k)$ respectively, where α_i represents the bargaining strength of the agent.[21] Comparing the pre- and post-merger profits of the merging parties yields that merger is profitable under either regime provided:[22]

$$\frac{\pi_r(N-k)}{\pi_r(N)} > k+1 \qquad r \in \{D, ND\} \qquad [6.11]$$

Now assume that the degree of concavity of demand is constant (ie $\beta(Q)=\beta$) and define $F_r(N,k,\beta)$ as the ratio of the post-merger to pre-merger equilibrium profits. That is

$$F_r(N,k,\beta) \equiv \frac{\pi_r(N-k)}{\pi_r(N)} \qquad r \in \{D, ND\} \qquad [6.12]$$

If $\beta \neq 1$[23] then

$$F_D(N,k,\beta) = \frac{N}{N-k}\left(\frac{N^2+N\beta+1}{(N-k)^2+(N-k)\beta+1}\right)^{\frac{2+\beta}{1+\beta}}\left(\frac{(N-k)^2+(N-k)\beta-\beta}{N^2+N\beta-\beta}\right)^{\frac{1}{1+\beta}} \qquad [6.13]$$

---

20 See Daughety (1990), Levin (1990) and d'Aspremont et al. (1983) for models that allow for the behaviour of the merging firms to change following the merger. For example, Daughety (1990) allows for the possibility that a set of follower firms may become leaders as a result of the merger.

21 For a derivation of these bargaining payoffs see footnote 17. In deriving these payoffs it is assumed the threat point payoff of the agents to equal zero (ie $\overline{A}=0$). Setting $\overline{A}=0$ implies that there will be no fixed costs savings under a delegation regime from using 1 instead of k + 1 agents. The reason for setting $\overline{A}=0$ is that it allows the profitability of merger under delegation and non-delegation regimes in the absence of any cost efficiency gains under either regime to be compared.

22 This condition ignores the strategic issues associated with merger via acquisition that are discussed in Kamien and Zang (1990).

23 See appendix 6C and Faulí-Oller (1997), respectively, for the delegation and non-delegation versions of $F_r(N, k, -1)$.

under a delegation regime (see appendix 6B) and

$$F_{ND}(N,k,\beta)=\left(\frac{N}{N-k}\right)^{\frac{\beta}{1+\beta}}\left(\frac{N+1+\beta}{N-k+1+\beta}\right)^{\frac{2+\beta}{1+\beta}} \qquad [6.14]$$

under a non-delegation (Cournot) regime (see Faulí-Oller 1997). Appendix 6C shows that under both regimes merger is profitable if there are enough firms involved in the merger (see lemmas 6.1 and 6.2 in appendix 6C). Furthermore, a comparison of the two regimes yields the following proposition.

Proposition 6.1

If marginal cost and the degree of concavity of demand are constant, and if industry marginal revenue is downward sloping, and if mergers do not create efficiency gains, then the minimum market share required for merger to be profitable:

(i) cannot be greater for a delegation regime than for a non-delegation regime;

(ii) is non-decreasing in β under a delegation regime;

(iii) goes to 100 per cent as $\beta \to \infty$ under a non-delegation regime; and

(iv) goes to 100 per cent $\left(1-\frac{\sqrt{1+4N}-1}{2N}\right)$ as $\beta \to \infty$ under a delegation regime.

For a proof of proposition 6.1 see appendix 6D.

Part (i) of proposition 6.1 implies that the dominant effect of mergers under a delegation regime is to reduce the number of rivals faced by any one firm. (ie reduced deterrence effect dominates output and slope effects). This reduction in rivalry reduces the aggregate responsiveness of any firm's rivals to the rent shifting efforts of that firm and thus reduces the benefits of aggressive incentive competition *for both merging and non-merging firms*. As a result, merger under a delegation regime serves to soften incentive competition. This softening effect enhances the profitability of merger by mitigating the output expansion of rival firms in response to the merger. Under non-delegation regimes, there is no incentive competition and thus this incentive softening effect is absent. As a result, delegation reduces the number of firms required for merger to be profitable. Evidence of the extent of this reduction is contained in table 6.1.

Table 6.1 **Minimum market share required for profitable merger[a]**

(per cent)

No. of firms pre-merger (N)	*Non-delegation regime*			*Delegation regime*		
	β=-1.99	β=0	β=10	β=-1.99	β=0	β=10
3	66.6	100	100	66.6	66.6	66.6
4	75.0	100	100	50.0	50.0	75.0
5	80.0	80.0	100	40.0	60.0	60.0
6	83.4	83.4	100	50.0	50.0	66.6
7	71.5	85.8	100	42.9	57.1	71.5
8	75.0	87.5	100	50.0	62.5	62.5
9	77.8	88.9	100	55.6	55.6	66.6
10	80.0	90.0	100	50.0	60.0	70.0
100	91.0	92.0	95.0	77.0	78.0	80.0
1000	96.9	97.0	97.4	89.7	89.8	90.1

a The results reported in this table hold for any demand function that exhibits constant concavity and for any level of marginal cost. In deriving these results, integer constraints were imposed. In particular, the lowest integer was found to satisfy the inequality given in equation [6.11].

The remaining parts of proposition 6.1 deal with the effect of concavity on the profitability of merger. Faulí-Oller (1997) has shown that under Cournot regimes, increases in concavity increase the minimum market share required for mergers that do not generate efficiency gains to be profitable. Part (ii) of proposition 6.1 confirms that this is also true of delegation regimes. Part (iii) of proposition 6.1 extends the analysis in Faulí-Oller (1997) to show that as the degree of concavity goes to infinity, only merger for monopoly is profitable under a non-delegation regime. The simulations in table 6.1 show that if markets are concentrated then $\beta = 10$ will be sufficient for the '*only merger for monopoly is profitable*' result to hold under non-delegation regimes. In contrast, part (iv) of proposition 6.1 points out that in order for this result to hold under a delegation regime both β and N must go to infinity.

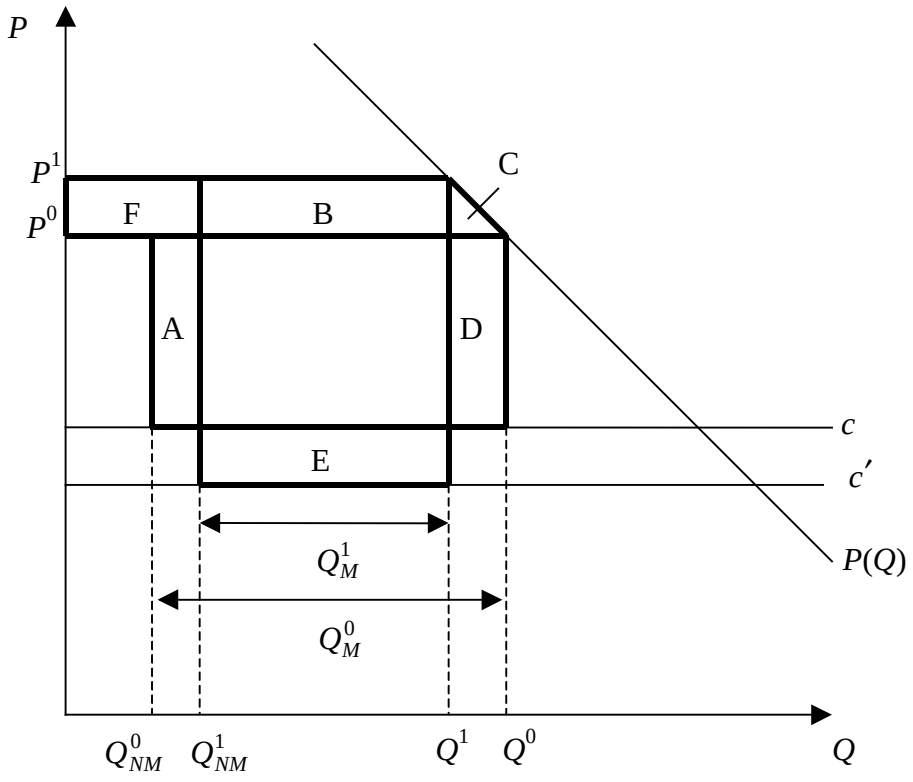
6.3 Effect of delegation on the welfare effects of mergers

The mergers considered in the previous section are all welfare reducing as they involve a reduction in output and no efficiency gains. Now consider the possibility of efficiency gains and the extent to which delegation increases the likelihood that profitable mergers increase welfare. Consider the following example illustrated in figure 6.1. Suppose that a merger results in a reduction in industry output from Q^0 to Q^1, a reduction in the output of the merging firms from Q_M^0 to Q_M^1 and an expansion of non-merging firm output from Q_{NM}^0 to Q_{NM}^1. Furthermore, suppose that the non-merging firms face a marginal cost of c before and after the merger and that

the merging firms experience a reduction in marginal cost from c to c' as a result of the merger. Such a merger improves welfare provided the cost efficiency gain (area E) exceeds the loss in allocative efficiency (areas C + D).

Now suppose that the antitrust authorities cannot observe area E. Given that all proposed mergers are presumably profitable then Farrell and Shapiro (1990) point out that a sufficient condition for a proposed merger to be welfare enhancing is that the external effects of the merger be positive. The external effects of a merger are given by the increase in non-merging firm profits (areas A + F) net of the reduction in consumer surplus (areas F + B + C). The external effects of a merger are thus positive provided area A exceeds area B + C. The height of area A depends on the mark-up over marginal cost, which is directly related, given profit maximising behaviour, to the output produced by the non-merging firms. The width of area A depends on the extent to which the non-merging firms respond to the output contraction by the merging firms. The size of area A is thus given by a weighted sum of the outputs produced by the non-merging firms, where the weights are determined by the responsiveness of non-merging firms to the output contraction of

Figure 6.1 **The welfare effects of merger**



the merging firms. The size of area B + C is determined mainly by the width of area B, which is given by the output of the merging firms. As a result, area A will exceed area B + C (ie the external effects of merger are positive) provided the weighted sum of the non-merging firm outputs (or market shares) is greater than the output (or market share) of the merging firms.[24]

Now use figure 6.1 and the results derived in the previous section to argue that delegation reduces the external benefits of merger. Under a delegation regime aggressive incentive competition reduces the mark-up and thereby reduces the effect of any output expansion by the non-merging firms (ie reduces the height of area A). Second, the incentive softening effects of mergers under delegation regimes reduce the responsiveness of the non-merging firms to the output contraction of the merging firms. This reduces the width of area A and results in higher prices which then increase the size of area B + C. This suggests that delegation reduces the external benefits of merger and thereby reduces the set of profitable mergers than are welfare enhancing. In particular, the market share of the merging parties that guarantees that a profitable merger is welfare enhancing (ie the 'safe harbour' market share) is lower under a delegation regime than under a non-delegation regime. Evidence that the 'safe harbour' market share is lower under a delegation regime can be obtained by comparing the results in Levin (1990)[25] with those contained in table 6.1. Levin (1990) shows that 50 per cent is a 'safe harbour' market share under non-delegation Cournot regimes, provided marginal cost is constant and demand is non-convex (ie $\beta \geq 0$). Table 6.1 documents two cases in which this 50 per cent rule is violated under a delegation regime. In particular, if demand is linear (ie $\beta = 0$) and there are either 4 or 6 firms in the pre-merger setting then a merger of 50 per cent of the firms is profitable and welfare *reducing*.

6.4 Conclusion

An old intuition in industrial organisation is that merger facilitates collusion because it reduces the number of competing units and thus makes it easier for firms to agree

[24] Given that the antitrust authorities cannot observe the cost efficiency gain then this rule is only useful if it is based on the pre-merger market shares. Farrell and Shapiro (1990) show that this is indeed the case provided the demand and cost functions satisfy certain conditions.

[25] The non-delegation welfare analysis contained in Levin (1990) and Farrell and Shapiro (1990) is made tractable by the fact that under Cournot regimes the response of the non-merging firms depends only on the output contraction of the merging firms and not on the number of merging firms. Under delegation regimes, the non-merging firm response depends both on the output contraction and the number of merging firms. This makes the delegation regime analysis unwieldy and precludes a more general comparison of the welfare analysis under the two regimes.

to the collusive outcome. A repeated game version of this argument is that merger gives each firm a larger share of the monopoly profits, which then results in each firm having more to lose by not acting collusively.[26] The delegation regime version of this argument presented here is that merger reduces the collective responsiveness of any firm's rivals to the rent shifting efforts of that firm, which then reduces the benefits of aggressive incentive competition. A consequence of incentive softening is that the merging firms capture more of the surplus from their output contraction. This implies that mergers are more profitable and that the positive external effects of a merger are reduced. The former of these two effects reduces the minimum market share that the merging parties must possess in order to profitably merge without efficiency gains, whereas the latter reduces the maximum market share that the merging parties can possess in order to guarantee that a profitable merge is welfare enhancing.

An important assumption of this model is that the products are assumed to be homogeneous. In the homogeneous product case, the merged entity always shuts down all but one of the products involved in the merger. If products were differentiated this may no longer prove to be profitable. Instead, a more prudent strategy for the merged entity might to be to hire a common agent to sell all of the products involved in the merger. An important distinction between this common agency regime and the one discussed in the common agency literature is that the terms of the common agent's contract are decided by a single firm selling multiple products, rather than by many competing firms selling one product each. Furthermore, the common agency literature considers a setting in which the common agent is a monopolist, as opposed to a multi-product common agent competing in an oligopoly setting. An interesting question for future research is whether a common agent is less responsive to rival rent shifting than are independent agents, and thus whether the incentive softening results derived in this paper for the homogeneous product case would continue to hold for differentiated products case.

Appendix 6A Derivation of the output stage comparative statics

Implicitly differentiating the set of first order conditions given in equation [6.5] with respect to w_i yields the following set of n equations:

$$\left(P'(Q) + q_i P''(Q)\right)\frac{\partial Q}{\partial w_i} + P'(Q)\frac{\partial q_i}{\partial w_i} - 1 = 0 \qquad [A.1]$$

[26] See chapter 6 in Tirole (1988).

$$\left(P'(Q) + q_j P''(Q)\right)\frac{\partial Q}{\partial w_i} + P'(Q)\frac{\partial q_j}{\partial w_i} = 0 \quad j = 1,\ldots,n \text{ where } j \neq i \qquad [A.2]$$

Now sum the above n equations and then substitute

$$Q = q_i + \sum_{j \neq i} q_j \text{ and } \frac{\partial Q}{\partial w_i} = \frac{\partial q_i}{\partial w_i} + \sum_{j \neq i}\frac{\partial q_j}{\partial w_i}$$

to obtain equation [6.6]. Substituting equation [6.6] into equation [A.1] and using $q_i = Q/n$ and $\beta(Q) = P''(Q)Q/P'(Q)$ yields equation [6.7].

Appendix 6B Derivation of $F_D(N, k, \beta)$

Substituting $\beta(Q) = \beta$ and equation [6.9] into equation [6.8] yields

$$P(Q) + P'(Q)\frac{Q}{n^2 + (n-1)\beta} - c = 0 \qquad [B.1]$$

Implicitly differentiate with respect to n[27] to obtain

$$P'Q' + \frac{Q'\left(P''Q + P'\right)}{n^2 + (n-1)\beta} - \frac{P'Q(2n + \beta)}{\left(n^2 + (n-1)\beta\right)^2} = 0$$

Re-arrange and use $\beta(Q) = P''(Q)Q/P'(Q)$ to obtain that the proportional increase in output due to a marginal increase in the number of firms is given by

$$\frac{Q'}{Q} = \frac{2n + \beta}{(n^2 + (n-1)\beta)(n^2 + n\beta + 1)} \qquad [B.2]$$

Each firm's equilibrium profits under a delegation regime are given by

$$\pi_D(n) = \frac{\left(P(Q(n) - c\right)Q(n)}{n}$$

Differentiating with respect to n yields

$$\pi'_D(n) = \frac{Q'(P - c + P'Q)}{n} - \frac{(P - c)Q}{n^2}$$

[27] For convenience n is being treated as a continuous variable. Since all functions are continuous then the integer solution can be obtained directly from the continuous solution.

Re-arrange to obtain that the proportional increase in profits due to a marginal increase in the number of firms is given by

$$\frac{\pi_D'(n)}{\pi_D(n)} = \frac{Q'}{Q}\left(\frac{P'Q}{P-c}+1\right) - \frac{1}{n}$$

Now substitute equations [B.1] and [B.2] to obtain

$$-\frac{\pi_D'(n)}{\pi_D(n)} = f_D(n,\beta) \qquad \text{[B.3]}$$

where

$$f_D(n,\beta) = \frac{(2n+\beta)(n^2+n\beta-(1+\beta))}{(n^2+n\beta-\beta)(n^2+n\beta+1)} + \frac{1}{n} \qquad \text{[B.4]}$$

Integrate both sides of equation [B.3] to obtain

$$\int_{N-k}^{N} -\frac{\pi_D'(n)}{\pi_D(n)}dn = \ln\left[\frac{\pi_D(N-k)}{\pi_D(N)}\right] = \int_{N-k}^{N} f_D(n,\beta)dn$$

The last two equalities imply

$$\frac{\pi_D(N-k)}{\pi_D(N)} = \exp\int_{N-k}^{N} f_D(n,\beta)dn = F_D(N,k,\beta) \qquad \text{[B.5]}$$

where $F_D(N,k,\beta)$ is obtained by integrating equation [B.4] and then taking the exponential. This procedure yields equation [6.13] if $\beta \neq -1$. If $\beta = -1$ then

$$F_D(N,k,-1) = \frac{N}{N-k}\left[\frac{N^2-N+1}{(N-k)^2-(N-k)+1}\right]\exp\left[\frac{1}{N^2-N+1} - \frac{1}{(N-k)^2-(N-k)+1}\right]$$

Appendix 6C Lemmas 6.1 and 6.2

Lemma 6.1

If $k_D(N,\beta) \in (0, N-1)$ is the lowest k to solve $F_D(N,k,\beta) = k+1$ then merger is profitable (resp. unprofitable) provided $k >$ (resp. $<$) $k_D(N,\beta)$, where $k \in (0, N-1]$ is an integer. If there does not exist a $k \in (0, N-1)$ that solves $F_D(N,k,\beta) = k+1$ then all mergers are profitable.

Proof

Let $g(k) = F_D(N, k, \beta)$. Then it can be shown that

$$g(0) = 1, \qquad \text{[C.1]}$$

$$g'(k) > 0 \text{ for all } k \in [0, N-1] \text{ and} \qquad \text{[C.2]}$$

$$g''(k) > 0 \text{ for all } k \in [0, N-2] \qquad \text{[C.3]}$$

Case 1

If there exists a $k_D(N, \beta) \in (0, N-1)$ such that $g(k) = k + 1$ when $k = k_D(N, \beta)$ then equations [C.1] to [C.3] imply that

$$\text{if } k_D(N, \beta) \in (0, N-2] \text{ then } g(k) \begin{matrix} < \\ = \\ > \end{matrix} k + 1 \text{ if } k \begin{matrix} < \\ = \\ > \end{matrix} k_D(N, \beta) \text{ , where } k \in (0, N-2] \qquad \text{[C.4]}$$

$$\text{and if } k_D(N, \beta) \in (N-2, N-1) \text{ then } g(k) < k+1 \text{ for all } k \in (0, N-2] \qquad \text{[C.5]}$$

Given that merger for monopoly is always profitable then

$$g(k) > k+1 \text{ if } k = N-1 \qquad \text{[C.6]}$$

Equations [C.4] to [C.6] imply that if $k \in (0, N-1]$ is an integer then $g(k) \begin{matrix} < \\ = \\ > \end{matrix} k + 1$ for $k \begin{matrix} < \\ = \\ > \end{matrix} k_D(N, \beta)$, where $k_D(N, \beta)$ is the lowest k that solves $g(k) = k + 1$.

Case 2

If there does not exist a $k \in (0, N-1)$ that solves $g(k) > k+1$ then equation [C.6] and the continuity of $g(k)$ imply that $g(k) > k+1$ for all $k \in (0, N-1]$.

Lemma 6.2

There exists a $k_{ND}(N, \beta) \in (0, N-1)$ such that $F_{ND}(N, k, \beta) = k+1$ when $k = k_D(N, \beta)$. The merger is profitable (resp. unprofitable) provided $k >$ (resp. $<$) $k_{ND}(N, \beta)$ where $k \in (0, N-1]$.

For a proof see Faulí-Oller (1997).

Appendix 6D Proof of proposition 6.1

(i) Faulí-Oller (1997) shows that the non-delegation version of $f_r(n,\beta)$ is

$$f_{ND}(n,\beta) = \frac{n-1}{n(n+1+\beta)} + \frac{1}{n} \qquad [D.1]$$

where

$$\frac{\pi_{ND}(N-k)}{\pi_{ND}(N)} = \exp \int_{N-k}^{N} f_{ND}(n,\beta)dn = F_{ND}(N,k,\beta) \qquad [D.2]$$

and where $F_{ND}(N,k,\beta)$ is given by equation [6.14] and is obtained by integrating equation [D.1] and taking the exponential. A comparison of equations [D.1] and [B.4] reveals that $f_D(n,\beta) > f_{ND}(n,\beta)$ for all $n \geq 2$ and for all $\beta > -2$. Given equations [B.5] and [D.2] this implies

$$F_D(N,k,\beta) > F_{ND}(N,k,\beta) \text{ for all } N \geq 2,\ k \in (0, N-2] \text{ and } \beta > -2 \qquad [D.3]$$

Case 1

If $k_D(N,\beta)$ is the lowest $k \in (0, N-1]$ to solve $F_D(N,k,\beta) = k+1$ then from the proofs of lemmas 6.1 and 6.2 it follows that if $k \in (0, N-1]$ is an integer then

$$\begin{matrix} < & < \\ F_r(N,k,\beta) = k+1 \text{ for } k = k_r(N,\beta) \text{ for } r \in \{D, ND\} \\ > & > \end{matrix} \qquad [D.4]$$

Equations [D.3] and [D.4] imply that if $k_{ND}(N,\beta) \leq N-2$ then $k_D(N,\beta) < k_{ND}(N,\beta)$, which implies that delegation reduces the number of firms required for profitable merger. If $k_{ND}(N,\beta) > N-2$ then *only* merger for monopoly is profitable under a non-delegation regime. Since merger for monopoly is always profitable under a delegation regime then delegation cannot reduce the set of profitable mergers.

Case 2

If there does not exist a k that solves $F_D(N,k,\beta)=k+1$ then any merger under a delegation regime is profitable (lemma 6.1). As a result, delegation cannot reduce the set of profitable mergers.

(ii) From equation [B.4] it follows that $\partial f_D(n,\beta)/\partial\beta<0$ for all $n\geq 2$ and for all $\beta>-2$. Given equation [B.5] this implies

$$\frac{\partial F_D(N,k,\beta)}{\partial\beta}<0 \text{ for all } N\geq 2,\ k\in(0,N-2] \text{ and } \beta>-2 \qquad [D.5]$$

Equation [D.5] and lemma 6.1 imply that: if $k_D(N,\beta)\leq N-2$ then $\partial k_D/\partial\beta>0$ (ie increases in β increase the number of firms required for profitable merger), and if $k_D(N,\beta)>N-2$ then only merger for monopoly is profitable before and after the rise in β (ie β has no effect on the number of firms required for profitable merger).

(iii) From equations [D.1] and [D.2] it follows that

$$\begin{aligned}\lim_{\beta\to\infty} F_{ND}(N,k,\beta) &= \exp\int_{N-k}^{N}\lim_{\beta\to\infty} f_{ND}(n,\beta)dn\\ &= \exp\int_{N-k}^{N}\frac{1}{n}dn\\ &= \frac{N}{N-k}\end{aligned}$$

As a result, the solution to $\lim_{\beta\to\infty}F_{ND}(N,k,\beta)=k+1$ yields $\frac{k+1}{N}=1$.

(iv) Following the procedure in (iii) yields that $\lim_{\beta\to\infty}F_D(N,k,\beta)=\left(\frac{N}{N-k}\right)^2$ which implies that the solution to $\lim_{\beta\to\infty}F_D(N,k,\beta)=k+1$ yields $\frac{k+1}{N}=1-\frac{\sqrt{1+4N}-1}{2N}$.

References

Baye, M.R., Crocker, K.J. and Ju, J. 1996, 'Divisionalisation, franchising and divesture incentives in oligopoly', *American Economic Review*, vol. 86, no. 1, pp. 223–36.

Bernheim, B.D. and Whinston, M.D. 1985, 'Common marketing agency as a device for facilitating collusion', *Rand Journal of Economics*, vol. 16, no, 2, pp. 313–28.

Bonanno, G. and Vickers, J. 1988, 'Vertical separation', *Journal of Industrial Economics*, vol. 36, no. 3, pp. 257–65.

Cheung, F.K. 1992, 'Two remarks on the equilibrium analysis of horizontal merger', *Economics Letters*, vol. 40, no. 1, pp. 119–23.

d'Aspremont, C., Jacquemin, A., Gabszewicz, J.J. and Weymark, J.A. 1983, 'On the stability of collusive price leadership', *Canadian Journal of Economics*, vol. 16, no. 1, pp. 17–25.

Daughety, A.F. 1990, 'Beneficial concentration', *American Economic Review*, vol. 80, no. 5, pp. 1231–7.

Deneckere, R. and Davidson, C. 1983, *Coalition formation in non-cooperative oligopoly models*, Working Paper no. 8203, Michigan State University, East Lansing, Michigan.

—— 1985, 'Incentives to form coalitions with Bertrand competition', *Rand Journal of Economics*, vol. 16, no. 4, pp. 473–86.

Farrell, J. and Shapiro, C. 1990, 'Horizontal mergers: An equilibrium analysis', *American Economic Review*, vol. 88, no. 1, pp. 107–26.

Faulí-Oller, R. 1997, 'On merger profitability in a Cournot setting', *Economics Letters*, vol. 54, no. 1, pp. 75–9.

Fershtman, C. 1985, 'Internal organizations and managerial incentives as strategic variables in competitive environments', *International Journal of Industrial Organization*, vol. 3, no. 2, pp. 245–53.

Fershtman, C. and Judd, K.L. 1987, 'Equilibrium incentives in oligopoly', *American Economic Review*, vol. 77, no. 5, pp. 927–40.

Gaudet, G. and Salant, S.W. 1991, 'Increasing the profits of a subset of firms in oligopoly models with strategic substitutes', *American Economic Review*, vol. 81, no. 3, pp. 658–65.

Horn, H. and Wolinsky, A. 1988, 'Bilateral monopolies and incentives for merger', *Rand Journal of Economics*, vol. 19, no. 3, pp. 408–19.

Kamien, M.I. and Zang, I. 1990, 'The limits of monopolization through acquisition', *Quarterly Journal of Economics*, vol. 105, no. 2, pp. 465–99.

Levin, D. 1990, 'Horizontal merger: the 50-per cent benchmark', *American Economic Review*, vol. 80, no. 5, pp. 1238–45.

Lin, Y.J. 1988, 'Oligopoly and vertical integration: Note', *American Economic Review*, vol. 78, no. 1, pp. 251–4.

—— 1990, 'The dampening-of-competition effect of exclusive dealing', *Journal of Industrial Economics*, vol. 39, no. 2, pp. 209–23.

McAfee, R.P. and Williams, M.A. 1992, 'Horizontal mergers and antitrust policy', *Journal of Industrial Economics*, vol. 40, no. 2, pp. 181–7.

O'Brien, D.P. and Shaffer, G. 1993, 'On the dampening-of-competition effect of exclusive dealing', *Journal of Industrial Economics*, vol. 41, no. 2, pp. 215–21.

Perry, M. and Porter, R. 1985, 'Oligopoly and the incentive for horizontal merger', *American Economic Review*, vol. 75, no. 1, pp. 219–27.

Salant, S.W., Switzer, S. and Reynolds, R.J. 1983, 'Losses from horizontal merger: The effects of an exogenous change in industry structure on Cournot-Nash equilibrium', *Quarterly Journal of Economics*, vol. 98, no. 2, pp. 185–99.

Scherer, F.M. and Ross, D. 1990, *Industrial Market Structure and Economic Performance*, 3rd edn, Houghton Mifflin, Boston, Massachusetts.

Sklivas, S. 1987, 'The strategic choice of managerial incentives', *Rand Journal of Economics*, vol. 18, no. 3, pp. 452–8.

Tirole, J. 1988, *The Theory of Industrial Organization*, MIT Press, Cambridge Massachusetts.

Vickers, J. 1985, 'Delegation and the theory of the firm', *Economic Journal*, vol. 95 (supplement), pp. 138–47.

Ziss, S. 1995, 'Vertical separation and horizontal merger', *Journal of Industrial Economics*, vol. 43, no. 1, pp. 63–75.

7 Carrot and stick games

*J. Jude Kline**

In a carrot game for a player, that player must help their opponent to get a desired response. In a stick game for a player, that player must hurt their opponent to get a desired response. It is shown that most all smooth monotonic two player games can be classified as: carrot for both, carrot for one and stick for the other, or stick for both. Carrot and stick games are then transformed into sequential move games and incentive games. A carrot game for a player implies mutual benefits will come from their moving first. A stick game for a player implies that if they move first, the other player will suffer. Incentive games based on a carrot game for each player bring mutual gains. Those based on a stick game for each player bring harm to at least one and possibly to both.

7.1 Introduction

In the industrial organisation literature there is often reference to the difference between strategic substitutes and strategic complements (Bulow, Geanakopolos and Klemperer 1985, Shy 1995). A game is said to involve strategic substitutes if the best response functions are downward sloping. Alternatively, a game is said to involve strategic complements if the best response functions are upward sloping. Quantity competition games typically involve strategic substitutes and price competition games typically involve strategic complements.

One well-known result is that under quantity competition, sequential movement benefits the leader at the expense of the follower (Dowrick 1985, Gal-Or 1985). For price competition, sequential movement benefits both firms and typically the follower benefits more. While one can describe this difference in terms of strategic complements and strategic substitutes, this description has virtually no economic content. Furthermore, to come to the conclusion that strategic complements imply sequential movement will always be mutually beneficial would be misleading. This

* The author would like to thank Lisa Hillsdon, Mamoru Kaneko, Fahad Khalil and Ko Nishihara for discussions, comments and suggestions.

paper gives an example of an advertising game with strategic complements where sequential movement never benefits both firms (example 7.3).

This paper introduces the notions of carrots and sticks to describe differences between games, such as quantity and price competition games. Loosely speaking, a carrot game is one in which a player must help out their opponent in order to get a desired response. A stick game, on the other hand, is one in which a player must harm their opponent in order to get a desired response. These definitions are inspired by the age-old question of whether to use a carrot or a stick to motivate a beast of burden.

Once it is recognised that price competition is a carrot game and quantity competition is a stick game, it is a trivial result that sequential movement will be mutually beneficial under price competition, while only unilaterally beneficial under quantity competition (proposition 7.1). In a carrot game, the leader offers a carrot to their opponent to get a desired response and both benefit. Under price competition, the carrot comes in the form of a higher price that translates into reduced competition. In a stick game, the leader will whack their opponent with a stick to get a desired response. The stick comes in the form of a greater output level by the leader in a quantity setting game. This drives the market price and profit of the follower down.

Price and quantity competition games are typically modelled as smooth monotonic games. Section 7.3 defines smooth monotonic games formally and shows that most all of them can be classified in terms of carrots and sticks. The section also gives conditions under which games with strategic complements are carrot games, and games with strategic substitutes are stick games. In those cases, the distinction between strategic complements and strategic substitutes has a clear economic meaning that derives from the distinction between carrot and stick games.

In the analysis of carrot and stick games the players are allowed to be asymmetric. In fact, a game may be carrot for one player and stick for the other. Such an example is a duopoly game where one firm chooses a price and the other chooses a quantity (example 7.2). It turns out that this game is a carrot game for the quantity setter and a stick game for the price setter. Whether or not sequential movement is mutually beneficial here depends on who moves first. The game will be mutually beneficial only in the case that the quantity setter moves first.

In relation to the literature on sequential versus simultaneous movement, Hamilton and Slutsky (1990) endogenised the choice between simultaneous and sequential movement. This paper does not explicitly give players the choice between simultaneous or sequential movement, but clearly the results have implications for that choice. The results are given conditional on whether a game has a carrot or a

stick nature, while Hamilton and Slutsky's results are given conditional on whether sequential movement is mutually or unilaterally beneficial. As is shown in proposition 7.1, the carrot or stick nature of a game determines whether or not sequential movement will be mutually beneficial.

Furthermore, as is seen in section 7.4, carrot and stick games have implications about other types of transformations of games, not just from simultaneous moves to sequential moves.

For example, many results in the industrial organisation literature about manipulating payoffs in games are easily understood as coming from the carrot or stick nature of the original game. Examples include owners of firms who pay managers to act more or less aggressively than profit maximisers (Fershtman and Judd 1987, Sklivas 1987, Vickers 1985, Basu 1995 and Basu, Ghosh and Ray 1997). Additionally, Bonanno and Vickers (1988), and Rey and Stiglitz (1995) show how vertical separation may provide firms with new commitment opportunities that yield higher equilibrium profits. Baye, Crocker and Ju (1996) show similar results for horizontal separation.

The main results of the papers cited in the previous paragraph can be predicted immediately from noting the carrot or stick nature of the original games. Take the example of the managerial incentives game studied by Sklivas (1987). The original game is a carrot game for each firm when the firms compete in prices and a stick game for each firm when they compete in quantities.

It is shown in a general setting that when players can manipulate incentives of their agents in an incentive game, they will choose to magnify the carrot or stick behaviour of their agents (proposition 7.3). Thus, if the original game is a stick game for one firm (player), then the manager (agent) of that firm will be given incentive to act more aggressively (more stick like) in the incentive game. Alternatively, if the original game is a carrot game for a firm, then the manager of that firm will be given incentive to act less aggressively (more carrot like) in the incentive game.

If both firms compete in prices, it follows from the carrot nature of the original game that an incentive game will lead to even more carrot like behaviour and greater firm profits (proposition 7.4). Alternatively, if the firms compete in quantities, the incentive game will lead to even more stick like behaviour and reduce the profits of at least one firm (proposition 7.5). Note the distinction between the result for carrot games and the result for stick games. If the original game is a carrot game for each player, then the incentive game will raise the welfare of *each*. However, if the original game is a stick game for each player, then the welfare of *one, but possibly not both* will be reduced.

For example, if the managers compete in quantities, but one has a cost advantage over the other, then the firm with the cost advantage can actually do better in the incentive game (example 7.5). The result that both firms are made worse off from the incentive game based on a stick game for each player is obtained if the players are symmetric (corollary 7.3).

The layout of the paper is as follows. Section 7.2 defines carrot and stick games and discusses the implications of these games for sequential movements. Section 7.3 concentrates on smooth monotonic games and relates the results for those games to strategic complements and strategic substitutes. Section 7.4 gives some welfare results for incentive games based on smooth monotonic games. Section 7.5 shows the relationship between carrot and stick games and Stackelberg solvable games. Conclusions are given in section 7.6. All proofs are in the appendix.

7.2 Carrots and sticks

Consider a two person normal form game $G = ((S_1, S_2), (u_1, u_2))$, where the subscripts 1 and 2 denote the players. The strategy set of player *i* for *i* = 1,2 is S_i and refers to a specific strategy of player *i* as s_i. The objective of each player is to maximise their payoff $u_i : S_1 \times S_2 \to R$, where *R* is the real line. Let $b_1(s_2) = \{s_1 \mid s_1 \in S_1$ and $\forall$ $s_1' \in S_1$, $u_1(s_1, s_2) \geq u_1(s_1', s_2)\}$. Call $b_1(s_2)$ the best response set of player 1 given s_1. Assume that for each s_2 the best response set is a singleton and thus $b_1(s_2)$ can denote the best response function of player 1 and $b_2(s_1)$ to denote the best response function of player 2. A *Nash equilibrium* in a game *G* is a pair of strategies s_1^* and s_2^* that simultaneously satisfy: (i) $u_1(s_1^*, s_2^*) \geq u_1(s_1, s_2^*)$ for all $s_1 \in S_1$, and (ii) $u_2(s_1^*, s_2^*) \geq u_2(s_1^*, s_2)$ for all $s_2 \in S_2$.

Define $\underline{v}_1$ to be the minimum Nash equilibrium payoff of player 1 in *G* and $\bar{v}_1$, to be the maximum Nash equilibrium payoff of player 1 in *G*. Define $\underline{v}_2$ and $\bar{v}_2$ in the same way for player 2 and assume that these values exist for both players. The game of figure 7.1 has a unique equilibrium (T, L), so $\underline{v}_1 = \bar{v}_1 = 2$, and $\underline{v}_2 = \bar{v}_2 = 3$. In general, each of $\underline{v}_1$, $\bar{v}_1$, $\underline{v}_2$ and $\bar{v}_2$ may be determined by a different Nash equilibrium.

Now define carrot and stick games for player 1. Carrot and stick games for player 2 are defined in an analogous way by exchanging the roles of the players. The remainder of the paper often gives results only for player 1.

Figure 7.1 **Game 1**

		Player 2	
		L	R
Player 1	T	2,3	6,1
	B	1,–3	3,2

Figure 7.2 **Game 2**

		Player 2	
		L	R
Player 1	T	1,1	3,–1
	B	0,0	2,2

(2.1) **Carrot game:** A *carrot game* for player 1 is one in which $u_1(s_1, b_2(s_1)) > \bar{v}_1$ implies $u_2(s_1, b_2(s_1)) > \bar{v}_2$, and there exists at least one $s_1 \in S_1$ for which $u_1(s_1, b_2(s_1)) > \bar{v}_1$.

(2.2) **Stick game:** A *stick game* for player 1 is one in which $u_1(s_1, b_2(s_1)) > \bar{v}_1$ implies $u_2(s_1, b_2(s_1)) < \underline{v}_2$, and there exists at least one $s_1 \in S_1$ for which $u_1(s_1, b_2(s_1)) > \bar{v}_1$.

Loosely speaking a carrot game is one in which for player 1 must help player 2 in order to get a desired response. A stick game is one in which player 1 must hurt player 2 in order to get a desired response.

While not every two player game can be classified as a stick or carrot game for each player, many games of interest to economists and game theorists can. For example, section 7.3 shows that a heterogeneous price setting game with linear demand and constant marginal cost is an example of a carrot game for each player, while the quantity setting version is a stick game for each player (example 7.1).

Look first at some simple games. The game of figure 7.1 turns out to be a stick game for player 1, but neither carrot nor stick for player 2. To check this, recall that the unique Nash equilibrium (T, L) determines $\underline{v}_1 = \bar{v}_1 = 2$, and $\underline{v}_2 = \bar{v}_2 = 3$. Clearly, since $b_2(T) = L$ and $b_2(B) = R$, the strategy $s_1 = B$ is the only strategy of player 1 that satisfies $u_1(s_1, b_2(s_1)) > \bar{v}_1$. Since $u_2(B, b_2(B)) = 2 < \underline{v}_2$, it follows that this is a stick game for player 1. For player 2, the game is neither stick nor carrot since there does not exist a strategy $s_2 \in \{L, R\} = S_2$ for which $u_2(b_1(s_2), s_2) > \bar{v}_2$.

By a similar type of analysis, the game of figure 7.2 can be shown to be a carrot game for player 1 and neither carrot nor stick for player 2.

What will happen if player 1 is allowed to move first? In the game of figure 7.1, they will choose *B* rather than *T*, inducing the desired response of *R* by player 2. Player 1's payoff will rise from 2 in the simultaneous move game to 3 in the sequential move game. Conversely, player 2's payoff falls from 3 to 2. Here the benefit to player 1 comes at a loss to player 2.

In the game of figure 7.2, if player 1 can move first they will again choose *B*, inducing a desired response of *R*. Here both players benefit from this change since the payoff to each increases from 1 to 2.

In general, the gain of one player from sequential movement may or may not be at the expense of another player. Note that the player moving first must do no worse from this change, since they can simply choose the strategy that yields their value in the original game *G*.

Given a game $G = ((S_1, S_2), (u_1, u_2))$, a sequential move game $G' = ((S_1, F_2), (u_1, u_2))$ is obtained by allowing player 1 to move first and player 2 to observe 1's strategy choice before making their choice. In the game G', a strategy of player 2, denoted by f_2, describes what player 2 will do for each strategy s_1 of player 1. The strategy set of player 2 is thus a set of functions $F_2 = \{f_2 \mid f_2 : S_1 \to S_2\}$. This captures the idea that player 2 can observe and react to player 1's choice. The appropriate equilibrium concept in this setting is a subgame perfect equilibrium.

Since the best response sets are assumed to be unique valued, a strategy profile (s_1^*, f_2^*) is a *subgame perfect equilibrium* in G', if and only if $f_2^*(s_1) = b_2(s_1)$ and $u_1(s_1^*, b_2(s_1^*)) \geq u_1(s_1, b_2(s_1))$ for all $s_1 \in S_1$. Another sequential move game can be obtained by allowing player 2 to move first.

Carrot games, on the one hand, are games where both players gain from sequential movement, while stick games involve one gaining at the expense of the other. This idea is expressed in the next proposition.

Proposition 7.1

Let *G* be a two person game and let G' be the sequential move game where player 1 moves first.

(2.3) If *G* is a carrot game for player 1, then for any subgame perfect equilibrium (s_1^*, f_2^*) of G', $u_1(s_1^*, f_2(s_1^*)) > \bar{v}_1$ and $u_2(s_1^*, f_2(s_1^*)) > \bar{v}_2$.

(2.4) If *G* is a stick game for player 1, then for any subgame perfect equilibrium (s_1^*, f_2^*) of G', $u_1(s_1^*, f_2(s_1^*)) > \bar{v}_1$, and $u_2(s_1^*, f_2(s_1^*)) < \underline{v}_2$.

If can be established that a game is a carrot or stick game for a player, proposition 7.1 shows immediately whether or not sequential movement with that player moving first will be unilaterally or mutually beneficial. Notice that the result states that sequential movement in a carrot game will improve each player's best Nash equilibrium payoff in *G*. In a stick game, sequential movement makes the player moving first better off than their *best* Nash equilibrium payoff in *G*, and the player moving second worse off than their *worst* Nash equilibrium payoff in *G*.

Notice that in each of the simple 2×2 examples given above, the game was neither a carrot nor a stick game for player 2. For 2×2 games, it turns out that they can be classified in terms of carrots and sticks for at most one player.

Lemma 7.1

Let *G* be a 2×2 game. If $\bar{v}_1$ and $\bar{v}_2$ are defined by the same Nash equilibrium, then *G* is a carrot or stick game for at most one player. Furthermore, if $\bar{v}_1$ and $\bar{v}_2$ are defined by different Nash equilibrium, then *G* is neither a carrot nor a stick game for either player.

In view of this result, it appears there may be a problem classifying games with finite strategy spaces as carrot or stick games for both players. This problem is really only one of 2×2 games. The next example, which is based on the game of figure 7.1, is meant to dispel the belief that such a result holds more generally.

The game of figure 7.3 is a stick game for player 1 for all $\alpha > 0$. However, for player 2 it is a stick game if $0 < \alpha < 2$ and a carrot game if $\alpha > 2$.

In the case when $\alpha > 2$, since the game is a carrot game for player 2 and a stick game for player 1, proposition 7.1 says that sequential movement will only be mutually beneficial if player 2 moves first. Other asymmetric examples are discussed in the next section.

Figure 7.3 **Game 3**

		Player 2		
		L	*R*	*N*
Player 1	*T*	2,3	6,1	0,0
	B	1,–3	3,2	0,0
	N	0,5	0,0	α,4

7.3 Smooth monotonic games

Now turn to smooth monotonic games — games commonly used throughout the industrial organisation and economics literature. It is shown that all smooth monotonic games where each player has something to gain from moving first, for example, those with only interior equilibrium, can be classified as either a carrot or a stick game for each player.

Let the strategy space of each player be some convex subset of the real line *R*. Let the payoff functions $u_1(s_1, s_2)$ and $u_2(s_1, s_2)$ and the best response functions $b_1(s_2)$

and $b_2(s_1)$ be differentiable. Use $b_1'(s_2)$, and $b_2'(s_1)$ to represent the first derivatives of the best response functions.

Smooth monotonic game

A two player game *G* is a smooth monotonic game, if and only if the following are true at all (s_1, s_2):

(3.1) **Smoothness**: the payoff functions $u_1(s_1, s_2)$ and $u_2(s_1, s_2)$ and the best response functions $b_1(s_2)$ and $b_s(s_1)$ are differentiable.

(3.2) **Monotonic externalities**: (i) either $\frac{\partial u_1(s_1, s_2)}{\partial s_2} < 0$ everywhere or $\frac{\partial u_1(s_1, s_2)}{\partial s_2} > 0$ everywhere; and (ii) either $\frac{\partial u_2(s_1, s_2)}{\partial s_1} < 0$ everywhere or $\frac{\partial u_2(s_1, s_2)}{\partial s_1} > 0$ everywhere.

(3.3) **Monotonic best responses**: (i) either $b_1'(s_2) < 0$ everywhere or $b_1'(s_2) > 0$ everywhere; and (ii) either $b_2'(s_1) < 0$ everywhere or $b_2'(s_1) > 0$ everywhere.

When it happens that $\frac{\partial u_1(s_1, s_2)}{\partial s_2} < 0$, player 2 imposes a negative externality on player 1. When the sign is positive, player 2 imposes a positive externality on player 1. The following is an example of a smooth monotonic game.

Example 7.1: Price or quantity competition

Two firms sell heterogeneous substitute products, face linear demand curves and constant zero marginal cost. For quantity competition, the inverse demand system is given by $p_i(q_i, q_j) = 1 - q_i - \frac{q_j}{2}$ for *i*=1,2 and $j \neq i$. For price competition the demand system is given by $q_i(p_i, p_j) = \frac{2}{3} - \frac{4p_i}{3} - \frac{2p_j}{3}$, for *i*= 1,2 and $j \neq i$.[1] The strategy space in either case is the strictly positive real line R^{++}.

It is straightforward to check that the games are smooth and monotonic. In the case of quantity setting, the externalities are negative since raising one player's quantity

[1] These come from the more general linear demand system given by Klemperer and Meyer (1986): $p_i = \alpha - \beta q_i - \gamma q_j$, for *i*=1,2, $j \neq i$, where $\alpha = \beta = 1$, and $\gamma = 1/2$.

lowers the market price faced by the other firm. In the price setting game, the externalities are positive since raising a player's price is a less competitive act. The best response functions, which are given in figures 7.4 and 7.5, are also clearly monotonic.

Figure 7.4 **Quantity competition**

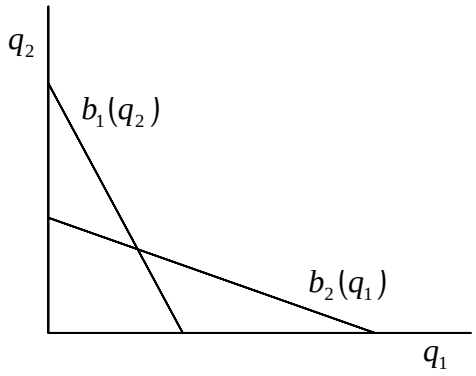
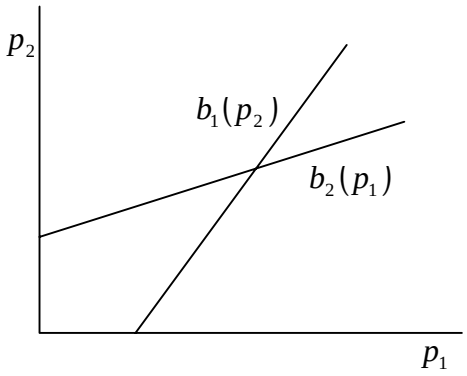


Figure 7.5 **Price competition**



In the price and quantity games of example 7.1 the players are very symmetric. In general, the players may be quite asymmetric. Two situations involving asymmetric players are given in examples 7.2 and 7.3.

Example 7.2: Mixed price/quantity competition

The demand and cost conditions and the strategy spaces are the same as in example 7.1 except here firm 1's strategy is a quantity and firm 2's strategy is a price. The demand system is now given by $p_1(q_1, p_2) = \frac{1}{2} + \frac{p_2}{2} - \frac{3q_1}{4}$ and $q_2(q_1, p_2) = 1 - \frac{q_1}{2} - p_2$.

The best response functions are: $b_1(p_2) = \frac{1}{3} + \frac{p_2}{3}$ and $b_2(q_1) = \frac{1}{2} - \frac{q_1}{4}$. The best response functions are monotonic, but one is negatively sloped and the other is positively sloped. These are given in figure 7.6. The externalities are also monotonic, but they are in opposite directions, since just as in example 7.1, price setters generate positive externalities and quantity setters generate negative externalities.

Figure 7.6 **Price/quantity competition**

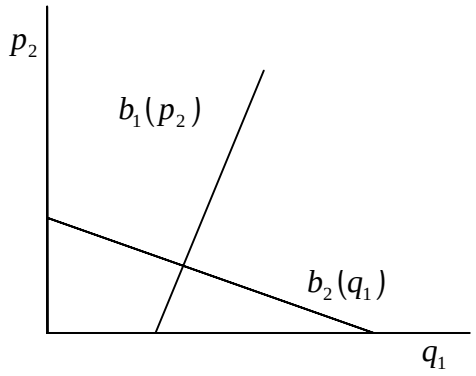
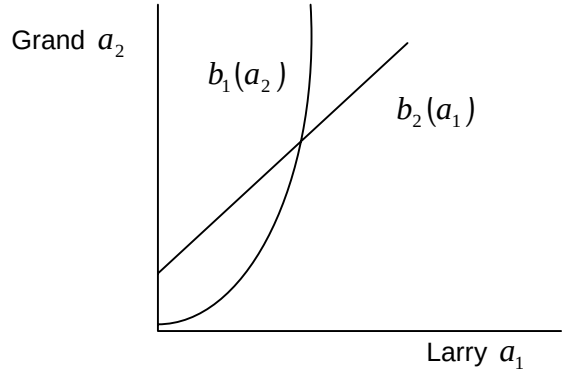


Figure 7.7 **Advertising game**



Example 7.3: Advertising

Two stores can each choose a level of advertising. Larry (player 1) is located inside a shopping mall and can only be reached by first passing a larger store called Grand (player 2). Larry's advertising level a_1 generates a positive externality $\left(\frac{\partial u_2}{\partial a_1} > 0\right)$ for Grand since more people will stop at Grand on the way to Larry's. Grand's advertising level a_2 imposes a negative externality $\left(\frac{\partial u_1}{\partial a_2} > 0\right)$ since Grand will steal customers from Larry.

More specifically, let Larry's utility function be given by $u_1(a_1, a_2) = b - \frac{a_2}{a_1} - a_1$ and let Grand's utility function be given by $u_2(a_1, a_2) = a_1 a_2 - \frac{a_2^2}{2} + 2a_2$. The strategy space of each player is again the strictly positive real line. Then it follows that the externalities move in opposite directions since $\frac{\partial u_1}{\partial a_2} = \frac{-1}{a_1} < 0$ and $\frac{\partial u_2}{\partial a_1} = a_2 > 0$. The best response functions shown in figure 7.7 are both positively sloped since $b_1(a_2) = a_2^{1/2}$ and $b_2(a_1) = a_1 + 2$.

Proposition 7.2 is the main result of this section. It states that all smooth monotonic games where each player has something to gain from moving first can be classified for each player as a carrot or stick game. Following the policy from section 7.2, the result is given for player 1 only.

Proposition 7.2

Let *G* be a smooth monotonic game with $u_1(s_1, b_2(s_1)) > \bar{v}_1$, for some s_1.

(3.4) If $sign\left[\frac{\partial u_1}{\partial s_2} b_2'(s_1)\right] = sign\left[\frac{\partial u_2}{\partial s_1}\right]$ then *G* is a carrot game for player 1.

(3.5) If $sign\left[\frac{\partial u_1}{\partial s_2} b_2'(s_1)\right] \neq sign\left[\frac{\partial u_2}{\partial s_1}\right]$ then *G* is a stick game for player 1.

Proposition 7.2 gives a simple way of determining whether a smooth monotonic game is a stick or a carrot game for a player when that player has something to gain from moving first. — simply compare $sign\left[\frac{\partial u_1}{\partial s_2} b_2'(s_1)\right]$ with $sign\left[\frac{\partial u_2}{\partial s_1}\right]$.

For smooth monotonic games, interior solutions suffice to ensure that each player has something to gain from moving first. To see why this is true consider the case at an interior Nash equilibrium (s_1^*, s_2^*). If player 1 were to alter their strategy choice marginally and player 2 were to respond optimally to this change, then the change in player 1's utility would be given by $\frac{\partial u_1(s_1^*, s_2^*)}{\partial s_1} + \frac{\partial u_1(s_1^*, s_2^*)}{\partial s_2} b_2'(s_1^*)$. Since (s_1^*, s_2^*) is a Nash equilibrium it follows that the first term $\frac{\partial u_1(s_1^*, s_2^*)}{\partial s_1} = 0$. Since externalities are non-zero and best response functions have non-zero slope in smooth monotonic games, it follows that the second term $\frac{\partial u_1(s_1^*, s_2^*)}{\partial s_2} b_2'(s_1^*) \neq 0$. Hence, player 1 has something to gain by moving first. For example, if $\frac{\partial u_1(s_1^*, s_2^*)}{\partial s_2} b_2'(s_1^*) < 0$ then player 1 has something to gain by lowering his strategy marginally.

An example of when non-interior equilibrium do not allow smooth monotonic games to be classified in terms of carrot and sticks is a zero sum game. It is a well known result that each Nash equilibrium of a two person zero sum game involves each player minimising their opponent's utility given their opponent's strategy. Monotonic externalities in a zero sum game imply that this minimisation occurs when each player chooses an endpoint of their strategy space. It is shown in section 7.5 that every zero sum two person game is not a carrot or a stick game for either player.

The simplicity of the comparison given in proposition 7.2 can be understood on an intuitive level by again considering small changes around an interior Nash

equilibrium point (s_1^*, s_2^*). From the earlier argument, on the one hand $\frac{\partial u_1}{\partial s_2} b_2'(s_1^*)$ measures the change in the utility of player 1 if they increase s_1 a small amount and player 2 responds optimally. On the other hand, $\frac{\partial u_2}{\partial s_1}$ measures the change in utility of player 2 from such a change. If the utilities of the players move in the same direction, then small changes that help player 1 will also help player 2. If the utilities move in opposite directions, then small changes that help player 1 will hurt player 2. The monotonicity of the externalities and best response functions ensures that this simple comparison is sufficient to determine if the game is a carrot or a stick game for a player.

The argument given in the last paragraph used an interior equilibrium to provide some intuition for the result. The proof of proposition 7.2 does not rely on any of the equilibria being interior.

Using proposition 7.2, the quantity setting game of example 7.1 can be shown to be a stick game for each player while the price setting version can be shown to be a carrot game for each player. Proposition 7.1 then gives the well known result that sequential movement will be mutually beneficial in the price setting game, but not in the quantity setting game.

When one firm chooses quantities and the other chooses prices, as in example 7.2, the game can be shown to be a stick game for the price setter and a carrot game for the quantity setter. Using proposition 7.1 in this example, sequential movement will be mutually beneficial if the quantity setter moves first, but not if the price setter moves first.

Furthermore, the price setter actually prefers to move second rather than first even though their reaction function is downward sloping. This is in contrast to Dowrick (1986, proposition 1) which states that if a firm's reaction function slopes downward, it will always prefer to move first rather than second. Dowrick's proposition requires that the externalities move in the same direction for both players, which is true if the goods are substitutes and the firms use the same strategic variables, as in example 7.1.

In example 7.2, the goods are no doubt substitutes, but the externalities move in opposite directions. Since this is a carrot game for the quantity setter, the price setter will benefit from sequential moves regardless of who moves first. Recall that moving first is always at least as good as simultaneous moves.

The ordering of profits for the price setter, firm 2, is $\pi_2^F > \pi_2^L > \pi_2^S$, where the superscripts stand for (F)ollower, (L)eader and (S)imultaneous, respectively. The exact values are: $\pi_2^F\left(\frac{3}{7},\frac{11}{28}\right)=\frac{121}{784}$, $\pi_2^L\left(\frac{19}{42},\frac{5}{14}\right)=\frac{175}{1176}$ and $\pi_2^S\left(\frac{18}{39},\frac{5}{13}\right)=\frac{150}{1014}$.

The advertising game of example 7.3 turns out to be a stick game for each player. Given the stick nature of the game, proposition 7.1 shows that sequential movement will never be mutually beneficial. A casual observation that the game involves strategic complements, however, might have erroneously suggested otherwise.

The following corollary, which is a direct result of proposition 7.2, relates strategic complements and strategic substitutes to smooth monotonic carrot and stick games.
$b_1(s_2)<0$

Call *G* a game with *strategic substitutes* whenever $b_1'(s_2)<0$ and $b_2'(s_1)<0$. Call *G* a game with *strategic complements* whenever $b_1'(s_2)>0$ and $b_2'(s_1)>0$.

Corollary 7.1

Let *G* be a smooth monotonic two person game.

(a) If *sign* $[b_1'(s_2)]$ = *sign* $[b_2'(s_1)]$ and *sign* $\left[\frac{\partial u_1}{\partial s_2}\right]$ = *sign* $\left[\frac{\partial u_2}{\partial s_1}\right]$, then *G* being a game with *strategic substitutes* is equivalent to *G* being a *stick game* for each player, and *G* being a game with *strategic complements* is equivalent to *G* being a *carrot game* for each player.

(b) If *sign* $[b_1'(s_2)]$ = *sign* $[b_2'(s_1)]$ and *sign* $\left[\frac{\partial u_1}{\partial s_2}\right] \neq$ *sign* $\left[\frac{\partial u_2}{\partial s_1}\right]$, then *G* being a game with *strategic substitutes* is equivalent to *G* being a *carrot game* for each player, and *G* being a game with *strategic complements* is equivalent to *G* being a *stick game* for each player.

The price and quantity setting games (example 7.1) are examples of (a) and the advertising game (example 7.3) is an example of (b).

7.4 Incentive games

Section 7.2 discussed the implications of the carrot and stick nature of two person games for transformations of those games into games with sequential movements.

The transformation from simultaneous moves to sequential moves involves changing only the information structure of the game.

The current section looks at a transformation that affects the player set, the information structure of the game and the payoffs. It starts with a two player simultaneous move game and transforms it to a four player two stage game. The new game is called an incentive game. The change to the information structure of the game comes from the introduction of the first stage. The second stage of the incentive game is identical to the original game in every respect except that the payoffs (incentives) of the players choosing there may be different. This section concentrates on smooth monotonic games.

One example of an incentive game is given by Sklivas (1987) (see also Fershtman and Judd 1987, Basu 1995 and Vickers 1985). In Sklivas' example, owners of firms manipulate the incentives of their managers in an attempt to obtain higher profits.

Bonanno and Vickers (1988) and Rey and Stiglitz (1995) consider situations of vertical separation[2] where producers can write exclusive contracts with retailers and manipulate the incentives of the retailers in a way that yields the producers higher profits. These examples turn out also to be incentive games.

Baye, Crocker and Ju (1996) model horizontal separation of firms into multiple competing divisions. The incentives of the divisions are affected by the breakup. So essentially, these are much like incentive games. However, since incentive games are defined here to consist of only one division per firm, the propositions of this section do not directly apply. Yet, the results here do help to describe what is going on in those situations.

Consider first the example given by Sklivas (1987). Some general results are then given for incentive games.

Example 7.4 Sklivas (1987)

There are two firms, each with an owner and a manager. Each owner will choose the incentives for their manager. The managers will later choose a quantity to produce.[3] The managers compete in quantities in a market with linear demand and constant marginal cost. Each manager's remuneration is a linear combination of

2 Vertical separation is just the reverse of vertical integration. An upstream firm and downstream firm vertically integrate into a merged firm, while a merged firm vertically separates into an upstream firm and a downstream firm.

3 Sklivas also looked at price competition. The current results apply to that case also. Only the quantity competition case is given here for ease of explanation.

sales and profit. Hence, each is given an incentive to maximise the following objective function: $O_i = (1-\alpha_i)\pi_i + \alpha_i R_i$, where π_i and R_i denote the profit and revenue (sales), respectively, of firm *i*=1,2, and α_i is the strategy of the owner. By choosing α_i different from zero, owners can manipulate the incentives of their managers away from profit maximisation. While this behaviour would never be optimal in monopoly or perfect competition, it may be in an oligopoly.

Letting C_i denote costs at firm *i*, it can be shown that $O_i = R_i - (1-\alpha_i) C_i$. This shows that raising α_i above zero effectively makes the manager act as if costs were lower than true costs. Lowering α_i below zero makes the manager act as if costs were above true costs.

After α_1 and α_2 have been chosen by the owners, each manager observes both managers' incentives and then competes in a quantity setting game.

If inverse demand is given by $P(Q) = 1 - Q$ and cost at each firm by $C_i(q_i) = cq_i$, then it is straightforward to derive that in the quantity setting game based on α_1 and α_2, the best response function of agent *i* will be $b_i(q_j) = \frac{1-(1-\alpha_i)c - q_j}{2b}$.

Notice that raising α_i will shift the best response function of that agent outward just as if marginal cost were reduced. Lowering α_i shifts the best response function inward. The best response functions are similar to those in figure 7.4.

Sklivas found that in the subgame perfect equilibrium of a two stage game where owners choose incentives in the first stage and managers compete in quantities in the second stage, the owners choose incentives that make the managers act as if costs are less than true costs. By making managers act as if costs are lower, owners induce them to produce more. Since an opposing manager can anticipate this over-production by a rival, the hope is that the opposing manager will cut back output. However, since both managers are induced to overproduce, the total industry output rises and the profits at each firm fall.

It is now shown how to obtain an incentive game from a two player smooth monotonic game $G = [(S_1, S_2), (u_1, u_2)]$.

In example 7.4, *G* is the game the managers play when each is given an incentive to maximise the firm's profit (ie $\alpha_1 = \alpha_2 = 0$).

In this section it is also assumed that the utility functions $u_1(s_1, s_2)$ and $u_2(s_1, s_2)$ are twice continuously differentiable, and for each strategy choice of an opponent, a

player's utility function is a strictly concave function of their own strategy. Strict concavity implies that $\frac{\partial^2 u_1(s_1,s_2)}{\partial s_1^2} < 0$ and $\frac{\partial^2 u_2(s_1,s_2)}{\partial s_2^2} < 0$ everywhere.

Incentive game

$\Gamma = \left[(S_{p1}, S_{p2}, S_{a1}, S_{a2}), (u_{p1}(\cdot), u_{p2}(\cdot), u_{a1}(\cdot), u_{a2}(\cdot))\right]$ is a *smooth monotonic stick/carrot nature preserving incentive game* based on a two player smooth monotonic game $G = [(S_1, S_2), (u_1, u_2)]$, if and only if, (4.1) to (4.9) are satisfied.

(4.1) (Principals' strategies) S_{p1} and S_{p2} are convex subsets of the real line *R* that include zero as an interior point.

(4.2) (Agents' strategies) $S_{a1} \equiv \{f_1 \text{ such that } f_1 : S_{p1} \times S_{p2} \to S_1\}$ and $S_{a2} \equiv \{f_2 \text{ such that } f_2 : S_{p1} \times S_{p2} \to S_2\}$, where S_1 and S_2 are both convex subsets of the real line *R* with the same smallest[4] element ε.

(4.3) (Principals' utility) For any strategy profile $(\alpha_1, \alpha_2, f_1, f_2)$, $u_{p1}(\alpha_1, \alpha_2, f_1, f_2) \equiv u_1(s_1, s_2)$, and $u_{p2}(\alpha_1, \alpha_2, f_1, f_2) \equiv u_2(s_1, s_2)$ where $s_1 \equiv f_1(\alpha_1, \alpha_2)$ and $s_2 \equiv f_2(\alpha_1, \alpha_2)$.

(Agents' utility) The conditions are given for agent 1.

(4.4) (Smoothness and concavity) $u_a(\alpha_1, \alpha_2, f_1, f_2)$ is twice continuously differentiable, and $\frac{\partial^2 u_a}{\partial s_1^2} < 0$ everywhere.

(4.5) (Agent 1's utility equals principal 1's utility if $\alpha_1 = 0$) For any strategy profile $(\alpha_1, \alpha_2, f_1, f_2)$ with $\alpha_1 = 0$, then $u_{a1}(\alpha_1, \alpha_2, f_1, f_2) \equiv u_{p1}(\alpha_1, \alpha_2, f_1, f_2)$.

(Effects of changing α) The conditions are given for principal/agent 1.

(4.6) (The best response function of each agent is independent of the incentives of the other agent) Let $b_1(s_2; \alpha_1)$ be the best response function of player 1 given $\alpha = (\alpha_1, \alpha_2)$. Then $\frac{\partial b_1(s_2;\alpha_1)}{\partial \alpha_2} = 0$ everywhere.

(4.7) (Raising α_1 harms principal 2) $\frac{\partial u_2(b_1(s_2;\alpha_1), s_2)}{\partial \alpha_1} \equiv \frac{\partial u_2(s_1, s_2)}{\partial s_1} \frac{\partial b_1(s_2;\alpha_1)}{\partial \alpha_1} < 0$ for all s_2 and α_1.

4 Often in economic examples the smallest element is often taken to be zero. The existence of a smallest number is only used in the proof of lemma 7.2 to show that the product of the slopes of the best response function are not greater than 1.

(4.8) (Preservation of the stick/carrot nature of *G*) For each $\alpha = (\alpha_1, \alpha_2)$ in the subgame $G(\alpha) = ((S_1, S_2), (u_{a1}(\cdot), u_{a2}(\cdot)))$ played by the agents, the direction of the externalities and the slopes of the reaction functions are equivalent to those in *G*.

(4.9) For each $\alpha = (\alpha_1, \alpha_2)$ the best response functions $b_1(s_2; \alpha_1)$ and $b_2(s_1; \alpha_2)$ satisfy $b_1'(s_2^*; \alpha_1) b_2'(s_1^*; \alpha_1) \neq 1$ whenever $s_1^* = b_1(s_2^*; \alpha_1)$ and $s_2^* = b_2(s_1^*; \alpha_2)$ are both met simultaneously.

In terms of example 7.4, the game *G* where each manager maximises its firm's profit is transformed into the four person incentive game Γ by turning each player into a team of two players, one principal (the owner) and one agent (the manager), and adding a preliminary stage before the game *G* is played. In this two stage game, the principals move simultaneously in the first stage. Each combination of the incentive choices $\alpha = (\alpha_1, \alpha_2)$ determines a subgame $G(\alpha)$ to be played by the agents in the second stage.

Condition (4.1) requires that each principal's strategy set includes zero as an interior point. The intention is to use zero as the case when the principal gives the manager the same incentives as the corresponding player in the original game *G*. The choice variable of each principal affects the incentives of the manager and the strategy sets may differ across players. By condition (4.7), it is assumed that whenever principal 1 increases α_1, this induces agent 1 to act in a way that lowers the utility of the other principal.

Condition (4.2) just says that each agent's strategy describes what to do in each subgame $G(\alpha)$, and that the strategy set in each subgame is the same as the one used by the corresponding player in *G*. This condition connects the choices of the agent in Γ to those of the player in *G*.

Condition (4.3) requires that the payoff function of each principal be the same as the payoff function of the corresponding player in *G*. Condition (4.5) requires the same thing for agent *i* only where the principal has chosen $\alpha_i = 0$. This condition connects the principal's and the agent's utility to that of the player in *G*.

Condition (4.4) is a smoothness and concavity assumption on the payoff of each agent, which will allow the use of the implicit function theorem to derive some comparative statics results about each agent's best response function.

Condition (4.6) says that each agent's best response function is not affected by the incentives of the rival. Nonetheless, the incentives of the rival agent will affect the payoff of an agent since it affects the rival's behaviour.

As was mentioned earlier, condition (4.7) states that raising α_1 gives agent 1 incentives that lower the utility to principal 2.

Raising α_1 can be diagrammatically represented by shifting player 1's best response function in or out. Consider figure 7.4. If player 1 inflicts a negative externality on player 2, then raising α_1 involves shifting player 1's best response function out. In this case, for each strategy choice of player 2, player 1's best response involves a higher level of strategy, which by the negative externality implies player 2 will be worse off.

If, alternatively, player 1 offers a positive externality to player 2, then raising α_1 corresponds to shifting player 1's best response function in. In figure 7.4 this means the best response of player 1 is smaller for each level chosen by player 2. Notice that the shifts need not be parallel, though they may be, and in fact they are, in the example of Sklivas (1987).

Condition (4.8) restricts incentive games by not allowing stick games to be transformed into carrot games or vice versa.

Condition (4.9) will be used in the comparative static results about changes in the equilibrium choices in the agents' subgame brought about by changes in α_1 or α_2.

A strategy combination $(\alpha_1, \alpha_2, f_1, f_2)$ is a *subgame perfect equilibrium* of incentive game Γ, if and only if (a) $(\alpha_1, \alpha_2, f_1, f_2)$ is a Nash equilibrium in Γ; and (b) $f_1(\alpha) = s_1$, and $f_1(\alpha) = s_2$ define a Nash equilibrium in $G(\alpha)$ for each α.

All of the propositions of this section are made under the following assumption, which is called assumption (A).

(A) Γ is a smooth monotonic stick/carrot nature preserving incentive game based on a smooth monotonic game G and $G(\alpha)$ has a unique Nash equilibrium which is interior for each α.

The assumption of uniqueness and an interior equilibrium has multiple important implications. First, it implies that the subgame perfect equilibrium strategies of the agents, f_1 and f_2, are uniquely determined. Second, the assumption of an interior solution in a smooth monotonic game G implies by proposition 7.2 that the game can be classified in terms of carrots and sticks for each player. By (4.8), the carrot/stick nature of the game will be the same in each subgame $G(\alpha)$. Last, it simplifies the comparative statics.

The following lemmas are used in the proofs of the results. The first is an implication of the unique interior equilibrium in each subgame. The second lemma uses this implication to derive some comparative statics results.

Lemma 7.2

Let *A* and *B* be convex subsets of *R*, each containing the same smallest element ε. Let $b_1 : A \to B$, and $b_2 : B \to A$ be differentiable functions. If the composite function $f(x) = b_1(b_2(x))$ has a unique fixed point x^* that is interior, then $f'(x^*) = b_1'(y^*)b_2'(x^*) \leq 1$, where $y^* \equiv b_2(x^*)$.

The intention is that $b_1(y)$ is the best response function of player 1 and $b_2(x)$ is the best response function of player 2. Whenever x^* is a fixed point of $f(x^*)$, it follows that (x^*, y^*) is a Nash equilibrium. Lemma 7.2 states that if a Nash equilibrium is unique and interior, then the product of the slopes of the best response functions must be no greater than 1 at the equilibrium point.[5] This result is used to obtain the comparative statics results presented in lemma 7.3. Results are given for changing α_1.

Lemma 7.3

Under assumption (A), let $(\alpha_1, \alpha_2, f_1, f_2)$ be a strategy for which $(f_1(\alpha'), f_2(\alpha'))$ is a Nash equilibrium in $G(\alpha')$ for each $\alpha' \in S_{p1} \times S_{p2}$. Then

(4.10) $sign\left[\frac{\partial f_1(\alpha)}{\partial \alpha_1}\right] \neq sign\left[\frac{\partial u_2}{\partial s_1}\right]$, and

(4.11) $sign\left[\frac{\partial f_2(\alpha)}{\partial \alpha_1}\right] \neq sign\left[\frac{\partial u_1}{\partial s_2}\right]$ if *G* is a carrot game for player 1, and $sign\left[\frac{\partial f_2(\alpha)}{\partial \alpha_1}\right] = sign\left[\frac{\partial u_1}{\partial s_2}\right]$ if *G* is a stick game for player 1.

(4.12) $sign\left[\frac{\partial b_1(s_2;\alpha_1)}{\partial \alpha_1}\right] \neq sign\left[\frac{\partial u_2}{\partial s_1}\right]$.

Lemma 7.3 connects the effect of changing α_1 in the incentive game to the payoff functions $u_1(s_1, s_2)$ and $u_2(s_1, s_2)$ in the original game *G*. This can be done since the restriction that the stick/carrot nature is preserved (4.8) implies that the direction of the externalities is preserved.

Whenever player 1 is said to be better off in Γ than in *G*, this means that in any subgame perfect equilibrium of Γ, principal 1's utility is higher than player 1's utility in the unique Nash equilibrium of *G*. Recall that principal 1 and player 1

[5] This result could have been obtained by the weaker assumption that only one player has a minimum or a maximum strategy.

have the same utility so they can be thought of as the same player and their utilities compared across games.

The first result for incentive games is that the carrot or stick behaviour of each agent will be magnified (proposition 7.3). If *G* is a stick game for a player, then the agent of that player will be induced to act more harmfully in Γ, that is, behave more stick like. Alternatively, if *G* is a carrot game for a player, then their agent will behave more carrot like in Γ. In this sense, the incentive game tends to magnify the carrot or stick behaviour of a player.

Proposition 7.3

Under assumption (A), the carrot or stick behaviour of each player is magnified in the incentive game. That is, for any subgame perfect equilibrium $(\alpha_1^*, \alpha_2^*, f_1^*, f_2^*)$:

- if *G* is a carrot game for player 1, then $\alpha_1^* < 0$; and
- if *G* is a stick game for player 1, then $\alpha_1^* > 0$.

Proposition 7.3 can be used to show that if the original game is a carrot game for a player, then their opponent will be better off in the incentive game. This result is rather intuitive since more carrot like behaviour will surely benefit an opponent.

Proposition 7.4

Under assumption (A), if *G* is a carrot game for player 1, then player 2 will be better off in Γ than in *G*.

Note that this proposition does not depend on the original game being a carrot game for player 2. Indeed, player 2 will benefit in the incentive game even if their original game is a stick game. In the price/quantity setting example of section 7.3, it can be predicted that the price setter will definitely benefit from any incentive game. This comes from the fact that quantity setters are carrot players in a price/quantity game and carrot players help their opponents in incentive games. Whether or not the quantity setter gains is ambiguous, since there will be offsetting effects.

Applying proposition 7.4 to a game that is a carrot game for each player, both players will be better off in the incentive game. So, for example, incentive games based on price competition will benefit both firms.

Corollary 7.2

Under assumption (A), if *G* is a carrot game for each player, then both are better off in Γ than in *G*.

In corollary 7.2, the only type of symmetry required is that G is a carrot game for each player. The players may have different strategy spaces and different payoff functions.

Sklivas (1987) looked at both price and quantity competition in symmetric games. Symmetry is defined later, but basically a game is symmetric if the players are *a priori* identical up to their names. Sklivas found that under price competition, both players were better off in the incentive game and under quantity competition both players were worse off.

The result for price competition follows directly from corollary 7.2 and requires only that the game is a carrot game for each player. The symmetry implicitly used by Sklivas about the payoff functions and strategy sets is not necessary. The result for quantity competition, however, requires such symmetry between the players. Without enough symmetry, for example, if the players have different payoffs or strategies, the result that both are worse off may not hold.

In general, if the original game is a stick game for each player, proposition 7.3 shows that each principal will magnify the incentives of their agent to harm the other principal. This implies that at least one player will be worse off.

Proposition 7.5

Under assumption (A), if *G* is a stick game for each player, then at least one player is worse off in Γ than in *G*.

To obtain the result that both players are worse off, more symmetry is needed between the players than simply that both are stick players. Use the following definition of symmetry given in Gal-Or (1985).

A two player game $G = [(S_1, S_2), (u_1, u_2)]$ is *symmetric*, if and only if $S_1 = S_2$, and $u_1(s_1, s_2) = u_2(s_2, s_1)$ for all $s_1, s_2 \in S_1$.

This definition of symmetry captures the idea that the two players are *a priori* identical up to their names. The price setting and quantity setting games of example 7.1 are symmetric. The advertising game (example 7.3), and the mixed price/quantity competition (example 7.2) are not symmetric, nor is example 7.5.

Based on the above definition of symmetry, define an incentive game Γ to be a symmetric incentive game, if and only if: (a) $S_{p1} = S_{p2}$ and $S_{a1} = S_{a2}$ and (b) $u_{p1}(\alpha_1, \alpha_2, f_1, f_2) = u_{p2}(\alpha_2, \alpha_1, f_2, f_1)$ and $u_{a1}(\alpha_1, \alpha_2, f_1, f_2) = u_{a2}(\alpha_2, \alpha_1, f_2, f_1)$ for all $(\alpha_1, \alpha_2, f_1, f_2)$.

This restriction requires that the game *G* is symmetric and also that the two principal-agent teams are identical *a priori* up to their names.

As was mentioned, the examples studied by Sklivas are symmetric incentive games.

In a symmetric game, attention is often given to symmetric equilibrium. In an incentive game Γ, an equilibrium $(\alpha_1, \alpha_2, f_1, f_2)$ is symmetric if $\alpha_1 = \alpha_2$ and $f_1 = f_2$. The following result for stick games corresponds to corollary 7.2 for carrot games. Notice the use of symmetry in this result.

Corollary 7.3

Under assumption (A), if Γ is a symmetric incentive game and *G* is a stick game for each player, then in any symmetric subgame perfect equilibrium of Γ, both players are worse off than in *G*.

Corollary 7.3 implies that both players are worse off in a symmetric incentive game based on quantity competition. The next example is an asymmetric quantity setting game where one player has a cost advantage. The player with the cost advantage is found to be better off in the unique subgame perfect equilibrium of the incentive game.

Example 7.5

Demand is linear and given by $P(q_1 + q_2) = 1 - (q_1 + q_2)$, and marginal costs are constant at a firm and are given by $c_1 = 0.65$ and $c_2 = 0.50$ for firms 1 and 2, respectively. The choice of an owner *i* is a parameter α_i, which is the weight placed on the profits in the manager's objective function. Manager *i*'s objective will be to maximise $O_i = R_i - (1 - \alpha_i)c_i q_i$.

For each (α_1, α_2) there is a unique equilibrium in the second stage game. Concentrate on equilibrium where the quantities are positive for both firms. Then it is straightforward to show that the equilibrium profits are given by:

$$\pi_i(\alpha_1, \alpha_2) = \frac{1}{9}\left[1 + (1 - \alpha_j)c_j - (2 + \alpha_i)c_i\left[1 + (1 - \alpha_j)c_j - 2(1 - \alpha_i)c_i\right]\right] \text{ for } i=1,2, \text{ and } j \neq i.$$

An interior equilibrium in the preliminary stage requires that $\frac{\partial \pi_i(\alpha_1, \alpha_2)}{\partial \alpha_i} = 0$ for i=1,2. This implies that the equilibrium α_i^*s are given by

$1 - \alpha_i^* = \frac{8c_i - 1 - 2c_j}{5c_i}$, for i=1,2 and $j \neq i$.

For the given values $c_1 = 0.65$ and $c_2 = 0.50$, $\alpha_1^* = \frac{5}{325} > 0$ and $\alpha_2^* = \frac{8}{25} > 0$. Hence, both players use more harmful incentives. The equilibrium quantities are given by $q_i^* = \frac{1 - 2(1-\alpha_i)c_i + (1-\alpha_j)c_j}{3}$ for i=1,2 and $j \neq i$. From this it is straightforward to show that, in moving from the original game G to the incentive game Γ, the equilibrium quantity and profits of firm 1 fall and the equilibrium quantity and profits of firm 2 rise. The firm with the cost advantage is made slightly better off even though both firms act more harmfully. The actual values are $q_1^G = 0.067$, $q_2^G = 0.2167$, $\pi_1^G = 0.044$, $\pi_2^G = 0.0469$, $q_1^\Gamma = 0.02$, $q_2^\Gamma = 0.32$, $\pi_1^\Gamma = 0.0002$ and $\pi_2^\Gamma = 0.0512$, where the superscripts G and Γ stand for the original game and incentive game, respectively.

Intuitively, more harmful incentives have a cost, which is determined here by the cost of output. For firm 1, the cost of bashing the opponent with a stick is higher than for firm 2 since firm 1's marginal cost is higher. This cost advantage leads firm 2 ultimately to bash the opponent more and firm 1 to bash less.

The example of vertical separation given in Bonanno and Vickers (1988) can also be shown to be an incentive game.[6]

Example 7.6

Two producers make imperfect substitutes and compete in prices. Each hires a retailer to sell its product and the producer can extract all the retailer's profit with a franchise fee.[7] The producer also chooses a per unit wholesale price ω_i to charge the retailer, which may or may not differ from the true marginal cost. The retailers then compete in retail prices, taking each other's wholesale prices as given.

In the terminology of incentive games, the original game is defined to be the one the retailers play when wholesale prices are chosen to equal marginal cost. In the first

6 So can the example of Rey and Stiglitz (1995) when franchise fees are allowed.

7 If a franchise fee is not allowed, then the objective of the producer may not be to maximise original profits and the propositions of this section cannot be applied.

stage of the incentive game, each producer *i* chooses a wholesale price ω_i, which determines the incentives of the retailer. In the second stage, each retailer chooses a retail price p_i to maximise its profit, which will differ from the firm's true profit when the wholesale price differs from marginal cost.

Let demand at retailer *i* be given by $D_i(p_1, p_2) = A - bp_i + dp_j$, for *i*=1,2 ($j \neq i$), where p_1 and p_2 are the retail prices. The marginal cost of each producer firm is a constant $c > 0$. The profit to retailer *i* is $\pi_i^R(p_1, p_2) = (p_i - \omega_i)D_i(p_1, p_2)$, since by assumption the retailer incurs no cost other than the wholesale price. This leads to the best response function $b_i(p_j; \omega_i) = \frac{(A + b\omega_i)}{2b} + \frac{dp_j}{2b}$. To get this in a form to use the propositions of this section, define $\alpha_i = c - \omega_i$ so that the true incentives are given when $\omega_i = c$ (ie $\alpha_i = 0$). Then clearly raising α_i is done by lowering wholesale price ω_i below marginal cost. Alternatively, lowering α_i is done by raising the wholesale price ω_i above marginal cost. Each principal *i* (producer) will choose α_i to maximise the true profit of firm *i*, $\pi_i = (p_i - c)D_i(p_1, p_2)$, since it can extract the retailer's profit by use of a fixed franchise fee.

The game played by the retailers is a carrot game for each pair (α_1, α_2). Applying corollary 7.2 gives the result of Bonanno and Vickers (1988) that both producers are made better off from vertical separation.

Baye, Crocker and Ju (1996) analyse horizontal separation. A firm horizontally separates by breaking up into multiple competing divisions. The authors start with two firms competing in quantities, so it is a stick game for each player. They then allow each firm to break up into multiple competing divisions in an previous stage. The effect of this breakup is to make the broken up firms more harmful to their rivals than before. While the propositions of this section do not directly apply to this example, the spirit does. Since the original game is a stick game for each player, each will tend to do more harmful things in the two stage game. Breaking up into multiple divisions is more harmful. Since the players are symmetric, it turns out that they are both worse off after the break up.

7.5 Relationship to Stackelberg solvable games

Stick games appear to be similar in some respects to zero sum games and carrot games appear to be similar to coordination games. This section shows that Stackelberg solvable games, which include zero sum games and coordination games, are neither carrot nor stick games. Hence, making use of the distinction between carrots and sticks means going beyond Stackelberg solvable games.

d'Aspremont and Gerard-Varet (1980) introduced the notion of Stackelberg solvable games. They were interested in games where changes to the order of moves, or the incentives of agents, do not affect the outcome of the game.

A game G is called *Stackelberg solvable* if it possesses at least one Nash equilibrium (s_1, s_2) such that:

(5.1) s_1 maximises $u_1(s_1, b_2(s_1))$, and

(5.2) s_2 maximises $u_2(b_1(s_2), s_2)$.

It follows directly from the definition that $\bar{v}_1$ and $\bar{v}_2$ in a Stackelberg solvable game G are determined by a pair (s_1, s_2) that satisfies (5.1) and (5.2).

Stackelberg solvable games include both zero sum games and coordination games. They also include games such as the prisoners' dilemma which is neither zero-sum nor a coordination game. Figure 7.8 is an example of a zero sum game and figure 7.9 is an example of a coordination game.

The game of figure 7.8 has a unique Nash equilibrium (T, L) and thus $\underline{v}_1 = \bar{v}_1 = 5$, and $\underline{v}_2 = \bar{v}_2 = -5$. The game of figure 7.9 has two Nash equilibrium, (T, L) and (B, R). Since they can be Pareto ranked, $\underline{v}_1 = \underline{v}_2 = 1$, and $\bar{v}_1 = \bar{v}_2 = 2$.

Figure 7.8 **Game 4**

		Player 2	
		L	R
Player 1	T	5,–5	10,–10
	B	3,–3	–2,2

Figure 7.9 **Game 5**

		Player 2	
		L	R
Player 1	T	2,2	–1,–1
	B	0,0	1,1

In a zero sum game, the players' objectives are diametrically opposed, while in coordination games they are perfectly aligned. In this sense, these games are at opposite extremes. However, if pre-play communication is allowed, but pre-play commitment to strategies is not, then zero sum games and coordination games share a similarity in that e*ach player is indifferent between the simultaneous move game, the sequential move game where player 1 moves first, and the sequential move game where player 2 moves first.*

Sequential moves might allow the player moving first to (a) commit to a strategy, and (b) choose a Pareto optimal Nash equilibrium strategy. Pre-game communication eliminates the value of (b), so concentrate on (a). It is not that (b) is unimportant. In fact, although (b) is of no value to players in zero sum games, players in a coordination game with multiple equilibrium may find great value in

(b). Pre-game communication solves the problem of coordinating on the best equilibrium and allows concentration on the value of commitment.

That the players will be indifferent between the three situations in a coordination game seems rather trivial if they can communicate beforehand. Since the Nash equilibrium in a coordination game can be Pareto ranked, the players will no doubt choose strategies that yield the maximum Nash equilibrium payoff of the simultaneous move game. In either of the sequential move versions of the original game, the player moving first will be compelled to choose the strategy that leads to their maximum Nash equilibrium payoff and the opponent will follow suit. It is because of the alignment of objectives and pre-play communication that sequential movements are of no value.

The argument that (a) has no value for two person zero sum games is not so straightforward. It is a result of the minimax theorem. Von Neumann and Morgenstern discuss this result in their book *Theory of Games and Economic Behaviour* (section 14.2). In a two person zero sum game, moving second 'appears' to be better since you can always find out your opponent's strategy and minimise against it. However, it turns out that there is no difference between moving first, second, or simultaneously when an equilibrium exists.

The main result of this section is as follows.

Proposition 7.6

If a game is Stackelberg solvable, then it is neither a carrot game nor a stick game for either player.

7.6 Conclusions

Carrots and sticks have provided insights into whether or not both players will benefit from transformations to games with sequential movements or to incentive games.

Section 7.2 showed that sequential movement is mutually beneficial in carrot games and unilaterally beneficial in stick games (proposition 7.1).

Section 7.3 defined smooth monotonic games. It showed every such game is either a carrot or a stick game for each player, provided the Nash equilibrium are interior (proposition 7.2). Carrot and stick games were compared to the notions of strategic substitutes and strategic complements (corollary 7.1). An advertising game

(example 7.3) was used to show how standard expectations about strategic complements and strategic substitutes might lead one astray.

The results obtained for smooth monotonic games were used in section 7.4 to obtain some further results for incentive games. Specifically, it was shown that if a game is a carrot game for one player, the opponent will benefit in any incentive game (proposition 7.3). This leads to the result that incentive games based on carrot games for both players generate outcomes that are mutually beneficial (corollary 7.2). For incentive games based on stick games, the results were weaker. The corresponding result to proposition 7.3, that stick players will make their opponents worse off in an incentive game, was not obtained. Proposition 7.5 showed that an incentive game based on a stick game for each player will harm at least one player, but not necessarily both.

Appendix 7

Proof of proposition 7.1

Let (s_1^*, f_2^*) be a subgame perfect equilibrium in G'.

If G is a carrot game then there exists a $s_1 \in S_1$, such that $u_1(s_1, b_2(s_1)) > \bar{v}_1$. Then $u_1(s_1^*, f_2^*(s_1^*)) = u_1(s_1^*, b_2(s_1^*)) > \bar{v}_1$ since (s_1^*, f_2^*) is a subgame perfect equilibrium. Thus by (2.1) it follows that $u_2(s_1^*, f_2^*(s_1^*)) > \bar{v}_2$.

If G is a stick game, by similar reasoning using (2.2) and the definition of a subgame perfect equilibrium, $u_1(s_1^*, f_2^*(s_1^*)) > \bar{v}_1$ and $u_2(s_1^*, f_2^*(s_1^*)) < \underline{v}_2$.

Proof of lemma 7.1

First consider the case where $\bar{v}_1$ and $\bar{v}_2$ are defined by the same Nash equilibrium (s_1, s_2). Suppose that G is a carrot or a stick game for player 1. Then it is shown that G cannot be a carrot or a stick game for player 2.

By the assumption for player 1, there exists a $s_1' \neq s_1$, such that $u_1(s_1', b_2(s_1')) > \bar{v}_1 = u_1(s_1, s_2)$. Since this is 2×2 game it follows that s_1' is the only other strategy for player 1. Also, $b_2(s_1') \neq s_2$, since if $b_2(s_1') = s_2$, then $u_1(s_1', s_2) > u_1(s_1, s_2)$ which contradicts that (s_1, s_2) is a Nash equilibrium. So $b_2(s_1') \equiv s_2'$ is the only other strategy of player 2. For G to be a carrot or stick game for player 2 requires that $u_2(b_1(s_2'), s_2') > \bar{v}_2 = u_2(s_1, s_2)$. If $b_1(s_2') = s_1'$, then (s_1', s_2') is a Nash equilibrium, which violates the assumption that the value $\bar{v}_2$ is defined by (s_1, s_2). Alternatively, if

$b_1(s_2') = s_1$, then $u_2(s_1, s_2') > u_2(s_1, s_2)$ which contradicts that (s_1, s_2) is a Nash equilibrium. Either case gives a contradiction, so *G* cannot be a carrot or stick game for player 2.

Next, suppose that $\bar{v}_1 = u_1(s_1, s_2)$ and $\bar{v}_2 = u_2(s_1', s_2')$ and $(s_1, s_2) \neq (s_1', s_2')$. Then it follows that both $s_1 \neq s_1'$ and $s_1' \neq s_2'$. If $s_1 = s_1'$, then $b_2(s_1)$ being a unique valued implies that $s_2 = s_2'$. Similarly, if $s_2 = s_2'$, then $b_1(s_2)$ being unique valued implies that $s_1 = s_1'$. Since (s_1', s_2') is a Nash equilibrium and best responses are unique, $b_1(s_2') = s_1'$, and $b_2(s_1') = s_2'$. Since s_1 defines $\bar{v}_1$, and s_1' is the only other strategy for player 1, it follows that $u_1(s_1', b_2(s_1')) \leq \bar{v}_1$, and thus *G* is neither a carrot nor a stick game for player 1. A similar argument holds for player 2.

The following lemma is used in the proof of proposition 7.2.

Lemma 7.4

Let *G* be a smooth monotonic two player game and let (s_1^n, s_2^n) be a Nash equilibrium in *G*.

(3.6) If $\left[\frac{\partial u_1}{\partial s_2} b_2'(s_1)\right] > 0$, then $u_1(s_1, b_2(s_1)) > u_1(s_1^n, s_2^n)$ implies $s_1 > s_1^n$.

(3.7) If $\left[\frac{\partial u_1}{\partial s_2} b_2'(s_1)\right] < 0$, then $u_1(s_1, b_2(s_1)) > u_1(s_1^n, s_2^n)$ implies $s_1 < s_1^n$.

Proof of lemma 7.4

The proof is of the contrapositive of the part of (3.6) following the comma. That is, under the assumption $\left[\frac{\partial u_1}{\partial s_2} b_2'(s_1)\right] > 0$ it is proved that $[\text{not } (s_1 > s_1^n)]$ implies $[\text{not } (u_1(s_1, b_2(s_1)) > u_1(s_1^n, s_2^n))]$. The contrapositive of the corresponding part of (3.7) can be proved in a similar manner.

Since $s_1 = s_1^n$ implies $u_1(s_1, b_2(s_1)) = u_1(s_1^n, s_2^n)$, the proof concentrates on the case when $s_1 < s_1^n$. Suppose $s_1 < s_1^n$. By the assumption that $\left[\frac{\partial u_1}{\partial s_2} b_2'(s_1)\right] > 0$, either $[\frac{\partial u_1}{\partial s_2} < 0$ and $b_2'(s_1) < 0]$ or $[\frac{\partial u_1}{\partial s_2} > 0$ and $b_2'(s_1) > 0]$. In the first case,

$u_1(s_1, b_2(s_1)) < u_1(s_1, s_2^n) \leq u_1(s_1^n, s_2^n)$ where the first inequality follows since $\frac{\partial u_1}{\partial s_2} < 0$ and $b_2(s_1) > s_2^n$, and the second inequality follows from the utility maximisation of player 1. The second case is proved in the same way.

Proof of proposition 7.2

In the following, consider an arbitrary s_1 chosen such that $u_1(s_1, b_2(s_1)) > \bar{v}_1$ and let (s_1^n, s_2^n) satisfy $u_2(s_1^n, s_2^n) = \bar{v}_2$ and $u_2(s_1', s_2') = \underline{v}_2$. Then $u_1(s_1, b_2(s_1)) > u_1(s_1^n, s_2^n)$ and $u_1(s_1, b_2(s_1)) > u_1(s_1', s_2')$. The following uses such $s_1, b_2(s_1), s_1^n, s_2^n, s_1', s_2'$.

The If part of (3.4) is true only if [$\frac{\partial u_2}{\partial s_1} > 0$ and $sign\left[\frac{\partial u_1}{\partial s_2}\right] = sign[b_2'(s_1)]$] or [$\frac{\partial u_2}{\partial s_1} < 0$ and $sign\left[\frac{\partial u_1}{\partial s_2}\right] \neq sign[b_2'(s_1)]$]. In the first case, statement (3.6) in lemma 7.4 says that $u_1(s_1, b_2(s_1)) > u_1(s_1^n, s_2^n)$ implies $s_1 > s_1^n$. But then $u_2(s_1^n, s_2^n) < u_2(s_1, s_2^n) \leq u_2(s_1, b_2(s_1))$ where the first inequality follows since $s_1 > s_1^n$ and $\frac{\partial u_2}{\partial s_1} > 0$ and the second inequality follows from the utility maximisation of player 1. In the second case, statement (3.7) of lemma 7.4 says that $u_1(s_1, b_2(s_1)) > u_1(s_1^n, s_2^n)$ implies $s_1 < s_1^n$. But then $u_2(s_1^n, s_2^n) < u_2(s_1, s_2^n) \leq u_2(s_1, b_2(s_1))$ where the first inequality follows since $s_1 < s_1^n$ and $\frac{\partial u_2}{\partial s_1} < 0$ and the second inequality follows by the utility maximisation of player 2.

The If-part of (3.5) is true only if [$\frac{\partial u_2}{\partial s_1} < 0$ and $sign\left[\frac{\partial u_1}{\partial s_2}\right] = sign[b_2'(s_1)]$] or [$\frac{\partial u_2}{\partial s_1} > 0$ and $sign\left[\frac{\partial u_1}{\partial s_2}\right] \neq sign[b_2'(s_1)]$].

In the first case, statement (3.6) of lemma 7.1 says that $u_1(s_1, b_2(s_1)) > u_1(s_1', s_2')$ implies $s_1 > s_1'$. But then $u_2(s_1, b_2(s_1)) < u_2(s_1', b_2(s_1)) \leq u_1(s_1', s_2')$ where the first inequality follows from $s_1 > s_1'$ and $\frac{\partial u_2}{\partial s_1} < 0$ and the second inequality follows from the utility maximisation of player 2. The result in the second case can be proved in the same manner.

Proof of lemma 7.2

Consider the differentiable, and thus continuous, one to one function $g(x)=x$ defined from *B* onto *B*. At the unique fixed point x^*, then $f(x^*)=g(x^*)$. Furthermore, since x^* is a unique fixed point of $f(.)$ and $f(.)$ is continuous, it follows that for all $x<x^*, g(x)<f(x)$. This is because if there was an $x<x^*$, with $g(x)\geq f(x)$, then since $g(\varepsilon)=\varepsilon$, and $f(\varepsilon)\geq\varepsilon$, it would follow by the intermediate value theorem that there was an $x'\in[\varepsilon,x]$ such that $f(x')=g(x')=x'$. But this contradicts the uniqueness of the fixed point x^*.

Since $g(x)<f(x)$ for all $x<x^*$ and $g(x^*)=f(x^*)$and x^* is interior, then for all $\delta<0$ in some neighbourhood of x^*, $\frac{g(x^*+\delta)-g(x^*)}{\delta}>\frac{f(x^*+\delta)-f(x^*)}{\delta}$. Taking the limit as $\delta\to 0$ gives $g'(x^*)\geq f'(x^*)$. But $g'(x^*)=1$, and $f'(x^*)=b_1'(y^*)b_2'(x^*)$.

Proof of lemma 7.3

The first order conditions are:

(A.1) $\frac{\partial u_{a1}(\alpha_1,\alpha_2,s_1,s_2)}{\partial s_1}=0$, and

(A.2) $\frac{\partial u_{a2}(\alpha_1,\alpha_2,s_1,s_2)}{\partial s_2}=0$

By assumption that an interior equilibrium exists for each α, there are solutions to (A.1) and (A.2) for all $(\alpha_1,\alpha_2,s_1,s_2)$. Assumption (4.4) allows the use of the implicit function theorem on either one of the first order conditions separately, for example, to find the slope of the best response function, since by strict concavity $\frac{\partial^2 u_{a1}}{\partial s_1^2}<0$ and $\frac{\partial^2 u_{a2}}{\partial s_2^2}<0$ (eg Fulks 1978).

To use the implicit function theorem on both equations simultaneously to calculate, for example, $\frac{\partial f_1(\alpha)}{\partial \alpha_1}$, requires that $D=\left[\left(\frac{\partial^2 u_{a1}}{\partial s_1^2}\right)\left(\frac{\partial^2 u_{a2}}{\partial s_2^2}\right)-\left(\frac{\partial^2 u_{a1}}{\partial s_2 \partial s_1}\right)\left(\frac{\partial^2 u_{a2}}{\partial s_1 \partial s_2}\right)\right]\neq 0$

at s_1 and s_2 that satisfy (A.1) and (A.2) simultaneously. Using the implicit function theorem on the first order conditions independently gives:

$$b_1'(s_2;\alpha_1) = -\frac{\left(\frac{\partial^2 u_{a1}}{\partial s_2 \partial s_1}\right)}{\left(\frac{\partial^2 u_{a1}}{\partial s_1^2}\right)} \text{ and } \quad b_2'(s_1;\alpha_2) = -\frac{\left(\frac{\partial^2 u_{a2}}{\partial s_1 \partial s_2}\right)}{\left(\frac{\partial^2 u_{a2}}{\partial s_2^2}\right)}.$$

Substituting these values into D and rearranging terms gives

$$D = \left(\frac{\partial^2 u_{a1}}{\partial s_1^2}\right)\left(\frac{\partial^2 u_{a2}}{\partial s_2^2}\right)\left[1 - b_1'(s_2;\alpha_1)b_2'(s_1;\alpha_2)\right]$$

By lemma 7.2 and (4.9), $b_1'(s_2;\alpha_1)b_2'(s_1;\alpha_2) < 1$ at s_1 and s_2 that satisfy both (A.1) and (A.2) simultaneously. This together with assumption (4.4) that the utility functions are strictly concave implies $D > 0$.

From the first order condition of player 1, using the implicit function theorem gives

(A.3) $$\frac{\partial b_1(s_2;\alpha_1)}{\partial \alpha_1} = \frac{\left(\frac{\partial^2 u_{a1}}{\partial \alpha_1 \partial s_1}\right)}{-\left(\frac{\partial^2 u_{a1}}{\partial s_1^2}\right)}$$

By assumption (4.7) that $\frac{\partial u_2\left(b_1(s_2;\alpha_1), s_2\right)}{\partial \alpha_1} \equiv \frac{\partial u_2}{\partial s_1}\frac{\partial b_1(s_2;\alpha_1)}{\partial \alpha_1} < 0$, it follows that

(A.4) $$sign\left(\frac{\partial b_1(s_2;\alpha_1)}{\partial \alpha_1}\right) \neq sign\left(\frac{\partial u_2}{\partial s_1}\right)$$

This proves (4.12). This result and (A.3) and strict concavity of the utility function gives

(A.5) $$sign\left(\frac{\partial^2 u_{a1}}{\partial \alpha_1 \partial s_1}\right) \neq sign\left(\frac{\partial u_2}{\partial s_1}\right)$$

Using the implicit function theorem on both first order conditions gives:

(A.6) $$\frac{\partial f_1(\alpha)}{\partial \alpha_1} = \frac{\left(\frac{\partial^2 u_{a1}}{\partial \alpha_1 \partial s_1}\right)\left(-\frac{\partial^2 u_{a2}}{\partial s_2^2}\right)}{D}, \text{ and}$$

(A. 7) $$\frac{\partial f_2(\alpha)}{\partial \alpha_1} = \frac{\left(\frac{\partial^2 u_{a1}}{\partial \alpha_1 \partial s_1}\right)\left(\frac{\partial^2 u_{a2}}{\partial s_1 \partial s_2}\right)}{D} = \frac{\partial f_1(\alpha)}{\partial \alpha_1} b_2'(s_1;\alpha_1)$$

Since $D > 0$ and $\left(-\frac{\partial^2 u_{a2}}{\partial s_2^2}\right) > 0$, it follows by (A.5) and (A.6) that:

(A.8) $$sign\left(\frac{\partial f_1(\alpha)}{\partial \alpha_1}\right) \neq sign\left(\frac{\partial u_{a2}}{\partial s_1}\right)$$

This proves (4.10). It also follows that

(A.9) $$sign\left(\frac{\partial f_2(\alpha)}{\partial \alpha_1}\right) \neq sign\left(\frac{\partial u_{a2}}{\partial s_1}\right)$$ if and only if $b_2'(s_1;\alpha_1) > 0$.

To get to (4.11), apply the results of proposition 3.1 for smooth monotonic carrot and stick games.

For a stick game for player 1, $$-sign\left[\frac{\partial f_1(\alpha)}{\partial \alpha_1}\right] = sign\left[\frac{\partial u_{a2}}{\partial s_1}\right] = -sign\left[\frac{\partial u_{a1}}{\partial s_2} b_2'(s_1;\alpha_1)\right].$$

The last term equals $-sign\left[\frac{\partial u_{\alpha 1}}{\partial s_2}\right]$ if and only if $b_2'(s_1;\alpha_1) > 0$. This proves result (4.11) for stick games. The result for carrot games is proved in the same way.

Proof of proposition 7.3

Let (f_1^*, f_2^*) be given such that $(f_1(\alpha), f_2(\alpha))$ is a Nash equilibrium in $G(\alpha)$ for each $\alpha \in S_{p1} \times S_{p2}$. Let (s_1^*, s_2^*) denote the unique Nash equilibrium in $G(\alpha^*)$. Assumption (4.3) states that $u_{p1}(\alpha_1, \alpha_2, f_1, f_2) \equiv u_1(s_1, s_2)$, and thus

(A.10) $$\frac{\partial u_{p1}(\alpha_1^*, \alpha_2^*, f_1^*, f_2^*)}{\partial \alpha_1} = \frac{\partial u_1(s_1^*, s_2^*)}{\partial s_1}\frac{\partial f_1^*(\alpha^*)}{\partial \alpha_1} + \frac{\partial u_1(s_1^*, s_2^*)}{\partial s_2}\frac{\partial f_2(\alpha^*)}{\partial \alpha_1}$$

First it is shown that

if $\alpha_1^* \geq 0$ then $\frac{\partial u_1(s_1^*, s_2^*)}{\partial s_1}\frac{\partial f_1(\alpha^*)}{\partial \alpha_1} \leq 0$, and

if $\alpha_1^* \le 0$ then $\frac{\partial u_1(s_1^*, s_2^*)}{\partial s_1} \frac{\partial f_1(\alpha^*)}{\partial \alpha_1} \ge 0$.

Let $s_1' \equiv b_1(s_2^*; 0)$. By strict concavity of $u_1(s_1, s_2)$ for each given s_2, then $\frac{\partial u_1(s_1', s_2^*)}{\partial s_1} = 0$, and $\frac{\partial u_1(s_1, s_2^*)}{\partial s_1} < 0$ for all $s_1 > s_1'$, and $\frac{\partial u_1(s_1, s_2^*)}{\partial s_1} > 0$ for all $s_1 < s_1'$.

Statements (4.10) and (4.12) of lemma 7.3 imply that

$$sign\left[\frac{\partial b_1(s_2^*; \alpha_1)}{\partial \alpha_1}\right] = sign\left[\frac{\partial f_1(\alpha)}{\partial \alpha_1}\right] \ne 0 \text{ for all } \alpha.$$

Now if $\alpha_1^* \ge 0$ and $\frac{\partial f_1(\alpha^*)}{\partial \alpha_1} < 0$ then using the facts of the previous two sentences, $s_1^* \le s_1'$, and thus $\frac{\partial u_1(s_1^*, s_2^*)}{\partial s_1} \ge 0$. If $\alpha_1^* \ge 0$ and $\frac{\partial f_1(\alpha^*)}{\partial \alpha_1} > 0$ then by a similar reasoning, $s_1^* \ge s_1'$, and thus $\frac{\partial u_1(s_1^*, s_2^*)}{\partial s_1} \le 0$.

Thus, it has been shown if $\alpha_1^* \ge 0$ then $\frac{\partial f_1(\alpha^*)}{\partial \alpha_1} \frac{\partial u_1(s_1, s_2^*)}{\partial s_1} \le 0$.

For $\alpha_1^* \le 0$ similar reasoning has shown that $\frac{\partial f_1(\alpha^*)}{\partial \alpha_1} \frac{\partial u_1(s_1, s_2^*)}{\partial s_1} \ge 0$.

Statement (4.11) of lemma 7.3 implies directly that the second part of the right hand side of (A.10) $\frac{\partial u_1(s_1^*, s_2^*)}{\partial s_2} \frac{\partial f_2(\alpha^*)}{\partial \alpha_1} < 0$ if *G* is a carrot game for player 1 while $\frac{\partial u_1(s_1^*, s_2^*)}{\partial s_2} \frac{\partial f_2(\alpha^*)}{\partial \alpha_1} > 0$ if *G* is a stick game for player 1.

Thus, if $\alpha_1^* \ge 0$ and *G* is a carrot game for player 1, then $\frac{\partial u_{p1}(\alpha_1^*, \alpha_2^*, f_1^*, f_2^*)}{\partial \alpha_1} < 0$.

Alternatively, if $\alpha_1^* \le 0$ and *G* is a stick game for player 1, then $\frac{\partial u_{p1}(\alpha_1^*, \alpha_2^*, f_1^*, f_2^*)}{\partial \alpha_1} > 0$. But then principal 1 can increase their utility in each case by changing α_1^* marginally in the direction of zero.

Proof of proposition 7.4

Let G be a carrot game for player 1. Then by proposition 7.3 the equilibrium incentives $\alpha^* = (\alpha_1^*, \alpha_2^*)$ must satisfy $\alpha_1^* < 0$. Let (s_1^*, s_2^*) be the equilibrium in $G(\alpha^*)$, let (s_1, s_2) be the equilibrium in $G(0)$ (the original game), and let (s_1', s_2') be the equilibrium in $G(\alpha_1^*, 0)$ (the game when principal 1 chooses equilibrium incentives and principal 2 chooses incentives for the original game).

Statements (4.10) and (4.12) of lemma 7.3 and $\alpha_1^* < 0$ then $s_1' < s_1$ if and only if $\frac{\partial u_2}{\partial s_1} < 0$. Hence, $u_2(s_1, s_2) < u_2(s_1', s_2) \leq u_2(s_1', s_2') \leq u_2(s_1^*, s_2^*)$ where the first inequality follows from the monotonic externalities and the second inequality follows from the fact that s_2 is available but s_2' is chosen in $G(\alpha_1^*, 0)$ and the last inequality follows since α_2^* is chosen and $\alpha_2 = 0$ is available.

Proof of proposition 7.5

By proposition 7.2, it follows that G can be a stick game for both players only if the best response functions of both players have the same slope. If they are downward sloping then the externalities must move in the same direction. If they are upward sloping, then externalities must move in opposite directions.

Examine the case where reaction functions are downward sloping and externalities are negative. That is, $b_1'(s_2) < 0$, $b_2'(s_1) < 0$, $\frac{\partial u_1}{\partial s_2} < 0$, and $\frac{\partial u_2}{\partial s_1} < 0$. The other cases can be shown by the same type of argument.

By proposition 7.3, the equilibrium $\alpha_1^* > 0$ and $\alpha_2^* > 0$. Let (s_1, s_2) be the equilibrium of $G(0,0) = G$ and let (s_1^*, s_2^*) be the equilibrium of $G(\alpha_1^*, \alpha_2^*)$. It is shown that $s_1^* \leq s_1$ implies $s_2^* > s_2$, and $s_2^* \leq s_2$ implies $s_1^* > s_1$.

If $s_1^* \leq s_1$, then $s_2^* = b_2(s_1^*; \alpha_2^*) > b_2(s_1^*; 0) \geq b_2(s_1; 0) = s_2$, where the equalities come from definitions. The strict inequality follows from applying (4.12) for changes in α_2 to obtain $\frac{\partial b_2(s_1; \alpha_2)}{\partial \alpha_2} > 0$. The weak inequality follows from $b_2'(s_1) < 0$ and $s_1^* \leq s_1$.

If $s_2^* \leq s_2$, then $s_1^* = b_1(s_2^*; \alpha_1^*) > b_1(s_2^*; 0) \geq b_1(s_2; 0) = s_1$, by the same reasoning given in the previous paragraph.

From this it can be concluded that either $s_2^* > s_2$ *or* $s_1^* > s_1$.

If $s_2^* > s_2$, then player 1 is worse off since $u_1(s_1^*, s_2^*) < u_1(s_1^*, s_2) \leq u_1(s_1, s_2)$, where the strict inequality follows from $s_2^* > s_2$ and $\frac{\partial u_2}{\partial s_1} < 0$, and the weak inequality follows from utility maximisation of player 1 in $G(0)$.

If $s_1^* > s_1$, then player 2 is worse off by the same type of argument.

Proof of corollary 7.3

Since G is a stick game for each player, then by proposition 7.5 at least one player is worse off in Γ. Since the players are symmetric it follows that they will be treated symmetrically in a symmetric equilibrium of Γ and thus both are worse off than in G.

Proof of proposition 7.6

Suppose the game is Stackelberg solvable and let $(\bar{v}_1, \bar{v}_2)$ be defined by (s_1, s_2). Conditions (2.1) and (2.2) are not satisfied for player 1 since by (5.1) there cannot be a pair $(s_1', b_2(s_1'))$ satisfying $u_1(s_1', b_2(s_1')) > \bar{v}_1$. The same argument holds for player 2 by (5.2).

References

Basu, K. 1995, 'Stackelberg equilibrium in oligopoly: An explanation based on managerial incentives', *Economics Letters*, vol. 49, no. 4, pp. 459–64.

Basu, K., Ghosh, A. and Ray, T. 1997, 'The babu and boxwallah: Managerial incentives and government intervention in a developing economy', *Review of Development Economics*, vol. 1, no. 1, pp. 71–90.

Baye, M.R., Crocker, K.J. and Ju, J. 1996, 'Divisionalisation, franchising, and divestiture incentives in oligopoly', *American Economic Review*, vol. 86, no. 1, pp. 223–36.

Bonanno G. and Vickers, J. 1988, 'Vertical Separation', *Journal of Industrial Economics*, vol. 36, no. 3, pp. 257–65.

Bulow J., Geanakopolos, J. and Klemperer, P. 1985, 'Multimarket oligopoly: Strategic substitutes and complements', *Journal of Political Economy*, vol. 93, no. 3, pp. 488–511.

d'Aspremont, C. and Gerard-Varet, L.A. 1980, 'Stackelberg-solvable games and pre-play communication', *Journal of Economic Theory*, vol. 23, no. 2, pp. 201–17.

Dowrick, S. 1986, 'Von Stackelberg and Cournot duopoly: Choosing roles', *Rand Journal of Economics*, vol. 17, no. 2, pp. 251–60.

Fershtman, C. and Judd, K.L. 1987, 'Equilibrium incentives in oligopoly', *American Economic Review*, vol. 77, no. 5, pp. 927–40.

Fulks, W. 1978, *Advanced Calculus*, 3rd edn, John Wiley & Sons, New York.

Gal-Or, E. 1985, 'First mover and second mover advantages', *International Economic Review*, vol. 26, no. 3, pp. 649–53.

Hamilton, J. H. and Slutsky, S. 1990, 'Endogenous timing in duopoly games: Stackelberg or Cournot equilibria', *Games and Economic Behaviour*, vol. 2, no. 1, pp. 29–46.

Klemperer P. and Meyer, M. 1986, 'Price competition vs. quantity competition: The role of uncertainty', *Rand Journal of Economics*, vol. 17, no. 4, pp. 404–15.

Rey, P. and Stiglitz, J. 1995, 'The role of exclusive territories in producers' competition', *Rand Journal of Economics*, vol. 26, no. 3, pp. 431–51.

Shy, 0. 1995, *Industrial Organization*, MIT Press, Cambridge, Massachusetts.

Sklivas, S.D. 1987, 'The strategic choice of managerial incentives', *Rand Journal of Economics*, vol. 18, no. 3, pp. 452–8.

Vickers, J. 1985, 'Delegation and the theory of the firm', *Economic Journal*, vol. 95, Supplement, pp. 138–47.

Von Neumann J. and Morgenstern, O. 1953, *Theory of Games and Economic Behaviour*, Princeton University Press, Princeton.

8 Endogenous co-leadership when demand is unknown

*Midori Hirokawa and Dan Sasaki**

Consider an oligopolistic industry where demand uncertainty resolves after at least one firm has engaged in production. Those firms who produce first behave as simultaneous leaders (co-leaders), whilst those who produce after demand becomes observable will be followers. Each follower simply plays an individual best response. On the other hand, each co-leader takes only the production quantities of other co-leaders as given. Using simple linear demand, this paper establishes that the number of co-leaders monotonically decreases in the magnitude of demand uncertainty relative to the expected level of demand. It also finds that, with demand uncertainty and the possibility of Stackelberg behaviour, whether the excess entry theorem applies depends upon the number of existing followers.

8.1 Introduction

When a new product is introduced to a market, the following is often observed. Firstly, 'products tend to be introduced nearly simultaneously by multiple firms', as Aron and Lazear (1990) point out. Secondly, the level of demand in a new market is difficult to infer *a priori*, but is revealed after one, or possibly more than one, firm starts selling the new product.

Potential firms face two problems. One is whether to enter or not. The other is, if it enters, the timing of production — whether to lead or to follow. This paper considers the latter first, and then proceeds to the former.

This paper first examines how many firms will undertake leadership simultaneously (henceforth called co-leaderships) *in equilibrium*, when the total number of entrants n is given. It finds that the number of co-leaders monotonically decreases in the

* This paper was financed in part by a research grant from the Faculty of Economics and Commerce at the University of Melbourne.

magnitude of demand uncertainty, but that at least one firm always becomes a leader.

Second, a similar observation can be made as to the *welfare-optimal* number of co-leaders, except that its range may differ from that of the *equilibrium* number of co-leaders. In linear oligopoly, the former runs only between 1 and $n/2$, as opposed to the latter running 1 through n. This contrast reflects the fact that it is socially desirable to have some followers who serve to absorb an unexpected demand shock. It also leads to a straightforward policy implication that, when demand uncertainty is so little that too many firms want co-leadership, the public planner should encourage some firms to wait to be followers instead.

This paper then proceeds to endogenise not only entry timing, but also entry *per se*, allowing n to vary. Intuitively, the expected demand level tends to determine the aggregate *ex ante* profitability of the market and thus the equilibrium *total* number of firms that can profitably enter the market, whilst the uncertainty of demand tends to determine the *behaviour* of firms and thereby the equilibrium number of *co-leaders* and *followers*.

Finally, this paper investigates the relation between the so-called 'excess entry theorem' and the current results based upon endogenous production timing. It theoretically re-examines the theorem in the context of the endogenous co-leadership model, by comparing the *profitability* of an additional entry measured by the profit of the entrant, and the *welfare increment* made possible by the entry.

This paper starts with a simple model laid out in section 8.2 where an exogenously given number of firms choose either to be leaders or to be followers, through their production timing decisions. Then, section 8.3 presents a two-stage model with endogenous entry, which requires a material cost in order to set up a firm, and is followed by the simple timing game analysed in section 8.2. Sections 8.2 and 8.3 conduct an equilibrium comparative statics analysis followed by a welfare comparative statics analysis. Finally, section 8.4 highlights the main qualitative findings and also locates this paper in the context of existing literature.

8.2 The basic model

There are $n \geq 3$ firms in the industry. They are *a priori* identical. There are two possible production timings. At the beginning of the game, each firm decides whether to produce immediately, or wait. Each firm can produce in only one of these two production timings. If a firm waits, then its quantity decision is also deferred. For simplicity, the marginal cost of supply is assumed to be a constant c per unit, and no time discounting is introduced.

An oligopoly market takes place after the second production period with the inverse demand function:

$$P = A - \sum_{j=1}^{n} q_j$$

where q_j $(j = 1,\ldots,n)$ is the production quantity of firm *j*. The intercept *A* is unobservable to the firms *until at least one firm starts production*, though its prior cumulative distribution $F(a)$ is commonly known to all *n* firms, with strictly positive finite mean $E[A]$ and variance $Var[A]$. This automatically implies that, as long as at least one firm produces immediately, the demand intercept *A* becomes observable to those who choose to produce late. Otherwise, if no firm produces early, then later all *n* firms will have to produce without knowing the state of demand *A*.

Therefore, by becoming a second-mover, a firm can make use of the observed state of demand to optimise its production quantity. On the other hand, if a firm produces before others, then the firm is entitled to a Stackelberg leadership or one of the co-leaderships.

Thus at the beginning of the game, firms face the trade-off between the strategic advantage of commitment and the ability to adjust to demand realisation.

The profit maximisation problem for each firm *i* $(i = 1,,\ldots n)$ can be solved as follows, depending upon the production timing decisions of firms. Let $\bar{q}_i$ denote the quantity supplied by firm *i* when the firm is a first producer $(i = 1,\ldots,m)$, and $q_i[A]$ the quantity supplied by firm *i* when it is a second producer $(i = m+1,\ldots,n)$, respectively. Note that only the quantity produced by a second producer can be made contingent upon the state *A*. To ensure the individually rational participation of all *n* firms, it is assumed that $0 < c < E[A]$, and

$$\inf[A] \geq \frac{nE[A] + c}{n+1} \qquad [8.1]$$

so that the realised prices are always positive (see appendix 8A).

(i) If *m*=*n*, then all *n* firms simultaneously choose their respective quantities according to

$$\max_{\bar{q}_i} E\left[\left(A - c - \sum_{j=1}^{n} \bar{q}_j\right)\bar{q}_i\right]$$

(ii) If m ($m=1,\ldots,n-1$) firms produce immediately, then the game is solved backward. The late movers $i=m+1,\ldots,n$ simultaneously solve

$$\max_{q_i[A]}\left[\left(A-c-\sum_{j=1}^{m}\bar{q}_j-\sum_{j=m+1}^{n}q_j[A]\right)q_i[A]\right]$$

Let $q_i[A,\bar{q}]$ denote the solution for this maximisation, given $\bar{q}=(\bar{q}_1,\ldots,\bar{q}_m)$. At the beginning of the game, the early producers $i=1,\ldots,m$ solve

$$\max_{\bar{q}_i} E\left[\left(A-c-\sum_{j=1}^{m}\bar{q}_j-\sum_{j=m+1}^{n}q_j[A,\bar{q}]\right)\bar{q}_i\right]$$

(iii) If $m=0$, no firm produces immediately, and therefore demand uncertainty does not resolve after the first production timing. Thereby in the second production period, all n firms produce simultaneously under demand uncertainty, entailing the same outcome as when $m=n$.

Equilibrium outcomes

The game specified in the above has the following pure-strategy equilibria.

Proposition 8.1

The equilibrium number of co-leaders m^* is given by

$$m^*=\begin{cases}1 & \text{if} \quad \dfrac{Var[A]}{(E[A]-c)^2}\geq v^*[2;n],\\ 2,\ldots,n-1 & \text{if} \quad v^*[m^*+1;n]\leq\dfrac{Var[A]}{(E[A]-c)^2}\leq v^*[m^*;n],\\ n & \text{if} \quad \dfrac{Var[A]}{(E[A]-c)^2}\leq v^*[n;n],\end{cases}$$

where $v^*[m;n]=\dfrac{(n-m+2)^2}{(m+1)^2(n-m+1)}-\dfrac{1}{m^2}$. Note that when $\dfrac{Var[A]}{(E[A]-c)^2}=v^*[m;n]$ (m=2,…,n), both $m^*=m-1$ and $m^*=m$ are equilibria.

The direct implications of this proposition can be summarised by the following.

Corollaries

- At least one firm will always take a lead.

- The equilibrium number of co-leaders monotonically increases in n and decreases in $\frac{Var[A]}{(E[A]-c)^2}$.
- For any m=1,…,n, there exists a prior demand distribution $F(a)$ such that exactly m firms undertake co-leadership in a pure strategy equilibrium.
- As demand uncertainty vanishes (ie $Var[A]\downarrow 0$), all n firms become co-leaders.[1]
- Given $\frac{Var[A]}{(E[A]-c)^2}$, the equilibrium number of followers $n-m^*$ also monotonically increases in n.[2]

This is proved in appendix 8A.

Economic intuition

The riskier it is to precommit with a quantity, the less firms undertake co-leadership. Note also that the degree of demand uncertainty is measured against the demand intercept *in excess of the production costs*. This means that, in a market where the mark-up is relatively low, the effect of demand uncertainty is magnified. Therefore, when the (expected) mark-up is low, few firms are willing to undertake co-leadership.

Recall that it is possible, when demand uncertainty is sufficiently low, for all n firms to be co-leaders, whilst it never occurs in any pure strategy equilibrium that all n firms decide to wait, even when demand uncertainty is high. Therefore, Stackelberg outcomes are likely in an industry where demand is *a priori* highly uncertain, and/or the marginal production costs are substantial relative to the price.

Finally, the last corollary implies that the simultaneous Cournot-Nash outcome, that all n firms act as co-leaders, becomes less likely as the number of firms n increases. In other words, Stackelberg outcomes with at least one leader and at least one follower become likely when there are many firms.

---

1 Some of the preceding contributions adopt a slightly different modelling technique that entails substantially different equilibrium results. For example, Matsumura (1995) assumes that q_i (i=1,…,n) should be determined simultaneously and therefore mutually independently irrespective of leader/followerships, so that followers' quantities cannot be made contingent upon leaders' quantities. In such a model, a firm may volunteer a follower's role even without demand uncertainty.

2 The second and last corollaries imply that, for any given demand distribution F(.) and marginal cost c satisfying inequality [8.1], if the equilibrium number of co-leaders is m^* when there are n firms in total, then the equilibrium number of co-leaders is either m^* or m^*+1 when the total number of firms increases to n+1.

Social welfare

In quantity competition, Stackelberg outcomes are generally welfare superior to Cournot-Nash outcomes. This is because Stackelberg behaviour leads to a larger expected total supply than the simultaneous Cournot-Nash outcome, and this reduces the welfare loss arising from oligopolistic market power. In the uncertain demand model presented in this paper, Stackelberg leader-follower behaviour enhances welfare through an additional effect — the followers can accommodate the fluctuant demand.

Proposition 8.2

The expected social welfare, defined as the sum of the profits of the *n* firms and the consumer surplus, reaches its maximum when $\overline{m}$ firms undertake co-leaderships, where

$$\overline{m} = \begin{cases} 1 & \text{if} \quad \dfrac{Var[A]}{(E[A]-c)^2} \geq \omega[2;n], \\ 2,\ldots,\text{int}\left[\dfrac{n}{2}\right] & \text{if} \quad \omega[\overline{m}+1;n] \leq \dfrac{Var[A]}{(E[A]-c)^2} \leq \omega[\overline{m};n], \end{cases}$$

where $\omega[m;n] = \dfrac{(n-m+2)^{-2}m^{-2} - (n-m+1)^{-2}(m+1)^{-2}}{(n-m+1)^{-2} - (n-m+2)^{-2}}$, and $\text{int}[z]$ denotes the largest integer not exceeding z. Note that, when $\dfrac{Var[A]}{(E[A]-c)^2} = \omega[m;n]$ (m=2,…, $\text{int}\left[\dfrac{n}{2}\right]$), both $\overline{m} = m-1$ and $\overline{m} = m$ are welfare maximal.

The implications of this proposition are summarised as follows.

Corollaries

- To optimise expected social welfare, at least one, and no more than $\dfrac{n}{2}$ firms should take a lead.
- The optimal number of co-leaders $\overline{m}$ monotonically decreases in $\dfrac{Var[A]}{(E[A]-c)^2}$.
- For any integer $m \in \left[1, \dfrac{n}{2}\right]$ there exists a prior demand distribution $F(a)$ such that the co-leadership of exactly m firms is socially optimal.

This is proved in appendix 8B.

Economic Implication

In conjunction with proposition 8.1, the equilibrium number of co-leaders and the welfare-maximal number of co-leaders monotonically decrease in the relative degree of demand uncertainty. As demand uncertainty grows, however, the former runs from 1 to *n*, whilst the latter ranges only between 1 and *n*/2.

In fact, there are two countereffects for maximising expected welfare. One is a deterministic effect — the total supply quantity increases in the number of co-leaders up to *n*/2, which leads to welfare enhancement. The other is a stochastic effect — the presence of a certain number of followers is helpful in absorbing demand fluctuation. It is intuitively clear that the latter effect grows as demand becomes increasingly uncertain. The expected welfare is optimised when the balance is struck between these two countereffects. Note, however, that only the latter effect is fully taken into consideration when firms follow their individual incentives. This is the reason why the *equilibrium* number of co-leaders spreads beyond the range of the welfare-maximal number of co-leaders.

This also leads to a policy implication. Especially when demand uncertainty is insignificant and thus few firms volunteer to wait, a policy measure encouraging more firms to wait and less firms to undertake co-leadership will enhance welfare.

8.3 Endogenous entry

Until now, the number of firms *n* in the industry has been taken as exogenously given. In this section, *n* is endogenised in order to derive truly endogenous Stackelberg leader follower relations, and furthermore to relate the results to the so-called excess entry theorem.[3]

Consider the following two-stage game. There are potentially an infinite number of entrepreneurs. All of them are *a priori* identical. The first stage is for entry decisions. Each entrepreneur independently decides whether to set up a firm or not. The cost required in order to establish a firm is a positive constant γ. Then, the second stage is the same game as in section 8.2.

[3] There are two versions of the excess entry theorem: the first-best excess entry theorem which assumes that firms' market behaviour posterior to their entry is regulable, and the second-best version which assumes that firms' post-entry behaviour is not regulable. The discussion here is of the latter.

To maintain consistency with the previous model, assume that the set-up cost of a firm γ is sufficiently low. That is

$$\gamma \leq \frac{(E[A]-c)^2}{18}$$

so that at least three firms can profitably enter (see equation [8.5] in appendix 8A).

Endogenous equilibrium outcomes

In this two-stage game with endogenous entry, Cournot-Nash and Stackelberg outcomes occur according to the following criteria.

Proposition 8.3

A Cournot-Nash outcome (ie the absence of followers) can arise as a pure strategy subgame perfect equilibrium if and only if

$$\frac{Var[A]}{(E[A]-c)^2} \leq 4\left(\text{int}\left[\frac{E[A]-c}{\sqrt{\gamma}}\right]\right)^{-2} - \left(\text{int}\left[\frac{E[A]-c}{\sqrt{\gamma}}\right]-1\right)^{-2} \qquad [8.2]$$

This is proved in appendix 8C. Intuitively, when *n* is sufficiently large

$$\left(\text{int}\left[\frac{E[A]-c}{\sqrt{\gamma}}\right]\right)^{-2} \approx \left(\text{int}\left[\frac{E[A]-c}{\sqrt{\gamma}}\right]-1\right)^{-2} \approx \frac{\gamma}{(E[A]-c)^2}$$

The inequality [8.2] roughly becomes

$$Var[A] \leq 3\gamma$$

Namely, endogenous equilibrium outcomes hinge upon the relation between the *absolute* magnitude of demand uncertainty and the entry cost. Simultaneous Cournot-Nash outcomes with all leaders and no followers will arise when demand uncertainty is low relative to the entry cost (inequality 8.2), whereas Stackelberg leader-follower outcomes occur when demand is substantially uncertain (not inequality 8.2).

Overall, the relation between $Var[A]$ and γ determines the nature of equilibria, either Cournot-Nash or Stackelberg, whilst the market size $E[A]-c$ contributes to the number of firms that can profitably enter.

‘Excess entry theorem’ revisited

Irrespective of the size of the set-up cost γ, whether or not the endogenous entry of an additional firm is socially desirable depends upon the dominance relation between the *profitability* of entry and its *welfare enhancement*. In the following, the interlink between so-called ‘excess entry’ and endogenous co-leadership is investigated.

According to the excess entry theorem, when an additional firm enters into an oligopolistic market, it takes into account only its *own* profits, not the negative externality of crowding, that is, the decline in the profits of other competitors caused by the entry of the firm. The theorem asserts that, as a result, whenever there is an entry cost, more firms tend to enter than is socially efficient.

For technical simplicity, it is assumed that the choice of each firm whether to become a co-leader or a follower is unaffected by the entry decisions of other firms.[4] More precisely, in computing the profitability as well as the welfare impact of the entry, the entry profile $\{m,n\}$ (the first coordinate is the total number of co-leaders, and the second the total number of firms) is compared either with $\{m+1,n+1\}$ in the case of the entry of an extra co-leader, or with $\{m,n+1\}$ in the case of the entry of an extra follower.

Lemma 8.1 (Leader)

The entry of an additional leader is always more profitable than it is welfare enhancing.

Lemma 8.2 (Follower)

The entry of an additional follower is more profitable than it is welfare enhancing, except when it enters as *the only* follower.

These two lemmas are proved in appendix 8D.

Clearly, the profit of each co-leader $\Pi^L[m;n]$ decreases in the number of co-leaders m and the profit of each follower $\Pi^F[m;n]$ also decreases in the number of followers $n-m$. Co-leaders enter as long as $\Pi^L[m;n]\geq\gamma$, and the entry of followers continues as long as $\Pi^F[m;n]\geq\gamma$. Hence, these two lemmas imply the following.

[4] In other words, conceptualise potentially infinitely many co-leader-entrepreneurs and also infinitely many follower-entrepreneurs, making entry decisions simultaneously.

Proposition 8.4

The excess entry theorem applies, *except* when a follower enters as *the only* follower.

This result can be intuitively interpreted as follows. When there is no follower yet, it is socially desirable for a new follower to enter, so that the follower can absorb any unpredicted demand fluctuation. When there is already one follower, this effect is no longer large enough to justify the entry beyond its profitability. On the other hand, the social effect of the entry of a new co-leader, entailing an increment in the expected aggregate supply, is consistently smaller than the private profitability of the entry.

8.4 Concluding discussion

Notwithstanding the overwhelming numbers of game-theoretic attempts to explain endogenous Stackelberg leader-follower relations (eg Hamilton and Slutsky 1990; Albæk 1990; Robson 1990) and of those to extend the analysis to general *n*-firm oligopoly (eg Anderson and Engers 1992; Nishijima 1995), only a small fraction of them (eg Matsuura 1995) have been devoted explicitly to the issue of strategic leadership undertaken simultaneously by more than one firm.

On the other hand, the incentives of firms to become followers can be closely related to demand uncertainty. A few recent contributions have attempted to imbed endogenous leader-follower behaviour in a model of oligopoly with unknown demand. Sadanand and Sadanand (1996) constructs a model where demand uncertainty is resolved by the passage of time. On the other hand, Hirokawa (1998) models demand uncertainty that resolves after *the production of at least one firm*.

The endeavour of this paper has been to consolidate these two streams of studies to establish a qualitative linkage between demand uncertainty and the incentives for simultaneous multiple leadership, referred to here as co-leadership. The following propositions have been shown:

- Among *n* firms, at least one, and at most all *n*, may volunteer to be co-leaders who take a lead *before* the uncertainty clears. The equilibrium number of co-leaders decreases monotonically in the degree of demand uncertainty. Stackelberg outcomes can thereby arise as pure strategy subgame perfect equilibria whenever the demand uncertainty is large enough to prevent the simultaneous co-leadership of all *n* firms.
- The socially optimal number of co-leaders also decreases monotonically in the degree of demand uncertainty, running from $n/2$ down to 1.

- Therefore, the policy implication regarding co-leadership becomes such that, when demand uncertainty is insufficient to sustain an equilibrium where enough firms wait to be followers, it is socially desirable to encourage at least some of them to wait.
- When entry decisions are strategic and therefore the number of competing oligopolists is endogenous, the expected level of demand determines the equilibrium number of oligopolists, whereas the magnitude of demand uncertainty determines the form of competition, that is, whether the production timing of firms is simultaneous (Cournot-Nash) or leader-follower (Stackelberg), and in the latter, how many of them become leaders and how many followers.
- On the excess entry theorem, Mankiw and Whinston (1986) and Suzumura and Kiyono (1987) show that socially excessive firm entry can occur in an unregulated Cournot (or Cournot-like) oligopoly. Okuno-Fujiwara and Suzumura (1993) show that this theorem holds in the presence of strategic R&D commitments. This paper has shown that the theorem can be applied to Stackelberg equilibrium with demand uncertainty, *except* when a follower enters as the only follower.

Appendix 8A Equilibrium comparative statics

Start from an implicit assumption that the optimal supply quantities for firms and the resulting market prices are always strictly positive. This maintains their linear-quadratic objective (profit) functions, which makes $E[A]$ and $Var[A]$ sufficient statistics for solving the optimisation problems of firms. This enables the explicit-form solutions for market behaviour of firms to be derived without narrowly specifying the prior demand distribution $F(\cdot)$. Then the positivity conditions for quantities and prices will be reexamined.

When firms $1,\ldots,m$ $(1 \le m \le n)$ produce immediately and firms $m+1,\ldots,n$ wait, a follower $i = m+1,\ldots,n$ produces the quantity

$$q_i[A] = \frac{1}{n-m+1}\left(A - c - \frac{m}{m+1}(E[A]-c)\right) \qquad [8.3]$$

whilst a co-leader $i = 1,\ldots,m$ produces

$$\bar{q}_i = \frac{E[A]-c}{m+1} \qquad [8.4]$$

entailing the price

$$p = c + \frac{1}{n-m+1}\left(A - c - \frac{m}{m+1}\left(E[A]-c\right)\right)$$

Profits are thereby

$$\pi^L = \frac{E[A]-c}{(m+1)(n-m+1)}\left(A - c - \frac{m}{m+1}\left(E[A]-c\right)\right)$$

for a co-leader, and

$$\pi^F = \left[\frac{1}{(n-m+1)}\left(A - c - \frac{m}{m+1}\left(E[A]-c\right)\right)\right]^2$$

for a follower. Their expectations are

$$\Pi^L[m;n] = E[\pi^L] = \frac{1}{(n-m+1)}\left(\frac{E[A]-c}{m+1}\right)^2 \qquad m = 1,\ldots,n \qquad [8.5]$$

and

$$\Pi^F[m;n] = E[\pi^F] = \frac{1}{(n-m+1)^2}\left(Var[A] + \left(\frac{E[A]-c}{m+1}\right)^2\right) \quad m = 1,\ldots,n-1 \quad [8.6]$$

respectively. Otherwise, if all *n* firms wait, the outcome will be identical to the case where all *n* firms produce immediately, namely,

$$\Pi^F[0;n] = \Pi^L[n;n] = \left(\frac{E[A]-c}{n+1}\right)^2$$

Note that, by [8.3], the positivity of the quantity $q_i[A]$ supplied by a follower is guaranteed by

$$\inf[A] \geq \frac{mE[A]+c}{m+1} \qquad [8.7]$$

which also suffices for the positivity of the price. Meanwhile, the quantity of a co-leader is positive by the assumption that $E[A] > c > 0$ (see equation [8.4]). Inequality [8.1] is sufficient for [8.7] for any $m = 0,\ldots,n$.

Hereby inequality condition [8.1] should be imposed as an assumption throughout this paper (see section 8.2). Without this assumption, the first-order conditions for the maximisation problems of firms would become non-linear, so that the closed-

form solutions would be difficult to obtain without further distributional assumptions on $F(a)$ (see Hirokawa 1998).

Hence:

- When firms 2,…,*n* decide to wait, firm 1 should produce immediately if and only if $\Pi^L[1;n] \geq \Pi^F[0;n]$. That is

$$\frac{(E[A]-c)^2}{4n} \geq \frac{(E[A]-c)^2}{(n+1)^2} \qquad [8.8]$$

- When firms $1,\ldots,m-1$ $(2 \leq m \leq n)$ decide to produce immediately and firms $m+1,\ldots,n$ decide to wait, firm *m* should also produce immediately if and only if $\Pi^L[m;n] \geq \Pi^F[m-1;n]$. That is

$$\frac{Var[A]}{(E[A]-c)^2} \leq \frac{(n-m+2)^2}{(m+1)^2(n-m+1)} - \frac{1}{m^2} \qquad [8.9]$$

- When firms $1,\ldots,m$ $(1 \leq m \leq n-1)$ decide to produce immediately and firms $m+1,\ldots,n$ decide to wait, firm *m*+1 should wait if and only if $\Pi^F[m;n] \geq \Pi^L[m+1;n]$. That is

$$\frac{Var[A]}{(E[A]-c)^2} \geq \frac{(n-m+1)^2}{(m+2)^2(n-m)} - \frac{1}{(m+1)^2} \qquad [8.10]$$

By the assumption that $n \geq 3$, condition [8.8] is always satisfied. Hence, $m = 0$ can never be an equilibrium. Exactly m^* firms co-lead in equilibrium if and only if [8.9] holds for $m = 2,\ldots,m^*-1$ and [8.10] holds for $m = m^*,\ldots,n$.

The remaining step is to show that the right-hand side of inequality [8.9], henceforth denoted by $v^*[m;n]$, is larger than the right-hand side of inequality [8.10], denoted by $v^*[m+1;n]$. For this purpose, it suffices to show that the threshold function

$$v^*[m;n] = \frac{(n-m+2)^2}{(m+1)^2(n-m+1)} - \frac{1}{m^2}$$

is decreasing in *m*. Its derivative is calculated as

$$\begin{aligned}
\frac{\partial v^*[m;n]}{\partial m} &= \frac{2}{m^3} - \frac{1}{(m+1)^2}\left(1 - \frac{1}{(n-m+1)^2}\right) - \frac{2(n-m+2)^2}{(m+1)^3(n-m+1)} \\
&< \frac{2}{m^3} - \frac{2}{(m+1)^3}\left((n-m+1) + 2 + \frac{1}{n-m+1}\right) \\
&< \frac{2}{m^3} - \frac{8}{(m+1)^3}
\end{aligned}$$

The far right hand side is negative whenever $m \geq 2$.

It is hereby shown that $v^*[m;n]$ decreases in *m* over the whole range *m*=1,…, $n-1$.

To show that the equilibrium number of co-leaders m^* increases with the total number of firms *n*, it suffices to show that $v^*[m;n]$ increases in *n*, which is self-evident. To show that the equilibrium number of followers $n-m^*$ increases with *n*, the condition $v^*[m+1;n+1] < v^*[m;n]$ must hold. Since $n \geq m+1$ and $n > 2$,

$$v^*[m;n] - v^*[m+1;n+1] > \frac{3\left((m+2)^2-(m+1)^2\right)}{(m+1)^2(m+2)^2} - \frac{2m+1}{m^2(m+1)^2} = \frac{4m^3-12m-4}{m^2(m+1)^2(m+2)^2} > 0$$

Appendix 8B Welfare comparative statics

With *m* co-leaders (*m*=1,…,*n*), social welfare becomes

$$w = \frac{1}{2}\left((A-c)^2 - \left[\frac{1}{n-m+1}\left(A-c-\frac{m}{m+1}(E[A]-c)\right)\right]^2\right)$$

of which the expectation is

$$\begin{aligned} W[m;n] &= E[w] \\ &= \frac{1}{2}\left(\left[1-\frac{1}{(n-m+1)^2}\right]Var[A] + \left[1-\frac{1}{(n-m+1)^2(m+1)^2}\right](E[A]-c)^2\right) \end{aligned} \quad [8.11]$$

Since the first-order difference

$$W[m+1;n] - W[m;n] = \frac{1}{2}\left(\left[\frac{1}{(n-m+1)^2} - \frac{1}{(n-m)^2}\right]Var[A] \right. \quad [8.12]$$

$$\left. + \left[\frac{1}{(n-m+1)^2(m+1)^2} - \frac{1}{(n-m)^2(m+2)^2}\right](E[A]-c)^2\right) \quad [8.13]$$

is strictly decreasing in *m*, the two necessary conditions for the welfare optimality of the co-leadership of $\overline{m}$ firms ($\overline{m} = 2,\ldots,n-1$)

- $W[\overline{m};n] \geq W[\overline{m}-1;n]$

$$\Leftrightarrow \frac{Var[A]}{(E[A]-c)^2} \leq \frac{(n-\overline{m}+2)^{-2}\overline{m}^{-2} - (n-\overline{m}+1)^{-2}(\overline{m}+1)^{-2}}{(n-\overline{m}+1)^{-2} - (n-\overline{m}+2)^{-2}} \quad [8.14]$$

- $W[\overline{m};n] \geq W[\overline{m}+1;n]$

$$\Leftrightarrow \frac{Var[A]}{(E[A]-c)^2} \geq \frac{(n-\overline{m}+1)^{-2}(\overline{m}+1)^{-2}-(n-\overline{m})^{-2}(\overline{m}+2)^{-2}}{(n-\overline{m})^{-2}-(n-\overline{m}+1)^{-2}} \qquad [8.15]$$

automatically *suffice* for optimality of $W[\overline{m};n]$. Note that the first necessary condition for welfare optimality (inequality [8.14]) is feasible only when $\overline{m} \leq \text{int}\left[\frac{n}{2}\right]$.

Finally, since $W[0;n]=W[n;n]$, m=0 can never be welfare optimal. As to the case of $\overline{m}=1$, the first inequality condition [8.14] is inoperative, so that the second condition [8.15] alone constitutes the necessary and sufficient condition for welfare optimality of the single-firm leadership $\overline{m}=1$.

Appendix 8C Endogenous Cournot-Nash versus Stackelberg

In order for all n firms entering endogenously to become leaders, by [8.5] (see appendix 8A) the entry cost γ must satisfy

$$\Pi^L[n;n]=\left(\frac{E[A]-c}{n+1}\right)^2 \geq \gamma \geq \Pi^L[n+1;n+1]=\left(\frac{E[A]-c}{n+2}\right)^2$$

hence

$$n=\text{int}\left[\frac{E[A]-c}{\sqrt{\gamma}}\right]-1$$

This, in conjunction with the last inequality condition in proposition 8.1, completes the proof of proposition 8.3.

Appendix 8D Relation to excess entry theorem

Proof of lemma 8.1 (Leader)

By [8.5] (see appendix 8A), when there are n+1 entrants out of which m+1 are co-leaders, each of the m+1 co-leaders can expect the following profit from the market

$$\Pi^L[m+1;n+1]=\frac{1}{(n-m+1)}\left(\frac{E[A]-c}{m+2}\right)^2$$

At the same time, the welfare contribution by the last co-leader is

$$W[m+1;n+1]-W[m;n]=\frac{1}{2}\left(\frac{1}{(m+1)^2}-\frac{1}{(m+2)^2}\right)\left(\frac{E[A]-c}{n-m+1}\right)^2$$

(see [8.11] in appendix 8B). Hence $W[m+1;n+1]-W[m;n]\le\Pi^L[m+1;n+1]$ if and only if

$$n\ge m-1+\frac{1}{m+1}+\frac{1}{2(m+1)^2}$$

This inequality is always met as long as $n\ge 3$ and $0\le m\le n$.

Proof of lemma 8.2 (Follower)

By [8.6] (see appendix 8A), when there are $n+1$ entrants out of which m are co-leaders, each of the $n-m+1$ followers expects the profit from the market to be

$$\Pi^F[m;n+1]=\frac{1}{(n-m+2)^2}\left(Var[A]+\left(\frac{E[A]-c}{m+1}\right)^2\right)$$

where m is the number of leaders. By this entry, welfare is enhanced by

$$W[m;n+1]-W[m;n]=\frac{1}{2}\left(\frac{1}{(n-m+1)^2}-\frac{1}{(n-m+2)^2}\right)\left(Var[A]+\left(\frac{E[A]-c}{m+1}\right)^2\right)$$

Thus, $W[m;n+1]-W[m;n]\le\Pi^F[m;n+1]$ if and only if $m\le n-1$.

References

Albæk, S. 1990, 'Stackelberg leadership as a natural solution under cost uncertainty', *Journal of Industrial Economics*, vol. 38, no. 3, pp. 335–47.

Anderson, S.P. and Engers, M. 1992, 'Stackelberg versus Cournot oligopoly equilibrium', *International Journal of Industrial Organization*, vol. 10, no. 1, pp. 127–35.

Aron, D.J. and Lazear, E.P. 1990, 'The industrial organization of markets with differentiated products: The introduction of new products', *American Economic Review*, vol. 80, no. 2, pp. 421–6.

Hamilton, J.H. and Slutsky, S.M. 1990, 'Endogenous timing in duopoly games: Stackelberg or Cournot equilibria', *Games and Economic Behavior*, vol. 2, no. 1, pp. 29–46.

Hirokawa, M. 1998, Endogenous Stackelberg equilibrium under demand uncertainty, unpublished.

Mankiw, N.G. and Whinston, M.D. 1986, 'Free entry and social inefficiency', *Rand Journal of Economics*, vol. 17, no. 1, pp. 48–58.

Matsumura, T. 1995, *How many firms become leaders?*, Discussion Paper no. 367, Institute of Social and Economic Research, Osaka University.

Nishijima, M. 1995, N-person endogenous leader-follower relations, unpublished.

Okuno-Fujiwara, M. and Suzumura, K. 1993, 'Symmetric Cournot oligopoly and economic welfare: A synthesis', *Economic Theory*, vol. 3, no. 1, pp. 43–59.

Robson, A.J. 1990, 'Duopoly with endogenous strategic timing: Stackelberg regained', *International Economic Review*, vol. 31, no. 2, pp. 263–74.

Sadanand, A. and Sadanand, V. 1996, 'Firm scale and endogenous timing of entry: A choice between commitment and flexibility', *Journal of Economic Theory*, vol. 70, no. 2, pp. 516–30.

Suzumura, K. and Kiyono, K. 1987, 'Entry barriers and economic welfare', *Review of Economic Studies*, vol. 54, no. 1, pp. 157–67.

SECTION 3

FINANCIAL CRISES AND BANKING

9 The new era of financial fragility

Ross H. McLeod[*]

9.1 What has changed?

The East Asian experience (McLeod and Garnaut 1998) suggests that there is merit in thinking of balance of payments crises in terms of two extremes, referred to here as 'old-style' and 'new-style' crises.

The old-style crisis originates in the current account. It is typically brought on by the emergence of relatively high inflation or adverse terms of trade shifts creating a gap between the current account deficit and capital inflow. In such circumstances, given a pegged exchange rate and a reluctance on the part of governments to allow the money supply to fall and interest rates to rise, reserves decline continuously,[1] eventually attracting the attention of speculators and those with unhedged exchange rate risks. At this stage, the crisis begins to involve the capital account as well. The decline in reserves then accelerates, eventually forcing a devaluation.[2] If sufficiently large, this puts an end to the problem, providing fiscal and monetary policies are adjusted as necessary in cases where they have been the initial source of disequilibrium. (The successful resolution of the payments disequilibrium may cause other problems, but experience indicates that the anticipated effects of devaluation are often greatly exaggerated.)

New-style crises originate in the capital account rather than the current account. Three features distinguish them from old-style crises. First, they can occur with great suddenness, despite the absence of any significant prior shift in readily observable economic variables. The only obvious explanation for this is a sudden and wide-ranging change in risk perceptions that induces economic entities to attempt to modify significantly their exposure to risk. Second, the magnitude of the

---

[*] This paper is based on Chapter 20 of *East Asia in Crisis: From Being a Miracle to Needing One?*, edited by Ross H. McLeod and Ross Garnaut (McLeod 1998b).

[1] Alternatively, offshore borrowing by the government to prevent reserves from falling may result in a continuous decline in the government's net foreign assets.

[2] Krugman (1998) refers to this as the 'canonical', or first-generation, crisis model.

impact (on reserves, the exchange rate and/or interest rates) may be dramatic. This reflects the emergence of a huge pool of internationally mobile, risk-sensitive financial capital during the last decade or so[3] and the corresponding (and closely related) reduction in the cost of shifting funds from one country to another. Third, the warning signs are far less clear for new-style than for old-style crises. Falling (rising) prices for a country's major exports (imports) are readily observable, as are increasing current account deficits, large budget deficits, excessive money growth, high inflation and falling international reserves. But although some risk premia can be observed in the financial markets, there are no obvious means of monitoring potential changes in risk perceptions, much less predicting the magnitude of their impact should they occur.[4] To put it another way, it is only known there has been a change in risk perceptions when this has already had its impact in the financial markets. The only way to get an idea of how great this change has been is to observe the magnitude of its impact.

In contrast to old-style crises, new-style crises can occur without any additive spillover effect to the current account.[5] In a pegged exchange rate regime, liquidity tightens as money is used to purchase foreign exchange; this should have a negative impact on aggregate demand resulting in increased net exports. If the exchange rate is floating, depreciation will raise competitiveness, again leading to increased net exports. In a hybrid regime, if the currency is allowed to depreciate somewhat, but interest rates are raised in order to prevent this going too far, both of these effects will be seen. On the other hand, if money is allowed to increase, these effects will tend to offset each other.

The outlook in relation to the current account becomes less optimistic, however, when the possibility of negative feedback on risk perceptions is considered. If the initial disturbance is strong enough to cause a significant devaluation or a big jump in interest rates, this will add to the perception of increased risk in holding claims over firms and banks that previously had large exchange rate or interest rate risk exposures. If the banking sector then begins to look shaky, the availability to it of

3 An important part of this phenomenon is the growing importance of securitised finance relative to bank loans. Bank finance is much more stable because of the need for banks to maintain long-term relationships with their borrowers. Securitised finance does not rely on such relationships — shares, bonds, commercial paper and the like are issued and later traded in impersonal markets.

4 Current attempts to find reliable indicators of vulnerability to currency crises can be thought of as seeking to monitor potential changes in risk perceptions. Such efforts will probably come to little — the risks are so diverse and subject to change in relative importance that this amounts to shooting at an impossibly fast-moving target.

5 The 'new-style' crisis description is in the same mould as the 'second-generation' models discussed in Krugman (1998).

funds may decline suddenly. Firms, including exporters, may not be able to obtain working capital and exports again may fail to grow, despite the stimulus from depreciation. Moreover, if several countries experience a new-style crisis simultaneously, as has been the case in East Asia, the picture becomes much more complicated. If each country is important to the others as an export market or as a competitor in markets elsewhere, the export response in all countries may again be weak and the impact on their current accounts of reduced aggregate demand and exchange rate depreciation will be muted.

In reality, there may be no cases that are purely old-style or new-style. In the East Asian crisis, different countries lie along the spectrum between the two extremes (and some have avoided both kinds of crisis). Of the countries considered in McLeod and Garnaut (1998), Indonesia seems closest to the new-style extreme, while Thailand — the first domino to fall — combines elements of both. Nevertheless, outlining these polar extremes is a useful heuristic device because it suggests the possibility that different policy responses may be appropriate to each case, and that it will be necessary to focus on different things than previously if new-style crises are to be avoided, or their destructiveness minimised, in the future.

The sections that follow build on these basic ideas.

9.2 The crisis as a failure of risk management

Risk management may be thought of as the art of avoiding nasty surprises. The widespread failure to foresee the East Asian crisis after so many years of seemingly solid economic performance — much less to imagine how severe it could be — certainly comes into the 'nasty surprise' category. To differing degrees, a state of euphoria existed in the countries of the region in which investors and governments failed to recognise the magnitude of the risks that existed and to manage those risks adequately.[6] This made certain countries highly vulnerable to a financial crisis — like the Titanic steaming confidently across the North Atlantic oblivious to the proximity of icebergs, and complacent about its capacity to absorb a collision with one.

Various categories of risk were important in different degrees in the countries studied in McLeod and Garnaut (1998). These risks included:

[6] No doubt they were encouraged by applause from the many champions and supporters of East Asia — this writer included — and the World Bank's proclamation of 'The East Asian Miracle' (World Bank 1993).

- exchange rate risk (the risk posed by adverse exchange rate movements for unhedged borrowers and investors);
- investment risk (the risk of low or negative returns from poorly conceived, badly managed and speculative investments);
- bank sector risk (the risk of loss for depositors at banks that were themselves exposed to risks of various kinds);
- policy risk (the risk of harmful misjudgments by policy makers when responding to changing economic circumstances); and
- in some countries, political risk (the risk that political developments could bring adverse policy changes, a breakdown of social stability, loss of political protection for privileged firms, and so on).

Political risk was overwhelmingly important in Indonesia, but much less so elsewhere.

Financial risks arise from holding assets and liabilities whose future value is uncertain by virtue of the various categories of risk just mentioned. Such risks are inescapable aspects of doing business and accumulating wealth (saving) and they contributed strongly to initiating the crisis that emerged in 1997. In so far as financial assets tend to proliferate and their aggregate value to grow far more quickly than the value of output and the physical capital stock as countries develop (Goldsmith 1969), it may be conjectured that financial risks might also tend to grow rather rapidly. Indeed, Warr (1998) and Athukorala (1998) can be interpreted as arguing implicitly along these lines. Warr draws attention to the rapid growth of financial assets held by foreign investors, and Athukorala to the rapid growth of banking assets (as reflected in the growth of deposits), in the rapidly developing Thai and Malaysian economies. Market processes develop mechanisms for dealing with financial risks,[7] but government policies may hinder them, as will be discussed shortly.

A fundamental distinguishing characteristic of financial risk is that the act of insuring or seeking protection against an adverse movement in the price of an asset or liability tends to cause this price to move in the same direction as that which is feared. Of particular relevance in the present context is that, if an expectation that the currency will depreciate takes hold, those who have borrowed in foreign currencies will want to buy such currencies in order to repay their loans immediately or to ensure they will not be affected later by the expected depreciation. At the same

[7] For example, stock markets, mutual funds, banks, life insurance and pensions providers all facilitate the diversification of financial risk. Derivatives markets permit financial risks to be hedged.

time, those who are holding assets denominated in the local currency will want to liquidate them and use the proceeds to buy foreign exchange. Finally, there will be other entities who, although not initially exposed to such risks, see the opportunity for profit in others' potential misfortune. If these entities form a strong view as to the likely direction of future movement in financial asset prices, they will deliberately seek an exposure from which they will gain if their prediction turns out to be correct.

The impact of the actions of all of these parties — if there is no central bank intervention or if the central bank is overwhelmed by the demand for foreign exchange — will be to cause the expected depreciation to occur. Likewise in the case of a run on a bank, the withdrawal of deposits on a large scale and in a short period is likely to hasten the collapse of the bank, in turn resulting in a loss of value of the remaining deposits.

With these ideas in mind, 'financial panic' may be described as an episode in which large numbers of economic entities suddenly change their financial behaviour (ie try to modify significantly their portfolios of assets and liabilities) simply because they observe other entities doing so. This is often described as 'herd behaviour', but it is not necessarily irrational given the cost of quickly acquiring better information about the assets and liabilities in question. Thus, when a herd of zebra suddenly run off in the same direction, it may be because one of them saw a lion approaching. The zebra who stays put because it has not yet seen the lion with its own eyes runs the risk of going from needing a meal to being one. And, to round out the analogy, the vulture that does not follow its flock is likely to miss out on the free meal of lion leftovers that one of its companions has spotted.

The precondition for a financial panic to occur — that is, of vulnerability to crisis — is a set of circumstances in which economic entities are aware of the possibility of changes in the prices of assets and liabilities, but are not insured against such changes (or in the case of speculators, deliberately exposed to them). Then, when they see some indication that the occurrence that concerns them might be about to happen, they act in a manner that amplifies any initial disturbance. All are aware that, unless they are among the first to hedge a newly perceived risk (or to take a speculative position), the opportunity to avoid the loss (or enjoy the gain) may quickly disappear.

In such circumstances, when hidden risks come to light or when the probability of losses or their magnitude is perceived to increase, the rush to hedge the risk and to speculate hastens the very change in asset and liability values that is of concern. McKibbin (1998) incorporates such an occurrence in a general equilibrium modelling framework. This would appear to be a large part of the East Asia crisis story. In Thailand, people began to lose confidence in the US dollar peg for the baht

given the rapidly declining export growth and outlook, growing concerns about the prospects of listed companies and a mounting excess supply of real estate. Thus, investors, both foreign and domestic, were beginning to shift funds to other countries, where investment opportunities seemed more promising, putting pressure on reserves. Speculators eventually saw what was going on and the peg to the dollar was soon tested and broken.

This had elements of, but was something more than, an old-style balance of payments crisis. Both export and import growth had been very volatile, making it difficult to interpret the state of the current account in the months leading up to the crisis. International reserves, although declining, seemed adequate (in terms of their equivalence in months of imports). Subsequently, it was learned, however, that the bulk of reserves had been sold forward by the central bank. The extent to which Thailand's crisis originated in the current account in the old-style sense, or was a new-style crisis precipitated by concerns about a much wider range of issues concerning the domestic economy as well, therefore, is not clear. In Indonesia's case, however, the initial disturbance seems to have been almost entirely in the capital account (McLeod 1998a). Inflation was falling, there had been no marked deterioration in the current account, and reserves had been increasing. But the sudden float and devaluation of the baht in neighbouring Thailand and then the ringgit in Malaysia triggered a re-evaluation of risk exposures of all kinds. This, in turn, brought on precisely the changes in asset and liability values that investors feared.

The first such re-evaluation in Indonesia related to exchange rate risk. Private sector foreign debt was of the order of $70 billion, 83 per cent of it unhedged. A further $27 billion in foreign currency loans had been obtained from domestic banks. Borrowers began to rush to buy dollars in early July 1997, as did domestic and foreign entities holding rupiah-denominated deposits and other financial assets. Those who had borrowed offshore to finance investment in the non-tradables sector, especially property, incurred large unrealised losses if the companies in question were listed, selling pressure began to force the share prices down, and it is probable that some of the sale proceeds were converted to dollars. The focus of concern soon widened to encompass the banks. They were exposed to property and, even if not directly exposed to exchange rate risk, many of their borrowers were, which meant that the banks faced much larger credit risks than previously imagined. Concerned depositors began to withdraw funds. Deposit rates had to be increased to staunch the flow, but lending rates could not be raised because large parts of the corporate sector were already technically insolvent.

Policy risk also came into play. The government, in its own way, became caught up in the panic and reacted in a way that added the threat of a downturn in economic

activity. It imposed a drastic liquidity squeeze, and soon announced plans to cut its own spending. It confirmed fears about the banks by closing a number of them. Political risk became important too. The government began to show obvious signs of internal conflict, partly because a new cabinet was due to be formed and a new vice president elected within a few months. With each of President Soeharto's successive five-year terms of office it became increasingly likely that the incumbent vice president would become the next president, raising enormous concerns about the potential impact of this inevitable transition. Backsliding in relation to reforms adopted by agreement with the IMF became apparent, increasing the risk that Indonesia would lose the high degree of international support it had enjoyed for the previous three decades. Eventually, the ethnic Chinese community began to be treated as scapegoats, with military leaders pointedly requesting them to repatriate funds they had sent offshore. Another aspect of risk had to be re-evaluated, with the memory of anti-Chinese rioting associated with the parliamentary elections a year earlier still fresh, and darker memories from longer in the past always in the background.[8]

For brevity, the discussion above has dealt with just two of the case study countries in order to highlight the distinction between the origins of the crisis in Indonesia — and, perhaps less obviously, Thailand — and the old-style balance of payments crises for which the means of prevention and the appropriate policy responses are well known. The fundamental lesson is that risk matters. It is necessary to be aware of risk, to manage it with care, and to have a clear plan of action for occasions when disturbances do occur; the achievements of decades of development may unravel at an alarming speed otherwise. The possibility of rapid, large-scale movement of financial capital across international borders needs to be taken as given, and choices need to be made as to macroeconomic and financial sector policies that will do most to enhance economies' resilience in the face of these movements. In particular, there is now an urgent need to ensure that government policy does not add to risk, but diminishes it. These issues are addressed in the remainder of this paper.

Financial sector policies

If financial risk is at the heart of new-style balance of payments crises, this raises important policy issues in relation to the process of financial reform. In countries in which significant financial deregulation occurred — of which Indonesia is the prime example (McLeod 1999) — the resulting rapid expansion in the financial systems has often been blamed as an important element in the crisis. Indeed, it is hard to believe that banks and non-bank intermediaries could pay sufficient attention to the

[8] The ethnic Chinese suffered greatly in the aftermath of the attempted coup in 1965.

risks involved in expanding their loan portfolios so rapidly.[9] It would be wrong to conclude, however, that financial deregulation should be reversed in the countries that have already moved in this direction and avoided by those yet to do so.

Financial system weakness prior to the crisis was also a problem in countries where there was still much scope for reform, including Japan (Horiuchi 1998) and Korea (Smith 1998), for example. And in some countries as yet largely unscathed by the crisis, such as China (Song 1998) and Vietnam (Leung and Doanh 1998), there was (and remains) an urgent need for reform, because the existing financial sector does not do a good job of allocating financial resources. The writer therefore remains in favour of financial reforms that give a much stronger role to market mechanisms. Many observers have come to the same conclusion, but add the proviso that reform must be preceded by a building up of the system of prudential regulation and supervision. This will be more easily said than done. The art of financial system supervision cannot be learned merely from textbooks or in the classroom. Many of the requisite skills will only be learned from experience, and it is unlikely that the necessary regulatory and supervisory capacity can be created in advance of introducing policy reforms.

There is certainly a fundamental problem that needs to be confronted. As Fane (1998) argues, people tend to see financial institutions, especially banks, as being guaranteed by governments — implicitly, if not explicitly. The experience of the crisis shows, by and large, that this belief is well founded, despite earlier government protestations to the contrary, governments have indeed stepped in to prevent losses to depositors and other creditors. This being the case, Fane would appear to be correct in arguing that this provides a strong case for government intervention in finance, the essence of which must be to ensure that the shifting of risk from banks' owners to the general public is minimised. The reform of prudential regulation can and should proceed immediately and it should certainly be an integral part of moves to deregulate financial systems that have been heavily controlled in the past. The removal of barriers to entry, branching, interest rate controls and credit ceilings is likely to unleash a strong latent demand for banking services, creating the prospect of high returns for institutions. It is essential that the rush to establish market share be tempered by the credible threat of substantial losses to shareholders if mistakes are made.

[9] It should not be presumed that Indonesian banks were bankrupt or that generally they were heading in this direction, prior to the huge depreciation of the rupiah. The level of non-performing loans was high for banks as a whole, but these were much more heavily concentrated amongst the state banks than private banks. There is little hard evidence to show that problem loans were running out of control in the latter.

Effective prudential regulation needs to be tightly focused. In Indonesia, by contrast, the central bank brought in a vast array of prudential regulations in 1991. It seemed clear at the time — and it has been demonstrated by subsequent events — that the regulators had no real sense of what was important. By listing countless aspects of bank management to monitor, they succeeded in monitoring nothing effectively — but at great cost.[10]

One thing the crisis has demonstrated is that the authorities need not restrict themselves to using local expertise in the monitoring process. The new Indonesian Bank Restructuring Agency is making use of foreign accounting and auditing firms to get a clearer understanding of the true financial condition of Indonesian banks (Johnson 1998). There is every reason for continuing to make use of expertise more readily available in foreign firms than in the domestic supervisory bureaucracy. Other countries would do well to study this initiative.

As Fane notes, the most fundamental issue is capital adequacy. Ultimately, if the banks have sufficient capital, their owners will have a strong interest in sound management — including risk management — and this will do more than anything else to protect the interests of depositors and taxpayers. He argues in favour of much higher capital adequacy ratios (ie ratios of capital to assets). The writer supports this recommendation, but would broaden it — in particular, to require specific attention to foreign exchange exposures.

Indonesia had a regulated maximum of 25 per cent for the 'net open position' (NOP) of banks, where the NOP is defined as the ratio of net foreign currency-denominated assets to capital. In retrospect, this was far too high. A bank with a NOP of this size would have its capital entirely wiped out if the cost of foreign currency increased by a factor of five, as occurred in Indonesia between July 1997 and January 1998. Fluctuations in bank capital ought to be kept far smaller than this in conceivable circumstances: a maximum NOP of 2.5 per cent would seem closer to the required order of magnitude than 25 per cent. Given the above mentioned concerns about the government as guarantor, banks should not be in the business of taking on large foreign exchange exposures, nor do they need to do so in order to meet their customers' requirements. Provided they are permitted to borrow and to place funds offshore,[11] they can meet their customers' demands for foreign currency loans and deposit instruments.

[10] The central bank's move to implement an 'enhanced credit monitoring system' capable of reporting on loans down to a value of only $15 000 is an indication of its inability to focus on what really mattered in this field (Hendrobudiyanto 1994, p. 164).

[11] Curiously, offshore borrowing by banks was tightly restricted in Indonesia.

It is not much use to have these kinds of requirements, all of which relate to capital adequacy, if capital is not measured accurately. Fane points out that in at least some of the East Asian countries, lags between loan repayments falling into arrears and the eventual reporting of charges against bank capital resulting from provisioning for possible losses were far too long. In these days of computerised management information systems and flighty capital movements, this is an unnecessary and dangerous shortcoming.

There is an interesting contrast here with the manner in which risk is managed in markets such as the futures markets. Futures trading is done 'on margin'. In other words, investors need only put up a small amount of their own capital in order to be able to hold futures contracts much greater in value. By leveraging in this manner, there is the possibility of enormous profits if the underlying price moves in the right direction. But there is also the possibility of enormous losses if the price moves in the opposite direction. To protect the counter-party to each contract, therefore, the managers of futures exchanges require additional margin deposits to be made by the investor, sufficient to cover the entire (unrealised) loss, whenever there is an adverse price movement. This requirement operates on a very short time horizon. Calls to margin might be required on the same day or at least the following day. Consideration might be given to requiring that the process of provisioning for losses in banks approach this kind of standard.[12] This would have the effect of requiring banks to have higher levels of capital, which is precisely what is needed in order to make financial systems more resilient in the future.[13]

Fane also draws attention to the desirability of permitting foreign banks to operate freely in competition with domestic banks. Protectionist sentiment in many countries has prevented much relaxation in this area, even when other reforms have gone ahead (Athukorala 1998). But it is important to note that governments in the troubled countries have not needed to call on public funds to bail out any of the foreign banks during the crisis. The same was true in Australia's earlier limited banking crisis in the early 1990s. Foreign banks were the biggest losers (as Gruen, Grey and Stevens (1998) point out), but their parents covered the losses without cost to Australian depositors or taxpayers. Likewise, the collapse of Barings Bank in 1995 as a result of speculative activities in its Singapore office imposed

[12] This would complement, rather than replace, the practice of adjusting loss provisions on a case-by-case basis when management becomes aware of changes in the borrower's business prospects, the value of collateral assets, and so on.

[13] One issue here is that, if banks face much more stringent controls of this type than hitherto, there is a danger that other kinds of institution will spring up that will not be subject to the banking regulations. The most appropriate solution to this problem would be to define a 'bank' as any institution that 'collects deposits from the general public'. The authorities could be given a good deal of discretion to determine what constituted such deposits.

considerable losses on its shareholders, but caused little harm to Singapore other than a loss of face for the supervisory authorities (D'Ingeo 1996).

Exchange rate risk

Growing fears of devaluation have been at the heart of the sudden change in risk perceptions in East Asia. Concerns about banks were initially mitigated by government promises of the safety of deposits (although later they began to play an important role in some countries), whereas assurances regarding exchange rate stability most likely were regarded as expressions of policy intentions rather than firm commitments.

Indonesians remained highly sensitive to the possibility of devaluation, notwithstanding the country's high and increasing reserves, because of lasting memories of the three large devaluations during 1978–86 and the psychological scars of hyperinflation during the mid-1960s — which so devalued the old currency that a new one had to be introduced. This sensitivity was apparent in the large differential between domestic and foreign interest rates, far in excess of both officially targeted depreciation and recorded depreciation over the previous decade. It was manifested also in relatively frequent bursts of speculation against the currency, driven by the flimsiest of rumours.

And yet there were enormous unhedged exchange rate risk exposures in Indonesia and in the other crisis countries — in retrospect, a gross failure of risk management. It would be wrong to conclude, however, that this was a case of market failure to which the appropriate response would be to impose various kinds of controls on private sector behaviour. Those who believe in markets do not argue that mistakes are never made, but that properly functioning markets penalise those who make them. They also argue that government intervention is often the real cause of problems that are blamed on market mechanisms. Therefore, this paper now turns to consider whether, and how, the present crisis may have been initiated by inappropriate government policies.

In various chapters of McLeod and Garnaut (1998) attention is drawn to the state of business euphoria that existed in countries that had been growing rapidly for many years on end. In less heady times, bankers and other funds managers sensibly adopt the attitude that they need to be persuaded to part with funds by way of well researched business proposals, offers of solid security, and a willingness on the part of applicants to put a substantial amount of their own funds at risk in the venture to be financed. However, when an economy has been racing ahead for years on end, there is a strong tendency for those same bankers and financiers to become aggressive salespeople, almost begging businesses to take funds off their hands. All

too often, collateral is in the form of real estate valued on the basis of extrapolated recent upward trends in property prices, and projects of all kinds are financed on very high gearing ratios.

The 'euphoria', or more precisely the lack of attention to risk, in the years preceding the crisis related not only to the inherent viability of projects themselves,[14] but also to aspects such as the cost of finance and the quality of security for loans. In particular, the exchange rate risk was often ignored. The problem with exchange rate policy in the long period leading up to the crisis was that its very success in maintaining stable rates (or rates of depreciation) against the dollar encouraged the belief that governments could guarantee such stability. In the days before new-style crises, governments could make these promises and live up to them provided they managed monetary policy properly (which in turn also required responsible fiscal policy). In this new era of financial fragility, this is no longer the case.[15] The mobility of the large pool of global financial capital has rendered such promises barely credible in countries without capital controls, especially in those that rely relatively heavily on foreign capital (other than direct investment) and where banks have mobilised large volumes of deposits.

At the same time, the credibility of government promises about the safety of deposits in banks is also rightly called into question in this new era. The mobile pool of financial capital so often mentioned in McLeod and Garnaut (1998) includes funds mobilised by domestic financial systems which can easily contribute strongly to new-style crises[16] — and compound them further if there is a sudden loss of confidence in domestic banks. The volume of funds held in banks in the developing countries of East Asia had grown far larger prior to the crisis than in the old-style crisis days.[17] As total bank assets increase relative to the size of the economy, the credibility of government guarantees of bank deposits is likely to decline (in the absence of measures to strengthen banks' resilience). In Indonesia, for example, the outer bound of the government's contingent liability by virtue of its implicit guarantee of bank deposits grew more than three-fold in the decade to June 1997, from 16 per cent to 50 per cent of GDP. The current crisis clearly demonstrates that risk exposure of such magnitude can have devastating implications for the budget. It

14 As implied by the demand for their output, required initial investment outlays, the cost of non-financial inputs, and so on.

15 In the special case of currency board countries, the promise is much more credible.

16 The Mexican crisis of 1994 was driven to a much greater extent by residents shifting funds offshore than by foreigners repatriating capital (Folkerts-Landau et al. 1995, p. 7).

17 Broad money growth at rates of the order of 20 per cent a year in many countries gives some indication of the increasing scale of banking activity.

is essential, therefore, that they be managed consciously rather than simply assuming — or hoping — that the guarantee will never need to be honoured.

Taking this line of thought further, it is worth asking to what extent the large exchange rate risk exposures that interacted with increased risk perceptions to spark the present crisis would have emerged in the absence of government promises (explicit and implicit) about the safety of banks or, more realistically, if banking regulations had required banks to have much higher levels of capital relative to risk assets and exchange rate exposures. Presumably, it would have been much less. If the shareholders of banks had borne much more of the risks relative to governments and depositors, they could have been expected to require such exposures — direct and indirect — to be kept much smaller.

In short, governments themselves contributed greatly to the emergence of the East Asian crisis by following policies that encouraged complacency in relation to risk management on the part of the private sector.

9.3 Handling of the crisis

Macroeconomic policy responses

The main policy weaknesses that appear to have created the preconditions for financial panic have been discussed. The next aspect of risk management that needs to be considered is the preparedness of governments to respond appropriately to sudden and unexpected financial disturbances. A major part of the explanation why the crisis became so much more severe in some countries than in others is to be found, not in what was happening prior to the initial disturbances, but in the way in which governments reacted to the sudden speculation against, and the decline of, their currencies.

Should they have reacted at all? The question may seem ingenuous. After all, several of the countries studied in McLeod and Garnaut (1998) suffered greatly as a result of the disruption to their economies, and it seems natural to imagine that their suffering would have been considerably worse in the absence of government measures to ameliorate the crisis. The notion of ‘doing nothing’ is a nonsense, of course. What is really at issue is whether governments would have done better to maintain the previous settings of macroeconomic policy, rather than to adjust them in light of the disturbance. No government maintained the previous settings in their entirety, but some came much closer to doing so than others. The countries that have fared worst seem to be those in which governments themselves were caught up in the panic, and implemented a range of poorly chosen policies as a result.

The question of altering policy settings was of less importance in the countries that maintained controls on their capital account, since freedom of capital movement is a precondition for new-style crises. Accordingly, the discussion in this section focuses on the more open economies facing actual or potential capital outflow and exchange rate depreciation. In the case of Hong Kong, the government's exchange rate policy throughout the crisis has been to 'do nothing'. This provides the opportunity to witness the operation of a currency board system under conditions of extreme stress. Some observers have argued that this system has been found wanting because of the dramatic surge in interest rates resulting from speculation against the Hong Kong dollar in October 1997, but this is not the relevant consideration. Proponents of currency boards argue precisely that it is better to have a fixed exchange rate and to live with some interest rate volatility (Walters 1998). It is natural to view with horror the jump in interest rates to almost 300 per cent in October, as a result of the sudden decline in base money as entities sold the domestic currency (Cheng, Wong and Findlay 1998). However, upon more sober reflection, it can be seen that only the overnight rate attained such giddy heights and that this extreme surge had run its course within a few days. Rates remained higher than usual for some time thereafter, but circumstances were also anything but usual. Nevertheless, Hong Kong's growth turned negative for the first time in thirteen years in the first quarter of 1998, so it cannot be said that the government's policy response has provided a painless path through the crisis.

The Taiwanese government's response can also be characterised as mainly passive. The government had adopted a floating exchange rate regime a decade earlier (Kuo and Liu 1998) and, although by no means a clean float, the government was quick to allow the currency to depreciate when market sentiment moved in that direction in the latter half of 1997. Of all the 'open' Asian countries considered here, Taiwan seems to have incurred the least damage from the crisis. At the other extreme from these two cases, Indonesia's government has adopted a far more activist response, implementing a wide array of policy changes — some of its own volition, some as a response to pressure from the IMF. Yet its economy clearly has suffered far more than any other. Clearly then, it is important to consider whether an activist approach was called for and, if so, what it should have encompassed.

The most fundamental objective of macroeconomic policy making is to maximise the economic welfare of the general public. Whatever else, this requires governments to do all they can to keep resources of both labour and capital fully employed.[18] Broadly speaking, there is no reason why even large changes in asset prices resulting from changed expectations should lead to unemployment. A stock

[18] Except where the variable social cost of keeping an enterprise operational exceeds the marginal social benefit of doing so.

market crash, for example, unlike an earthquake, has no physical implications — the real assets underlying company shares are still in place. It is physically possible then to maintain output at previous levels, and there is every reason why this should be the aim of government policy. The same is true of a currency crash affecting firms with foreign currency liabilities.

Two problems arise in such circumstances, however, that may need to be dealt with. First, there is a wealth loss on the part of those whose assets have just been devalued or whose liabilities have just increased, which is likely to have a negative impact on consumption. But our experience of the 1987 share market crash suggests that this did not turn out to be the great disaster predicted by many at that time. World economic growth was not seriously affected — certainly not on the scale of the disruptions seen recently in East Asia. It is easy to suggest plausible reasons why this was so. First, the individuals concerned may simply have chosen not to reduce consumption, but to reduce the expected value of assets to be bequeathed to their heirs. Second, to the extent they felt it necessary to respond by reducing their own consumption, this could be spread over many years into the future. Finally, governments responded sensibly by loosening monetary policy to some extent. For all these reasons, the overall impact on aggregate demand was slight and a large surge in unemployment was avoided.

Second, there is the problem of the banks' reaction. If there are significant falls in share or property prices or significant increases in firms' debt burdens, the value of collateral available to support bank lending declines accordingly. Banks will therefore want to cut back on lending, other things equal. At the same time, however, to the extent consumer demand declines because of the wealth effect, the inflow of deposits to banks should increase. These two effects tend to counteract each other, and there may be little change in the level of spending financed by bank loans (assuming no tightening of monetary policy by the government). The issue of how best to handle troubled banks will be discussed shortly.

Provided banks have not been too heavily involved in lending secured by shares, there is no inevitability about a crash in the share market having major implications for economic activity.[19] Property crashes are more problematical since property is widely used as collateral by banks and other financial institutions. Exactly analogous problems arise when banks become heavily exposed to entities with large foreign currency liabilities if the exchange rate depreciates significantly. In both cases, however, it is important to emphasise that these are financial, not physical, losses.

[19] Japan appears to be the only one of the case study countries in McLeod and Garnaut (1998) where this has been a serious problem.

From the point of view of its owner, a property crash that halves the value of a building or an exchange rate decline that doubles the burden of debt may seem little different from a fire that causes damage that will cost as much to repair. From a broader perspective, however, the real level of services output from the building need not decline at all in the first two cases (although the price of such services will need to do so), whereas it necessarily declines in the third. Changes in the value of assets and liabilities affect national wealth and its distribution among individuals, but they need not cause the level of production of goods and services to fall if prices are sufficiently flexible. There is little or nothing a government can do to offset reductions in national wealth that result from changes in financial asset and liability values (although it can certainly affect the distributional impact). What it must see as its objective, therefore, is to ensure that the output of goods and services does not fall unnecessarily. If price flexibility is lacking, this may require some loosening of fiscal and monetary policies.

This is the main feature that distinguishes the 1987 stock market crash from that which occurred in 1929. And it is an important feature distinguishing the countries that have suffered most greatly from those that have been affected relatively little by the East Asian crisis. Although this has been the result of a currency crash, rather than (or leading to) a stock market crash, it makes little difference whether wealth declines as a result of a fall in the value of assets or a rise in the value of liabilities. In both circumstances governments need, if anything, to err on the side of boosting demand, not adding to its decline. In East Asia in 1997, governments tended to respond to the initial decline of their currencies, not by loosening, but by tightening fiscal and monetary policy. In some cases, there was a doctrinaire belief in the inherent desirability of balanced budgets which overrode any sympathy policy makers may have had for the anti-cyclical demand management approach recommended by Keynes. If tax revenues were falling because of private sector wealth reduction, therefore, government spending would need to be cut commensurately. This was certainly the government's view in Indonesia (McLeod 1998a), and it appears to have been an important factor contributing to the negative growth recorded in Malaysia and Hong Kong in the March quarter of 1998 (Athukorala 1998 and Cheng, Wong and Findlay 1998). In addition, there was a view that currency and share markets would react adversely to anything other than fiscal austerity.

Concern about the current account was also a key consideration. The balance of payments was suddenly in deficit, and a common view was that this called for adjustment on the imports side in order to prevent further depreciation (McLeod 1998a and Athukorala 1998). Cutting government spending was therefore seen as desirable because it would reduce the demand for imports both directly and indirectly (through the negative multiplier effect on private sector spending).

Finally, tightening liquidity in order to push up interest rates and thus stem, or reverse, the outflow of capital was often thought necessary.

All of this was precisely the reverse of what was required if economies were to remain fully employed. An alternative approach would have been to allow the budget to go into deficit (so that the economy would benefit from the automatic stabilising mechanism under which taxes and other revenues fall as incomes fall), and (for non-currency board countries) to allow the exchange rate to depreciate even further (thus allowing the market mechanism to determine the accommodating changes in spending and importing made necessary by the sudden outflow of capital).[20] While a modest increase in base money growth would also have been appropriate, most of the monetary impact of the budget deficit (including that part of the deficit resulting from the cost of government guarantees of bank deposits) would have needed to be offset by issuing bonds or central bank certificates and perhaps by selling off some international reserves.

An important implication of the shift from old-style to new-style balance of payments crises is that, if the IMF is to have any useful role to play, will need to be very different from previously. In the past, putting an end to a balance of payments crisis typically required a devaluation and a politically difficult tightening of fiscal and monetary discipline. The near exhaustion of reserves during the crisis allowed the IMF to come in with the offer of loans that would provide temporary liquidity during the wait for export recovery, conditional on the government making the necessary macroeconomic policy adjustments.

In new-style crises, it is not export recovery that is awaited, but a recovery of confidence. It seems most unlikely that the financial resources the IMF can make available to governments that choose to support their currencies when under speculative attack will increase as rapidly as the global pool of capital available for undertaking such speculation.[21] In short, the old approach will become increasingly less feasible, and the IMF will need to reinvent itself if it is to remain relevant. This may mean shifting to a role more akin to that of a ‘management consultant’ when brought in to assist governments in times of economic crisis, away from that of stern disciplinarian seeking to impose its will on policy makers.

---

[20] A genuinely floating exchange rate has the advantage that it does not ‘offer speculators an easy target’ since the government is not committed to defending any particular exchange rate (Krugman 1998).

[21] This is an extension of Krugman’s (1998) observation that the European currency crises of 1992-93 ‘demonstrated the near-irrelevance of foreign exchange reserves in a world of high capital mobility’. Recall that the IMF could pledge only $10 billion to assist Indonesia in November 1997, while Indonesia’s reserves themselves were more than twice that amount.

Finance sector policy responses

The treatment of weak financial institutions in the context of the unfolding crisis has also been important. At a superficial level, Thailand and Indonesia followed a similar course. In both countries, huge amounts of public funds have been poured into troubled financial institutions. Moreover, a number of institutions were closed abruptly in both countries early on. In Thailand's case there have been subsequent runs on non-bank financial institutions and banks, but the side effects of the closures seem to have been considerably less disruptive than in Indonesia, where they had a devastating impact — causing a run on most of the private domestic banks and eventually almost paralysing the banking system. It is important to account for this difference. The risk of failures amongst financial institutions will always be present — and will be immeasurably higher in the midst of a financial crisis — and good risk management practice requires planning for, and therefore understanding of, such contingencies.

After the first group of bank closures in Indonesia was announced at the end of October 1997, it quickly became common knowledge that the number closed reflected a compromise between the government and the IMF. It could just as easily have been forty as the actual sixteen. This made depositors concerned about the condition of the remaining banks and led to large-scale withdrawals in the ensuing weeks — notwithstanding the government's promise to cover the claims of all depositors up to a maximum of Rp20 million and its assurances that no other banks would be closed.[22] Deposits migrated to state banks and to foreign banks, but it was not practicable for either group to take over the private banks' borrowers, so the wheels of finance began to seize up.

In Thailand, many more financial institutions closed down than in Indonesia, but they were not banks. They were finance companies that were not part of the payments system. They drew the bulk of their funding from the wholesale market in the form of promissory notes and bills of exchange, rather than individuals' deposits, and they did not provide working capital finance to the business sector. Unlike in Indonesia, therefore, their closure did not have significant implications for the banking system.

Creditors of enterprises that default must always weigh the relative merits of forcing bankruptcy and the cessation of activity, on the one hand, and installing new management in the hope that the ultimate loss will be less severe if operations are continued, on the other. In a context in which a government effectively guarantees banks' liabilities, it is in the same position as such creditors when any bank fails,

[22] This lack of credibility was a clear indication of public consciousness of what was referred to earlier as political risk.

and it is not clear that any good purpose is served by closing down troubled banks, rather than keeping them going under new management.[23] Closing a bank requires its employees to be dismissed — most of them having been in no way responsible for the bank's troubles — yet its deposit and lending customers must still be serviced. Shifting customers to other banks involves considerable disruption.

It is essential, of course, to turn off the tap if government funds are being wasted, but the basic issue is whether the bank in question can cover its variable costs if it continues to operate, not the size of losses it may have accumulated in the past. The question is especially important in conditions of systemic crisis, where not just a single institution is in trouble, but many. In the Indonesian case, the banking law gave the government power to replace a bank's management and to explore various avenues for replenishing lost capital. At a later date, banks taken over in this manner could have been sold as going concerns, or their assets progressively sold off and their deposit business transferred to other institutions. Closure of banks without any warning seems in retrospect to have been unnecessary and counterproductive.

9.4 Conclusion

In this paper it has been argued that the world has entered a new era of financial fragility. This new era has been ushered in by a huge and growing pool of highly mobile financial capital, including funds mobilised by rapidly expanding banking and financial systems in many developing countries, and by the global trend to openness in regard to capital flows. These factors provide the preconditions for new-style balance of payments crises that demand different policy responses from old-style ones.

The new-style crises are driven by changes in risk perceptions, and efforts to prevent them must be guided by the principle of avoiding policies that inadvertently encourage private sector actors to be complacent in relation to financial risk. In respect of exchange rate risk exposures, such complacency might be discouraged by adopting genuinely flexible exchange rates, or by aiming for stability relative to a genuinely diversified basket of foreign currencies rather than just the US dollar (which would amount to flexibility of rates against any single currency in which borrowing might be undertaken). But perhaps the most promising possibility involves increasing significantly the amount of capital shareholders of financial institutions have at stake relative to the size of those institutions' activities. The notion underlying such policy reforms is the axiom that capitalism does not work

[23] The writer owes this point to George Fane.

well if only the gains resulting from business decisions accrue to the investors who make them and not the losses.

New-style crises also require a quite different policy response. Except if countries have adopted the currency board system, exchange rates should be permitted to be market determined against a backdrop of steady or slightly more expansionary fiscal and monetary policy settings. The remedy for old-style crises — devaluation, accompanied by fiscal and monetary tightening — is entirely inappropriate. The notion that it is sensible to 'do nothing' in the face of a new-style crisis clearly will take time to be accepted by governments, but, like first-time bungee jumpers, they will eventually come to realise that although the value of currencies may plunge and bounce alarmingly, ultimately their equilibrium values will depend on how carefully their quantity is managed.

References

Athukorala, P. 1998, 'Malaysia', Chapter 5 in McLeod, R.H. and Garnaut, R. (eds) 1998, *East Asia in Crisis: From Being a Miracle to Needing One?*, Routledge, London.

Cheng, Y., Wong, M. and Findlay, C. 1998, 'Singapore and Hong Kong', Chapter 10 in McLeod, R.H. and Garnaut, R. (eds) 1998, *East Asia in Crisis: From Being a Miracle to Needing One?*, Routledge, London.

D'Ingeo, M. 1996, *Can consumers rely on the Bank of England?*, Which? Ltd, <www.which.net/nonsub/cacampaigns/cpr/sep1996cpr/boe/contents.html>, accessed 20 June 1998.

Fane, G. 1998, 'The role of prudential regulation', Chapter 17 in McLeod, R.H. and Garnaut, R. (eds) 1998, *East Asia in Crisis: From Being a Miracle to Needing One?*, Routledge, London.

Folkerts-Landau, D., Takatoshi I., et al. 1995, *International Capital Markets: Developments, Prospects, and Policy Issues*, International Monetary Fund, Washington DC, August.

Goldsmith, R.W. 1969, *Financial Structure and Development*, Yale University Press, New Haven.

Hendrobudiyanto 1994, 'Bank soundness requirements: A central bank perspective', Chapter 9 in McLeod, R.H. (ed.) 1994, *Indonesia Assessment 1994: Finance as a Key Sector in Indonesia's Development*, Research School of Pacific and Asian Studies, Australian National University, Canberra, and Institute of Southeast Asian Studies, Singapore.

Horiuchi, A. 1998, 'Japan', Chapter 12 in McLeod, R.H. and Garnaut, R. (eds) 1998, *East Asia in Crisis: From Being a Miracle to Needing One?*, Routledge, London.

Johnson, C. 1998, 'Survey of recent developments', *Bulletin of Indonesian Economic Studies*, vol. 34, no. 2, pp. 61–6.

Krugman, P. 1998, 'Currency crises', <web.mit.edu/krugman/www/crises.html>, accessed 27 March 1998.

Kuo, S.W.Y. and Liu, C.Y. 1998, 'Taiwan', Chapter 11 in McLeod, R.H. and Garnaut, R. (eds) 1998, *East Asia in Crisis: From Being a Miracle to Needing One?*, Routledge, London.

Leung, S. and Doanh, L.D. 1998, 'Vietnam', Chapter 7 in McLeod, R.H. and Garnaut, R. (eds) 1998, *East Asia in Crisis: From Being a Miracle to Needing One?*, Routledge, London.

McKibbin, W.J. 1998, 'Internationally mobile capital and the global economy', Chapter 14 in McLeod, R.H. and Garnaut, R. (eds) 1998, *East Asia in Crisis: From Being a Miracle to Needing One?*, Routledge, London.

McLeod, R.H. 1998a, 'Indonesia', Chapter 2 in McLeod, R.H. and Garnaut, R. (eds) 1998, *East Asia in Crisis: From Being a Miracle to Needing One?*, Routledge, London.

—— 1998b, 'The new era of financial fragility', Chapter 20 in McLeod, R.H. and Garnaut, R. (eds) 1998, *East Asia in Crisis: From Being a Miracle to Needing One?*, Routledge, London.

—— 1999, 'Control and competition: Banking deregulation and re-regulation in Indonesia', *Journal of the Asia Pacific Economy*, vol. 4, no. 2, pp. 258–96.

Smith, H. 1998, 'Korea', Chapter 4 in McLeod, R.H. and Garnaut, R. (eds) 1998, *East Asia in Crisis: From Being a Miracle to Needing One?*, Routledge, London.

Song, L. 1998, 'China', Chapter 6 in McLeod, R.H. and Garnaut, R. (eds) 1998, *East Asia in Crisis: From Being a Miracle to Needing One?*, Routledge, London.

Walters, A.A. 1998, 'Exchange rate regimes for the future', Chapter 18 in McLeod, R.H. and Garnaut, R. (eds) 1998, *East Asia in Crisis: From Being a Miracle to Needing One?*, Routledge, London.

Warr, P.G. 1998, 'Thailand', Chapter 3 in McLeod, R.H. and Garnaut, R. (eds) 1998, *East Asia in Crisis: From Being a Miracle to Needing One?*, Routledge, London.

World Bank 1993, *The East Asian Miracle, Economic Growth and Public Policy*, Policy Research Report no. 91/93, Oxford University Press, Oxford.

10 Australia and the crisis in East Asia: The role of theory in the formation of policy

Steven Kates

Much of the depth of the crisis in the economies of Asia is attributable to a failure of economic theory. The Asian region is now going through a singularly difficult period whose fault is in substantial part due to the economic models adopted by the various policy advisers.

This paper will concentrate on two economies in particular, Australia and Japan, because they provide polar examples of the use made of fiscal policy in dealing with recession. In one case, Australia is an economy that has gone from strength to strength even with each and every one of its major regional trading partners having experienced large-scale economic difficulties. In the other, Japan has turned a mild cyclical downturn into on-going recession whose end is nowhere in sight.

In one sense it is a tale of two economies, but in a more important sense it is a tale of two economic theories — classical and Keynesian.

10.1 The failure of Keynesian theory

It is the very conception of how economies grow that divides modern economics from the economics of the pre-Keynesian period. In pre-Keynesian times, concern was seldom, if ever, expressed about the lack of demand in aggregate. Demand deficiency was never seen as a relevant issue, so that when an economy went into recession, no one looked to it to understand the problem or to develop a solution.

Recessions were rather seen as related to structural problems. In understanding the causes of economic downturn, the fundamental problem was seen as a failure on the part of producers to anticipate correctly the demands of those who were expected to buy. Production was, therefore, not understood as having been excessive relative to demand, but was instead comprised of the wrong assortment of goods and services.

The failure to sell was thus not due to a lack of demand in general, but to a lack of demand for the particular goods and services produced.

This is a difference of the most fundamental importance. If you believe that the problem is lack of demand then the obvious solution is to do what is required to increase the level of demand. A practical solution to recession would then be to increase the level of public expenditure in order to increase the level of activity. This is where the usual assortment of Keynesian approaches is introduced, which would encompass increases in public works and reductions in the level of taxation.

If, however, one conceives of the problem in terms of structure, so that the problem is seen as due to a failure of those producing to correctly anticipate what others would like to buy at prices that cover all costs of production, then public spending and tax cuts are inappropriate. Because the problem is structural, the solution is structural readjustment.

Indeed, if the problem is structural, public spending will make the problem worse in at least two ways. First, it will prolong the period of adjustment by providing income to those who have misjudged where demand actually lies. By pumping purchasing power into the economy, demand is created for some portion of the goods that could not previously be sold with the result that resources are held where their return is actually negative. Public outlays help to disguise the need to readjust and slow the process down.

But second, by increasing the level of public spending, the economy is distorted even further from its equilibrium growth path. Public spending is virtually certain not to comprise just those goods and services that would be produced in a free-exchange economy. Instead, a vastly different array of goods and services are produced, which then leads to the evolution of a less productive economic structure than would have been the case had market forces been allowed to work themselves out.

It is not just the final goods and services produced, but the intermediate goods and services that go into the production process that are also distorted by public spending. The result is that the greater the level of public spending, the more entrenched this sub-optimal structure becomes and the more difficult it becomes to reverse the process and return to a settled growth path.

Is this just theory? It is not. The experience of the Japanese recovery has been before our eyes for nearly a decade. It has shown the consequence of public spending and growing deficits. What the Japanese have done is demonstrate that classical theory has more to offer than Keynesian. Keynesian policies have not

worked. They have made a tragic mess of the Japanese economy, which has gone deeper into recession the more relentlessly Keynesian policies have been applied.

10.2 Australia and classical theory

And importantly, the contrary experience of Australia has involved applying classical remedies to its economy, resulting in the most extraordinary return to growth. It is part of the classical canon that when recessions occur, part of the recovery process is to reduce public spending and allow the economy to go through its period of readjustment as quickly as possible.

This has been the policy adopted in Australia. A massive deficit has been turned into a massive surplus in the space of three years, almost entirely through reductions in public spending. If one applies the standard Keynesian model to so thorough a reduction in the level of spending, the result is a fall in output and a rise in unemployment.

It should be borne in mind that these reductions in public spending did not occur after the economy had begun to grow strongly, but in advance. It was not the supposed Keynesian type approach in which the stimulus is removed once the economy has begun to recover. Instead, public spending was cut well in advance of any upturn. If there were any validity to Keynesian theory, the consequence of this massive removal of fiscal stimulus would have been a sharp downturn in activity and a large rise in unemployment.

Instead, the Australian economy went into overdrive. In spite of the tremendous fall in activity in Asia, the Australian economy has continued to grow at extraordinarily rapid rates. There has been no let up in this growth rate and most surveys of the Australian economy indicate that growth will continue into the foreseeable future.

10.3 Statistics

It is worth having a look at some statistics to have some indication of the kind of growth Australia has achieved. The national accounting data covering the four quarters to March 1998 are shown in appendix 10A.

There was a huge reduction in public spending in the Australian budget handed down in August 1996. This fall in public spending is reflected in the data on government consumption, which rose by only 0.4 per cent, and in the figures for public gross fixed capital expenditure, which fell by 3.9 per cent over the year.

Clearly, the economy was receiving no stimulus from higher levels of public spending.

Moreover, it is clear that the balance of payments had been severely damaged by the loss of trade into Asia. The accounts data show exports rose by only 2.8 per cent over the year, while imports rose 14.2 per cent. It is thus unmistakable that overseas sales, far from having contributed to growth, actually held growth back.

Nevertheless, GDP rose over this period by an amazing 4.9 per cent and, concentrating only on the non-farm sector, economic growth was 5.3 per cent. This would be an extraordinary performance, even in the best of times. That it occurred during a period when the Australian economy was surrounded by a large scale economic downturn amongst its major trading partners in Asia requires explanation. It is far from obvious why the combination of a cut in public spending and major recessions within the economies of its trading partners should, nevertheless, coincide with such robust growth in Australia.

Yet this is only a conundrum if one believes that public spending helps maintain growth, while reductions in public spending subtract. If one instead starts from the presumption that cuts to unproductive public spending actually free resources for use in more productive activities then the answer is readily to hand. The cuts to spending have created opportunities for growth that would not otherwise have appeared and the momentum provided by lower levels of public spending has driven the economy forward. Public spending had not been a stimulus, but had been an anchor holding growth back. Removing this wasteful public spending actually hastened growth.

This is the mirror image of what has taken place in Japan. The Japanese increased public spending and the economy has fallen apart. Australia has reduced its level of spending and the economy has taken off into rapid rates of growth.

This is an occurrence so far from modern economic thought that it simply will not stick in some minds. It is, to some ways of thinking, an impossibility that higher government spending should lead to lower levels of activity while lowering the level of government spending should cause the economy to take off. Yet, the evidence is before our eyes. What is missing is an ability to understand why this is so.

Pre-Keynesian economic theory understood these issues extremely well. There would have been no mystery to any economist educated prior to the publication of Keynes's *General Theory*. It would have been clear beyond question that the problem in Japan was not too little demand nor that an appropriate policy response would involve higher levels of unproductive public spending. It would have been understood that the solution administered in Japan would lead more or less directly

to the situation in which Japan now finds itself — mired in debt and unable to raise its level of activity towards anything like its previous rates of growth.

10.4 Historical framework

The essence of the issue at stake is the validity of a proposition once called the law of markets or in more modern terms, Say's Law. The Keynesian Revolution was addressed towards refuting Say's Law, which Keynes characterised by the phrase 'supply creates its own demand'. Say's Law was said by Keynes to mean that everything produced would be sold and, therefore, recessions were an economic impossibility. Keynes's *General Theory*, therefore, introduced demand deficiency into economic theory and in this way, at least according to Keynes, provided economics with its first ever explanation for involuntary unemployment.

This idea that classical economists, for a century and a half, failed even to notice that they had no theory to account for involuntary unemployment is so far-fetched that it is worth perhaps just providing a reminder of what the actual words in the most influential economics text of the twentieth century actually were. After defining involuntary unemployment, Keynes went on to state:

> So long as the classical postulates hold good, unemployment, which is in the above sense involuntary, cannot occur. Apparent unemployment must, therefore, be the result either of temporary loss of work of the 'between jobs' type or of intermittent demand for highly specialised resources or of the effect of a trade union 'closed shop' on the employment of free labour … Obviously, however, if the classical theory is only applicable to the case of full employment, it is fallacious to apply it to the problem of involuntary unemployment — if there be such a thing (and who will deny it?) … We need to throw over the second postulate of the classical doctrine and to work out the behaviour of a system in which involuntary unemployment in the strict sense is possible. (Keynes, *General Theory*, pp. 16–7)

That demand failure was the issue Keynes wanted to raise is also not in doubt. This is how it was stated, also in the *General Theory*:

> The idea that we can safely neglect the aggregate demand function is fundamental to the Ricardian economics, which underlie what we have been taught for more than a century. Malthus, indeed, had vehemently opposed Ricardo's doctrine that it was impossible for effective demand to be deficient; but vainly. For, since Malthus was unable to explain clearly (*apart from an appeal to the facts of common observation*) how and why effective demand could be deficient or excessive, he failed to furnish an alternative construction; and Ricardo conquered England as completely as the Holy Inquisition conquered Spain. (Keynes, *General Theory*, p. 32 — emphasis added)

It would be an interesting question indeed, comparing Australia and Japan, to ask where 'an appeal to the facts of common observation' would lead us today. We have

seen the expenditure policies applied in Japan and we have seen them fail miserably. And it was the role of Say's Law to specifically warn policy makers that this would be the case. Say's Law specifically tells economists and policy makers that whatever might have caused recession and unemployment, it would never have been deficient aggregate demand. Therefore, no alleviation of recession can ever come from unproductive expenditure, which will, in fact, make conditions worse rather than better.

These issues are dealt with in *Say's Law and the Keynesian Revolution* (Kates 1998). To understand the Keynesian Revolution and the harm it has done, it is necessary to go back through the development of Say's Law and to recognise the important role it played prior to the publication of the *General Theory*. Therefore, on one level, it is a text on the history of economics, but it is also of modern relevance detailing why Japanese fiscal policy would inevitably fail.

The disappearance of Say's Law has deprived policy makers of one of the most profound economic conclusions ever reached — that demand deficiency is never the cause of recession and demand stimulation is never a valid means to cure recessions when they occur.

Do you want proof that the classics were right and Keynes and his modern followers were wrong? First, visit Japan and then Australia. The one is in recession and the second has recorded a massive improvement in growth, even with every one of its major regional trading partners in recession. How does one account for this other than by recognising that Keynesian economic theory, which remains integral to modern macroeconomic thought, is utterly wrong, while classical economic theory, with its acceptance of Say's Law, is overwhelmingly right.

10.5 Policy advice

The kind of advice the Japanese are receiving, of course, is straightforwardly Keynesian. It is built on models for which aggregate demand is a feature part of the system. Whether it is based on some primitive Keynesian-cross analysis or built on an IS-LM model or is derived from bringing together an aggregate demand curve with an aggregate supply curve, it all comes to the same thing. The conclusion reached is that the recession in Japan can be overcome through higher levels of public spending.

Stanley Fischer, the First Deputy Managing Director of the IMF, provided a very clear statement of the basic impulse underlying this view. What he said was this:

> Japan's economic performance is of course a matter of grave domestic concern. But given the prominent role of Japan in the world economy, and especially in Asia, it is

also a legitimate matter for concern by Japan's neighbours and by the international community. *There is little disagreement about what needs to be done. There is an immediate need for a substantial fiscal expansion.…*

On fiscal policy, *the recent suggestion of a package of 16 trillion yen, about 3 per cent of GDP, would be a good starting point.* But, unlike on previous occasions, the program that is implemented should be close to the starting point. *The well-known reservations about increases in wasteful public spending are correct:* that is why much of the package, at least half, should take the form of tax cuts. Anyone who doubts the effectiveness of tax measures need only consider the effectiveness of last year's tax increases in curbing demand.

The IMF is not famous for supporting fiscal expansions. And it is true that Japan faces a long-term demographic problem that has major fiscal implications. *But after so many years of near-stagnation, fiscal policy must help get the economy moving again. There will be time to deal with the longer-term fiscal problem later.* (Fischer 1998 — emphasis added)

It is as worrying where he states that there is 'an immediate need for substantial fiscal expansion' as where he states 'there is little disagreement about what needs to be done'. There is indeed little disagreement and that is itself a very great worry. Economist are (almost) all Keynesians now. Take this example from an editorial in *The Economist* of 28 February 1998 under the heading 'Fiscal paralysis':

The [Japanese] government says it cannot afford a big stimulus because its finances are perilous. It is true that Japan's gross public debt has risen to 87% of GDP, but net debt amounts to only 18% of GDP, the smallest among the G7 economies. The general-government budget deficit, 2½% of GDP, is smaller than its European counterparts'. Rightly, the Japanese are worried about the future pension liabilities implied by their rapidly ageing population. But now is not the time to sort the problem out. Far better to cut the budget later, when the economy has recovered its strength.

Some advice! And both take the same view that Japan should spend now and clean up the mess after. In Stanley Fischer's view, 'there will be time to deal with the longer-term fiscal problem later'. *The Economist* thinks 'now is not the time to sort the problem out. Far better to cut the budget later, when the economy has recovered its strength'. Ignore all danger signs, they argue and just plunge ahead. How bad can any set of advice be, but given the economic models from which they work, it all makes sense to them. The pity of it is that their models are completely useless in making heads or tails of what is actually going on and what needs to be done to fix things up.

10.6 Closer look at Australia

That is why recognising how well Australia has done with a very different set of policies is so instructive.

As part of the Australian Chamber of Commerce and Industry's (ACCI) regular quarterly *Survey of Investor Confidence*, a number of questions were added to gauge more accurately the effect that the Asian crisis would have on Australia. From the data, there is every reason to believe that for firms that are dealing directly with Asia there will be important consequences, but for Australia overall it should be able to ride through this storm without excessive damage to the economy. These results are shown in appendix 10B. What these survey results show is the following:

- For firms that sell directly into Indonesia, Thailand, Malaysia, Singapore and Korea, there is expected to be a big fall in actual and expected sales.
- For firms selling into Japan, there is expected to be a fall in sales and big fall in expected sales, but not to the same extent.
- In Australia there is some sign that a slowing is occurring, but it is not excessive. In terms of the national economy, businesses are somewhat more pessimistic than they were, but not greatly so.
- Most interestingly, the tables containing the results for firms that trade directly with Asian economies, in comparison with the full survey results, show that for economic activity in general, investment and full-time employment, those firms trading directly into Asia are very pessimistic while for the survey in general the results can only be described as optimistic. It is noteworthy that the full survey results contain the results from those firms trading directly with Asia.

Thus, Australia has a very good chance of coming out of the Asian crisis relatively unscathed, although not untouched. The national accounts show growth in aggregate continues, but there are also clear problems with the balance of payments. The ACCI forecasts indicate that there will be a slowing of the Australian economy, but hardly a recession. The survey results indicate a rebound in 1999-2000.

10.7 Policy conclusions

This is a story of two economies that took different policy routes to deal with recessionary conditions, with the main difference being the manner in which fiscal policy was employed. Japan, following Keynesian advice, raised its level of spending, which has led to contraction and stagnation. Australia, taking an approach consistent with pre-Keynesian doctrine, cut its level of spending, which has led to the most robust recovery seen in years, and which has occurred even with the Asian crisis affecting so many of its major regional trading partners.

One cannot make sense of either result using standard Keynesian models which are the models used almost universally throughout the developed world for policy formation. The contention of this paper is that it is standard economic theory that is

wrong and that is leading to the development of policies that, when adopted, make matters worse. Japan is wrecking its economy based on what ought to by now be a discarded and discredited economic theory.

Instead, Keynesian remedies continue to be applied and continue to provide no success. But until those policies are reversed, Japan will continue to flounder and any recovery it does manage to achieve will be limited at best. It is time to toss out the Keynesian economic analysis, which has dominated economics for the past half century, and to take a closer look at the theories of the cycle that these failed Keynesian theories have replaced.

Appendix 10A

Table 10.1 **National accounts, four quarters to March 1998 (per cent)**

Private Consumption	4.5
Government Consumption	0.4
Private Gross Fixed Capital Expenditure:	
• Dwellings	12.5
• Non-Dwelling Construction	-3.2[a]
• Equipment	14.4[a]
• Total Private	6.4[a]
Public Gross Fixed Capital Expenditure:	
• Public Enterprises	-13.7[a]
• General Government	9.1
• Total Government	-3.9[a]
Total Gross Fixed Capital Expenditure	7.3
Domestic Final Demand	4.5
Gross National Expenditure	7.6
Exports	2.8
Imports	14.2
GDP(Average)	4.9
Non-Farm Product	5.3

[a] Adjusted for sales from the public sector to the private sector.

Source: ABS(*Australian National Accounts: National Income, Expenditure and Product*, Cat. no. 5206.0.)

Appendix 10B: ACCI Survey on the Effects of Instability in Asia on the Australian Economy, January 1998

Indonesia, Thailand, Malaysia, Singapore or Korea

What has happened to your sales to these Asian economies during the past three months?

	%
Actual Change in Sales Volume	
Large increase	3.8
Small increase	9.2
No change	25.2
Small reduction	25.2
Large reduction	36.6

What do you expect to happen to your sales to these economies during the next twelve months?

	%
Expected Change in Sales Volume	
Large increase	2.3
Small increase	12.1
No change	10.6
Small reduction	28.0
Large reduction	47.0

Japan

What has happened to your sales to Japan during the past three months?

	%
Actual Change in Sales Volume	
Large increase	1.1
Small increase	12.4
No change	36.0
Small reduction	36.0
Large reduction	14.6

What do you expect to happen to your sales to Japan during the next twelve months?

	%
Expected Change in Sales Volume	
Large increase	2.2
Small increase	16.5
No change	23.1
Small reduction	39.6
Large reduction	18.7

National economy

Because of the economic instability now occurring in Asia, in what way do you expect to change your own level of production during the next twelve months?

	%
Expected Change in Own Level of Output	
Large increase	0.6
Small increase	7.3
No change	59.2
Small reduction	27.1
Large reduction	5.8

What do you expect to happen to the level of economic activity in Australia during the next twelve months as a result of the economic instability now occurring in Asia?

	%
Expected Change in National Output	
Small increase	9.1
No change	10.0
Small reduction	66.3
Large reduction	14.7

Comparison of Full Survey Results with Firms Affected by Asia

Economic activity

	Affected by Asia	Full Survey
Expected Level of Activity		
Much higher	0.0	2.3
Somewhat higher	27.1	39.0
About the same	36.5	43.3
Somewhat lower	29.4	14.1
Much lower	7.1	1.4

Investment

	Affected by Asia	Full Survey
Expected Level of Investment		
Much higher	5.9	4.9
Somewhat higher	18.8	21.9
About the same	38.8	53.0
Somewhat lower	28.2	14.3
Much lower	8.2	5.8

Full-time employment

	Affected by Asia	Full Survey
Expected Number of Employees		
Much higher	2.4	2.1
Somewhat higher	5.9	16.1
About the same	51.8	62.2
Somewhat lower	36.5	18.8
Much lower	3.5	0.9

References

Fischer, S. 1998, The Asian Crisis, the IMF, and the Japanese economy, Paper presented at the Asahi Shimbun symposium "The Asian Economic Crisis and the Role of Japan", Tokyo, 8 April.

Kates, S. 1998, *Say's Law and the Keynesian Revolution: How Macroeconomic Theory Lost its Way*, Edward Elgar, Cheltenham.

Keynes, J.M. 1936, *The General Theory of Employment, Interest and Money*, Macmillan, London.

SECTION 4

INDUSTRY STRUCTURE AND POLICY IN AUSTRALIA

11 The future of Australian motor vehicle manufacture

*Harry Clarke, Darcy McCormack and Joseph Sunderland**

Future economic prospects of the Australian passenger motor vehicle (PMV) manufacturing industry are examined using 'new trade theory'. While recognising the need for improved cost efficiencies through expanded plant capacities, this paper emphasises the importance to PMV producers of operating in design niches that reflect specific Australian advantages. These lie in the production and design of 'upper-medium' type vehicles. The Australian PMV industry is characterised by highly skilled human capital, well-developed design capabilities and relatively low labour costs. This gives it an advantage over producers whose success is mainly dependent on production cost advantages and provides the basis for an Australian push into new export markets.

11.1 Introduction

This paper discusses economic prospects of the Australian passenger motor vehicle (PMV) industry. It highlights expected developments and analyses strategies for industry viability in a competitive, global context.

The Australian PMV industry displays a number of paradoxical attributes. While there is well-recognised excess capacity globally and industry protection has been substantially reduced in Australia, plans for record investment have been announced recently by firms in the Australian industry. The four local producers will spend $4 billion over the next few years. This follows a decade of major industry rationalisation, which has seen a substantial reduction in the number of Australian PMV producers, models and assembly plants. Improving profits and rising

* None of the views expressed here necessarily represents the views of any of the authors' employing institutions. Initial work on this topic was done while Joseph Sunderland was a Research Assistant in the School of Business at La Trobe University. The authors acknowledge the financial assistance provided by a School of Business Research Grant and thank Julie Rowe for assistance.

investment in the face of tariff cuts reflect the enhanced performance of the Australian industry. Overall industry profits topped $517 million in 1997. Exports in 1996 were about 40 000 vehicles.

The steady reduction of tariffs, the continuing development of low-cost vehicle producers in Asia and increasingly competitive conditions in Australian-supplied markets have clarified where Australia's comparative advantages lie. Current 17.5 per cent tariffs will fall to 15 per cent by 2000 and, although there will be a tariff freeze for the following five years, the Government has indicated it will legislate to cut tariffs to 10 per cent in 2005. While, with industry rationalisation, an increasingly optimistic future for the Australian industry can be envisaged, there are global and domestic factors that temper this optimism. This paper examines the strategic rationale for industry behaviour using new theories of trade and investment.

Section 11.2 below discusses industry structure and the role of costs and market focus in ensuring commercial viability of the Australian PMV industry. Section 11.3 applies 'new trade theory' to the industry while Section 11.4 sets up strategic guidelines. Section 11.5 examines strategies and prospects individual producers, while Section 11.6 summarises industry prospects. Section 11.7 discusses future government policy toward PMV assembly. Section 11.8 draws conclusions.

11.2 Industry structure and strategy

With substantially reduced protection, the Australian PMV industry can now be viewed as a part of a global industry dominated by multinationals and subject to some price competition. The combination of product differentiation and widespread excess capacity gives the global industry the appearance of monopolistic competition. However, the Australian component of this structure has significant market power with respect to an important part of total Australian motor vehicle demand (the 'upper-medium' vehicle segment). Hence, the Australian PMV industry is a differentiated oligopoly, with only a few firms competing partly by price, but more importantly via product differentiation.[1] This gives individual firms in

[1] VFACTS (1995) specifies five segments of the industry: the micro car segment, the small car segment (eg Toyota Corolla); the medium segment (eg 4-cylinder versions of the Toyota Camry); the upper-medium segment (eg Holden Commodore, Ford Falcon and 6-cylinder versions of the Camry and Magna) and the luxury segment (eg Holden Statesman/Caprice and Ford Fairlane/LTD).

Australia market viability which is exploited by designing vehicles for specific consumer tastes, thereby occupying particular market niches.[2]

Product differentiation implies that PMV assemblers have some ability to exploit market power. Their ability to eventually succeed in doing so relies on meeting niche market demands and production cost efficiencies. Two major strategies available to multinational PMV corporations are cost and focus strategies. These are now discussed for the Australian industry.

Cost strategy

With substantially lower Australian tariffs, the ability of local PMV producers to be cost competitive is critical. Emerging producers in less developed countries, such as Korea, have lower labour costs than Australian producers, while established producers in developed countries, such as Japan and the United States, enjoy somewhat higher productivity because of greater scale economies. A variety of producers can therefore supply the Australian market with competitive vehicles. However, evidence on labour costs and productivity indicates that local producers are reasonably competitive (appendix 11A).

Increasing output from existing assembly plants is a widely advocated approach to reducing unit costs. Australian producers are often seen to operate at the low volume end of the cost-efficiency scale. Consequently, increasing plant volumes is important if local producers are to compete successfully with imported vehicles. Attempts to realise scale economies in Australia have resulted in industry rationalisation and a concentration in the 'upper-medium' vehicle segment. As competition for market share has become more important for producers than seeking very distinctive market niches, most producers have concentrated on providing larger vehicles.

However, the importance of cost efficiencies should not be overemphasised, since most Australian plants operate close to cost-minimising scale (assumed around 100 000 vehicles annually). While some cost advantages would stem from further volume increases, such cost reductions would be small relative to the effect of other parameters affecting costs, such as tariffs or local taxes (appendices 11B and 11C). They would also be small relative to brand loyalty effects and the premiums paid for

[2] As Markusen et al. (1995, p. 205) state, 'automobile models vary in size, horsepower, comfort, performance and appearance … If preferences are heterogeneous, so that imported goods and similar domestic goods are seen by consumers as imperfect substitutes, the products can command different prices and profit markups in various markets. This is because each variety of a product will face a distinct demand curve in each market, allowing its producer to act with some market power'.

vehicles closely meeting individual model preferences. While some emphasis should be on technical innovation, improved work practices and scale economies, it is also important to meet market niches with an appropriate 'focus' strategy.

Focus strategies

While some cost efficiencies can be achieved by producing only a single model, there is then the risk that producers will fail to match consumer preferences by producing an insufficiently diversified range of vehicles. Following the Button Plan, specialised production plans have been strongly pursued in Australia. However, this has recently changed with some producers now moving toward *mixed plans,* which seek to supply different market segments by producing more than one vehicle type.

The economics of the case for mixed plans are straightforward. If different consumers seek different vehicle types then production plans must tradeoff technology-determined economy of scale cost advantages against the premium consumers will pay for vehicles closely matching their preferences. An *optimal degree of product diversity* must be found. A further advantage of diversification is that, with uncertainty in the identification of preferences, producers offering a product range can diversify away from products that are poor market performers.

The ability of firms to recognise and respond to different consumer tastes in different markets is their *focus* strategy. This concentrates on identifying tastes of potential buyer groups and targeting such groups with specific product lines that are differentiated and that meet different niche demands in separate markets without creating cost inefficiencies.

Models must be distinct to gain market power, but must attract large enough markets for viability. In the PMV industry, this tradeoff is often resolved by designing a distinct model to capture a local market segment. This distinct model is then often adapted to accommodate the specific tastes of export markets. For example, Holden developed a special hybrid short-wheelbase Statesman/Commodore 6-cylinder model for luxury vehicle buyers in Asia. By not involving production of entirely new models, these model variants meet distinct niche demands enabling producers to diversify at low cost.

Although Australian tastes are generally geared towards larger vehicles, it is widely believed Asian preferences are primarily for smaller vehicles (eg IC 1997a, p. 196). On the other hand, North American tastes are mainly for larger vehicles. In terms of creating differentiated demands in markets where tastes are similar to those in

Australia and hence where economies of scope seem greatest, the North American market is potentially a better long-term prospect than Asia.[3]

Asian consumer preferences have converged somewhat toward developed country preferences. Therefore, the assumption that Asian demands will always favour small cars is questionable.[4] Higher protection in several potential Asian markets provides an additional strong reason for the targeting of non-Asian markets (see table 11.5 in appendix 11B at the end of this paper). Trade liberalisation under APEC, in the longer term, may reduce Asian barriers to Australian exports.

11.3 International trade and the PMV industry

By producing for both domestic and export markets with similar tastes, firms enhance scale economies. The ability of local PMV manufacturers to expand exports will rely increasingly on lowering costs and meeting these new market niches.

Trade and economies of scale

Economies of scale exist in the PMV industry. New trade theory shows that such economies can imply a rationale for international trade not based on traditional comparative costs. Countries may establish a base in producing a niche vehicle product because of *first-mover* advantages, despite lack of comparative advantage. This explains why trade can occur between countries *all of which* produce PMVs. As long as scale economies are present, there are potential gains from production specialisation with international trade, even if PMV prices in different countries were the same under autarchy.

Even if vehicle types are highly differentiated, international trade allows scale economies and production specialisation. Imports meet the demand for vehicles not produced locally (eg small cars in Australia), thereby generating intra-industry trade. Trade based on economies of scale can therefore explain why Australian car makers can sensibly aim to increase exports even without comparative cost advantages.

---

[3] This is supported by the IC (1997a, p. 196): 'A high proportion of vehicle demand in Asia is accounted for by small passenger cars … Hence, there appears to be limited export demand in these markets for the vehicles traditionally produced in Australia (upper-medium vehicles with comparatively large engines)'.

[4] The IC (1997a, p. 196) argue that 'There may be some niche market demand in Asian countries. Given that large vehicles are by-and-large not produced in these countries (excluding Japan), this demand must be filled almost entirely by imports from Europe, Japan, North America and Australia'.

Trade and industry rationalisation

The existence of scale economies implies that exporting Australian vehicles and importing foreign-produced vehicles helps rationalise the local PMV industry by encouraging local specialisation. Such specialisation is the key to an efficient and internationally competitive Australian industry. Past high tariff protection encouraged firms to become inward-looking and to produce a wide range of vehicles in small volumes. The 1984 Button Plan saw the Australian PMV industry as small and excessively diversified. Rationalisation called for fewer producers, assembly plants and models. Increased exports would enable more focused production plans and greater competitiveness.

Thus, in the early 1980s, there were five Australian vehicle producers manufacturing twelve models in eight assembly plants. Exports were only around 2 500 vehicles annually. By 1996, the industry had four car makers, five models and four assembly plants with exports of 40 000 vehicles.

This shift in focus of the Australian PMV industry is towards specialisation in production of larger, high-power vehicles, with smaller vehicles being imported from Korea and Japan. Increased trade in PMVs has helped rationalise the Australian industry. Local car makers have focused on lowering costs through higher production of fewer vehicle varieties.

Trade and pro-competitive gains

With scale economies, cost-efficient local markets will only support a few firms and be imperfectly competitive. Therefore, a reason for opening up the industry to international trade is to promote efficiency by introducing competition. These are the pro-competitive gains from trade (Markusen et al. 1995, p. 181). Such gains take the form of expanded vehicle outputs and lower prices, the price reductions reinforced by the decreasing average costs associated with increasing returns. Competition is increased by exposing local markets to greater international competition.

Importantly, the existence of pro-competitive gains makes it clear that, from society's viewpoint, increases in profitability are not an inevitable feature of an improved PMV market environment. Losses, and indeed firm exit, may be an essential by-product of generating such gains.

Trade and firm exit

Through trade, the total number of firms competing internationally can increase while the number of firms producing in an individual country decreases. Over the past fifteen years, Australian PMV producers have increased their export (and import) activities. Although the industry has come to involve fewer players during this period (Leyland ceased assembly in 1984 and Nissan in 1992), more competition has occurred because of increased trade. The exit of high cost firms signals rationalisation, not necessarily future problems for the industry.

Trade and product diversity

Given scale economies, few countries can manufacture unique vehicles for every class of consumer. Instead of trying to create 'compromise' vehicles (such as in Australia, the medium-sized four-cylinder Camry), trade allows international specialisation with designs more closely reflecting individual preferences. For example, Australian produced vehicles have two major (non-price) characteristics — their size and their power. For illustrative purposes, suppose there are three main vehicle types — a small, low-power vehicle (eg the Mirage), a mid-sized, medium-power vehicle (eg the Toyota Camry) and a large, high-power vehicle (eg the Ford Falcon). Australian car buyers may prefer small low-power (low cost) vehicles or large high-power (high cost) vehicles depending on whether they are low or high income earners.

Without international trade, if Australia provided high and low-income earners with their 'ideal' car by producing both Falcons and Mirages, the volume of sales of each model would be lower and the average cost would be higher than for a single 'compromise' model (the Camry). However, if Australia produced only the Camry, high and low-income earners would not consume their ideal car. High-income earners would buy a smaller, lower-powered car than their ideal, while low-income earners would have access only to a vehicle that is more expensive and larger than their ideal.

With trade between Australia and a second country that produces small low-powered cars (eg Mirages produced in Japan), Australia can now specialise by producing only Falcons, which it also exports to Japan, and by importing Mirages from Japan. Each country is more specialised, so that even if average costs did not fall and vehicle prices remained the same, consumers would enjoy gains from trade due to increased product diversity. In fact, average costs are likely to fall because of scale economies arising from production of a single Australian model.

As noted above, Australian consumers generally select larger vehicles (appendix 11C). Such vehicles have uniqueness and have been produced under protected conditions for the Australian market over many years. Local producers have acquired 'learning-by-doing' advantages in their production, particularly in terms of product design and development. Even Mitsubishi and Toyota seem to have recognised the case for such specialisation in Australia, as they transform the Magna and Camry respectively into 'larger' vehicles by emphasising six-cylinder, rather than the original four-cylinder, versions of these vehicles. Evidence supports this shift in focus. Between 1993 and 1995, combined domestic sales of the Magna 4 and Camry 4 fell 20 per cent, whereas sales of the Magna 6 and Camry 6 increased 40 per cent (DIST 1995, pp. 15–6). Mitsubishi earlier this year discontinued production of the Magna 4.

Trade and excess capacity

Over the past decade, new PMV sales in Australia have been fairly static. Total sales in 1997 of 722 427 vehicles, despite an increase of 11 per cent over 1996, were only 4 per cent higher than in 1985 (DIST (1995, p. 13), Lynch and Saunders (1998)). It seems the Australian market has limited growth potential over the next decade or so. Thus, major growth in vehicle sales for Australian producers must come through exports.

Global PMV markets are characterised by substantial (and increasing) excess capacity. While demand in most markets is stable, PMV production is increasing, creating more excess capacity and exerting downward pressure on prices. Both North America and Japan have 3 to 4 million units of surplus vehicle capacity, while Europe has between 5 and 6 million units of excess capacity (Keller 1996, p. 14). New firms entering the industry, particularly in LDCs (eg Korea, Malaysia and India), exacerbate this situation. Much of the growing LDC capacity is export-oriented.

Australian producers have announced plans to expand capacity and increase exports. However, a constraint on this expansion is the global excess capacity. Australian producers offering vehicles that are close substitutes to those in excess supply will face very competitive conditions. Successful trade strategies require differentiation towards a design niche that reflects Australian advantages. This is most likely in the medium-large vehicle segment.

Trade and set-up costs

Australian producers attempting to penetrate new export markets face establishment costs associated with marketing and distribution. Consider Mitsubishi, for example. To succeed with its Diamante (Verada) export program to the United States, Mitsubishi needs to spend heavily on advertising and marketing. Consumers must be well informed about the vehicle, including after-sales service and be assured of the availability of spare parts. If they feel the vehicle may only be available for a limited time, their expectations may be that spare parts will be expensive and resale values low. Marketing efforts are vital to offset these beliefs and to establish commercial viability.

There are also high distribution costs. Even though the Diamante is distributed through Mitsubishi's dealer network in the United States, with so few vehicles targeted for the United States market, distribution costs per unit are high. Such costs must be accounted for by producers when considering which export markets are viable. A concentrated focus has advantages over a more diffuse one. The 'fit' of the model with the dealer's profile and existing range is also important.

The Mitsubishi example highlights some issues in formulating an export strategy. Advertising and marketing costs need to be optimised by targeting only a few plausible markets. Product distribution costs are subject to scale economies and are lowered by exporting volume. Net returns can be optimised by focusing on markets where parent firms already have a well-established dealer network and reputation.

In short, when targeting a market for vehicle exports, a producer must not only identify consumer tastes and contemplate marginal export expansions. It should also be concerned with consequent fixed set-up costs which imply the need for a focused effort.

11.4 Strategic guidelines for industry

PMV firms must optimise the tradeoff between maximising economies of scale and meeting tastes through product diversity. This is a key to being globally efficient. Constraints imposed by this tradeoff are weaker if a viable export market has been developed and if plant technology allows economies of scope via the production of model variants on the same production line.

Most Australian firms now forecast increased productive capacity over coming years and will thereby achieve further scale economies. Moreover, the ability of producers to successfully pursue design niches is enhanced by the Australian industry's substantial design capabilities. Indeed, some global vehicle makers are

successfully using their Australian operations to provide design services for overseas parents and for related foreign companies. These parent firms recognise the strength of Australian producers in design, engineering and innovative manufacturing.[5] Holden has successfully established itself as a design and development centre for General Motors in the Asian region, with considerable skills and capability built up from adapting overseas designs to Australian conditions.[6] Toyota has recently announced plans to set up a design centre in Melbourne and Ford Australia was recently recognised by then Ford Global Operations Manager, Jac Nasser, as having the most efficient design centre in Ford.[7]

One means for lowering costs while addressing national tastes is through multiple (related) model production to realise economies of scope. In Australia, there are five basic models being produced by four firms. If firms can produce related models, they gain economies of scope and production flexibility with respect to uncertain tastes — they can diversify risk associated with uncertain demands. While such measures contradict an original objective of the Button Plan — to reduce the number of locally-produced models — they have 'product niche' and economy of scope advantages.

By supplying a larger market, producing for export enables firms to boost scale economies, but risks sacrificing a domestic design niche. This risk is reduced if export markets have similar tastes to domestic markets. International trade in

---

5 'Australia's design and manufacturing capability is based on a strong R&D base and a training and education system that has produced a skilled workforce … The automotive skills developed … give Australia a particular advantage over Asian countries as a technology base … The fact that a number of multinational companies have recently given their Australian subsidiaries responsibility for design and production of certain products, or nominated them as their technical bases for the Asian region, supports this conclusion'. (IC 1997a, p. 323)

Ford Australia also praised the design capabilities of the Australian industry: 'The industry is the leading volume manufacturer of a complex product … The industry has powerful technology linkages to world centres of design and production … [It] is a 'University' of advanced manufacturing methods and technologies' (IC 1997a, p. 437). Also from Ford: 'one of the key advantages Australia has as a manufacturing location for Ford is its unique product development capability. Australia is one of the few countries in the Ford world that has a unique capability of being able to design and engineer a car from a clean sheet of paper'. (IC 1997a, p. 317)

6 Holden comments: 'Australian design and development of passenger motor cars is now very competitive with overseas. When the Export Facilitation benefit is taken into account Australia becomes a relatively attractive prospect as an automotive design location' (IC 1997b, footnote 11).

7 Nasser states: 'We now believe that the tool room in Geelong is the most efficient in the world and that includes our German tool rooms. It is bidding for business around the world and winning it' (Fulcrum-Analysis 1997b). Nasser was also indirectly quoted as saying that the turnaround had been such that Ford Australia management was now a benchmark within Ford Motor Company for its speed and efficiency in getting products and concepts to market.

manufactures, such as PMVs, is often more dependent on demand than factor endowments (Linder 1961). PMV production is motivated by the perception of potential demand and entrepreneurs will perceive such demands most accurately in home markets. Perceptions of foreign demands are more difficult due to barriers of distance, language and culture. A low risk strategy is to produce models that first meet domestic demand and are then exported to countries with similar demand patterns. Demand in the domestic Australian market should condition the direction of the export drive. Australian car makers should then target countries with similar tastes and per capita incomes. This entails market research.

In summary, the future strategic plans of Australian PMV producers should include:

- efforts to increase plant capacities to maximise scale economies;
- producing multiple (related) model variants to lower the risk associated with uncertain demands and to maximise economies of scope cost advantages;
- identifying the particular tastes in specific export markets and modifying vehicles to meet these tastes; and
- choosing export markets where motor vehicle tastes are broadly similar to domestic tastes and where a strong, accessible dealership network exists.

11.5 Strategies and prospects of individual producers

Ford

Ford in recent years has committed itself to continued production in Australia through a $1.2 billion investment program over five years. $800 million of this was for development of the AU Falcon, with other expenditure aimed at expanding Ford's Broadmeadows operations. Ford's expenditure on developing the Falcon is substantial because, unlike the Commodore which draws upon several overseas GM designs, the Falcon is totally designed and manufactured in Australia.

Export plans

Ford produces a single, large-model vehicle and has relatively small export sales. In 1998, Ford exported 4 900 Falcons to New Zealand and South Africa. The recent abolition of tariffs on imported new cars to New Zealand will mean that the Falcon (and other Australian cars), which already enjoyed duty free entry to New Zealand under the *Closer Economic Relations Trade Agreement*, will now face increased competition from non-Australian cars in that market.

Although Ford is currently pursuing a domestic market niche, it is moving increasingly towards a more export-oriented strategy to reduce excessive reliance on the domestic fleet market. The recent announcement that approval will been given for additional expenditure of between $100 and $200 million to engineer the Falcon for left-hand drive suggests Ford is planning a serious bid at exporting the Falcon to the Middle East, Europe[8] and South America.[9] An alternative expansion option was for a second locally-produced model, although this now seems unlikely.

Domestic plans

Ford relies heavily on the Falcon for domestic sales. Until the release of the VT Commodore, the Falcon was Australia's top-selling vehicle. However, with more than 60 per cent of Falcon sales destined for the fleet market, Ford may face increasing pressure for fleet sales from the Camry and the Magna.

Production plans

Although Ford is only producing the Falcon range in Australia, annual output at its Broadmeadows plant is around 100 000 units. Consequently, it is reasonable to assume that Ford is currently able to exploit good scale economies. It also seems probable, however, that the opportunities for reduced costs would increase with a greater productive capacity (in excess of 100 000 units annually).

Overall future strategy

Currently Ford's strategy involves:

- increasing domestic sales with the new Falcon (Ford's success depends greatly upon fleet sales);
- becoming more export-oriented (vastly improved export opportunities for the Falcon are likely); and

[8] In contrast to Holden which faces difficulties exporting to Europe because of fuel economy legislation there, Ford is in an advantageous position to sell its new AU Falcon there. The Falcon could replace the Ford Scorpio, which was the largest Ford model sold in Europe until production stopped last year. Unlike Holden, Ford does not have a range of other large vehicles in Europe and can, therefore, meet current fuel reduction legislation in force there which directs producers to reduce fuel consumption across model ranges.

[9] Ford's Global Automotive Operations President, Jac Nasser, is quoted indirectly as saying that Ford Australia hoped to build between 20 000 and 60 000 export Falcons a year. With 60 000 sales, the exports would be worth more than $1 billion annually (Doak and Lynch 1997).

- boosting economies of scale either by increasing exports of the Falcon or producing a second model.

The current Falcon has an estimated product life cycle of only 5–6 years (rather than the usual 8–10 year life). In 2003-4, a completely new Falcon will be released by Ford Australia based on a rear-wheel-drive world platform. It seems likely that Ford Australia will play a significant role in the design and development of this platform for use in large Ford models worldwide. It is believed that Ford headquarters in Detroit recently 'signed-off' approval for development of the next model Falcon (McDonald 1998, p. 14).

Holden

Holden has established itself as a major force in the Australian PMV industry. Recent successes culminated in the 1996 announcement of a five-year, $1.4 billion investment program. Almost half of this was for retooling for the launch of the VT Commodore and for the introduction of a second assembly line at the company's plant. In August 1998, Holden commenced local assembly of the previously-imported medium sized Vectra on this second line.

Export plans

Holden's current exports are small relative to those of Mitsubishi and Toyota but Holden is becoming increasingly export-focused. The VT Commodore and WH Statesman/Caprice have been designed for both left and right-hand drive, so exports to the Middle East, Latin America and elsewhere have commenced (Lynch 1997a, Kable 1998 and Gover 1998). These vehicles are re-badged as Chevrolets, replacing the popular Chevrolet Caprice in the Middle East.

Most Vectras will be produced for export markets in Asia, although some may be sold in South Africa. The vehicle is designed to meet needs of consumers in Asian markets, with Holden realising the potential for selling small and mid-sized vehicles in Singapore, Hong Kong, Malaysia and Thailand.

Along with the Vectra, Holden regards the Commodore as the key to cracking Asia-Pacific export markets and the main link between the Australian manufacturer, its parent company and GM's Global Network. GM will reintroduce large rear-wheel drive vehicles in the United States, but will do this with only one rear-drive global 'platform' for cost efficiency. With its experience in developing the Commodore, it was expected that Holden might provide the world car design for use by various GM brands. However, this plan was recently rejected by GM headquarters in Detroit,

where sources say the Commodore is longer and wider than wanted for some markets. A GM executive stated 'this platform is more likely to come out of North America than anywhere else' (McCormick 1997, p. 16).

In Europe, GM is already marketing three large cars (the Opel Omega, the Saab 9-5 and the Cadillac Seville). This suggests that the export potential of the Commodore there is limited. In addition, European Union fuel economy legislation demanding a gradual reduction in average fuel consumption across a car producer's range of vehicles prejudices the introduction of another large GM vehicle there.

Overall, GM International has given Holden access to markets in Asia, the Middle East and Latin America with an estimated export potential of 50 000 vehicles per year.

Domestic plans

Holden has a two-pronged attack when it comes to Australian consumers. First, it is relying upon local Commodore sales to lift its domestic market share — recent figures show the VT Commodore has been Australia's top-selling car since its release in September 1997. However, Holden is concerned with threats by Toyota, Mitsubishi and Ford to its share of the fleet market. This currently accounts for 75 per cent of Commodore sales. Therefore, Holden will introduce the Vectra with significant domestic sales in mind.

Production plans

With Holden now producing the Vectra, overall capacity at its South Australian plant could climb to 160 000 units per year. Consequently, opportunities for exploiting economies of scale are sound.

Overall future strategy

Having primarily pursued a domestic market niche with the Commodore, Holden is now trying to broaden its appeal. Its plans entail:

- increasing domestic sales and boosting local market share with the Commodore (aimed primarily at the fleet market) and the Vectra; and
- becoming more export-oriented (increasing Commodore exports to markets such as the Middle East and Latin America and exporting 20 000 to 30 000 Vectras annually).

The viability of Holden's local operations seems assured given the design and manufacturing standards in place at Holden's plant in Elizabeth, South Australia.[10] Holden is now the regional design and engineering centre for every GM Opel vehicle selling in fifteen Asia-Pacific markets, including Japan, China, Thailand, India, Indonesia and Malaysia (RACV 1997, p. 7).

Mitsubishi

Of the four Australian PMV producers, Mitsubishi is finding it most difficult to compete, although there are some early signs of improved performance. A record post-tax loss of $66 million in 1996 was followed by a return to net profitability of $106 million in 1997 (table 11.7) and a small profit ($3.6 million) in 1998 . Mitsubishi has not announced plans for a major Australian expansion. Key decisions on whether to develop the new Magna locally need to be made soon.

Export plans

With the lowest Australian market share of all four local producers, Mitsubishi relies heavily on exports. Mitsubishi Australia has some market advantage in that it is the world source of the Magna (repackaged internationally as the Diamante).

At present, the Magna is exported mainly to the United States, with early export targets of 25 000 vehicles projected to generate around $750 million in annual revenues. However, Mitsubishi Australia is experiencing problems in the United States related to dealership and engineering compliance issues. It has been forced to work on incentive campaigns to bolster North American sales. Exports from Adelaide in 1998 fell 9549 units from 17 500 in 1997 (Lynch 1998a). Mitsubishi's export plans for the United States are supported by $30 million worth of Diamante exports to Taiwan, the first Australian company to break into this market.

---

[10] Holden Chief Jim Wiemels' comments on Holden's assembly plant in Elizabeth, South Australia and Holden's overall strategy are interesting:

'The plant has become so efficient, it is being used as a training base for executives who will be operating the new GM factory in Thailand … Make no mistake, Elizabeth is a class act, a world-leading class act' (Fulcrum-Analysis 1997a, p. 18).

'Holden is shifting from its domestic market focus to a global perspective … GM's determination to succeed in the Asia-Pacific depends on the contribution that Holden can make. Without Holden's resources, GM's timetable for growth in the region would be postponed for many years' (Fulcrum-Analysis 1997a, p. 18).

GM International Vice-President Peter Hannenberger described Holden as 'a value-added kind of company with an unbelievably capable engineering department that is very creative and can do a lot with a relatively little amount of money' (Kable 1998).

Mitsubishi Motors Australia has stated that if it could achieve exports of 25 000 units per year, it would reap substantial scale economies with fixed costs per car falling 30 per cent (Colebatch 1996, p. 1).

Domestic plans

The ability of Mitsubishi to boost the output of its Adelaide plant relies heavily on domestic success. It is forecast that domestic sales of the Magna must exceed 45 000 vehicles in order to expand output at Mitsubishi's plant to 75 000 units per year.

Like Toyota, Mitsubishi is pinning most of its hopes for future local success on the fleet market. Since mid-1997, Mitsubishi has targeted the fleet market more tightly with the Magna, sharpening its focus on the six-cylinder model. Improved security systems, a speed-alert warning and in-built wiring for a hands-free phone were added earlier — all features aimed at fleet buyers (Tuckey 1997b, p. 78). Mitsubishi expects government and business fleet purchases to account for 65 per cent of its domestic sales — roughly equal to Commodore and Falcon sales.

Production plans

With no major plans for expansion, Mitsubishi will continue to produce only the Magna (and Verada) for Australia and overseas markets.

Overall future strategy

Mitsubishi's future investment plans are modest compared with the three other local producers. These plans include:

- lifting domestic sales (particularly to fleet markets) of the Magna and exporting 25 000 units (primarily to the United States — 18 000 units); and
- increasing total production of the Magna/Diamante to 75 000 units per year to achieve sought-after scale economies.

Mitsubishi seems more intent on boosting economies of scale than diversifying its product range. A heavy reliance on one model for essentially two markets (the United States and Australia) is risky for Mitsubishi's long-term viability.

Limited future investment plans and weak recent export performance to the United States by Mitsubishi Australia give the impression of a company with an uncertain Australian future. Despite this lack of investment commitment and the current export problems, the signs look cautiously optimistic for Mitsubishi. Plans to increase plant capacity and exploit valuable scale economies have been

complemented by improving domestic vehicle sales and the possibility of ongoing improvements in profitability — 1997 after-tax profits were $106 million.[11]

Toyota

Despite recording losses on Australian operations for more than seven years (table 11.7), Toyota has pledged its commitment to local manufacturing with investment plans worth $1 billion over the next seven years. The decision further integrates Toyota's Australian operations into the company's global manufacturing plans.[12]

Toyota's Australian operations are regarded highly by its Japanese parent. Until mid-1999 at its Altona plant, four base models of the Corolla and Camry were built on a single production line, an achievement unique in Toyota's worldwide operations (TMCA 1996, p. 1).

Export plans

Toyota is a prime example of a firm whose strategy includes measures to boost economies of scale, while simultaneously addressing different domestic and international market tastes. By the year 2000, Toyota plans to double exports of its popular Camry to 28 000 vehicles. This will involve the production of an additional 20 000 units, which will lift total production to 90 000 units annually — slightly above the 'break-even' level in a plant capable of producing more than 100 000 units per year (Tuckey 1997b, p. 77).

Domestic plans

Although Toyota is clearly developing an export-oriented strategy, it remains focused on meeting the Australian market. In particular, Toyota has set its sights firmly on the lucrative fleet market for vehicle sales. In August 1997, Toyota launched the new Camry, with heavy selling emphasis on the four-cylinder version as a fleet car.

---

[11] Mitsubishi's general manager of merchandising, Charles Iles, stated in 1997 'We will make a profit this year. Our export volumes have been good and we have gained 2 per cent market share in Australia this year, with the Magna doing well' (Doak and Lynch 1997).

[12] The Hon. John Moore (1997, p. 1) states that 'The decision confirms that Australia has become a key production base for the Asian and Mid-East markets … and it reflects a recognition by Toyota that its Australian operations make a major contribution to the company's global network'.

As part of its investment plans, Toyota also announced that in 2000 it would begin production of the six-cylinder Avalon, a large family vehicle that Toyota believes will challenge the domestic market supremacy of Ford and Holden.

Production plans

Corolla production ceased in mid-1999. The plant is now preparing for the production of the Avalon scheduled to commence in May 2000.

Overall future strategy

Toyota's future in Australia looks positive. With a massive commitment to investment over the next seven years, Toyota aims to:

- double exports of the Camry by the year 2000 (Toyota currently produces both left- and right-hand drive vehicles at its Altona plant and is targeting overseas markets that have broadly similar tastes in medium-to-large vehicles, such as the Middle East, South East Asia and Oceania);
- increase domestic sales of the Camry (Toyota aims to do this by producing four and six cylinder Camrys, with the four-cylinder version aimed at fleet buyers.);
- begin production of the six-cylinder Avalon (initially, the Avalon will challenge the Falcon and the Commodore and will assist Toyota in regaining market leadership in Australia, but since it is currently only built in the United States, Australian production over time should involve exports to third country markets); and
- increase total output to 140 000 vehicles per year. This rise in total production through both increased exports and higher domestic sales will boost economies of scale at Toyota's plant.

11.6 General industry prospects

In the light of the above firm strategies, the industry outlook appears positive though complex, given different degrees of production specialisation and export market penetration among PMV producers.

The ability to exploit economies of scale is clearly critical for local producers. At present, only Ford and Holden have the capacity to exploit good scale economies. However, three of the four manufacturers have future strategies allowing them to increase volumes — the exception is Mitsubishi.

Holden and Toyota will have a two model line-up — in addition to the large six-cylinder range, each will have a medium-sized competitor. In contrast, both Ford and Mitsubishi will probably continue to rely on one model for both domestic and export sales.

The dependence on fleet sales by all four producers is worth highlighting. In 1996, each local model relied upon government and private fleet purchases for more than 50 per cent of sales (table 11.8). The upper-medium segment relied on fleet purchasers for almost 75 per cent of sales in 1997, with one third of these being government purchases (IC 1997a, p. 34). With the development of the 6-cylinder versions of the Camry and Magna, the dependence of local producers on this upper-medium segment — and hence fleet purchases — has increased. However, as more governments and agencies outsource fleet management duties, previous methods for selecting fleet vehicles will come under the spotlight. The choice of fleet vehicles is based on fuel economy, running costs, resale value and employee feedback (Tuckey 1997a, p. 10). Recent figures indicate that the Korean-built Hyundai Excel (Australia's top-selling 4-cylinder car) is by far the cheapest vehicle in terms of overall costs (table 11.9). It is likely that more attention will be paid to cars with low purchase prices and running costs, such as the Holden Astra and the Korean Hyundai Excel, rather than the 'big sixes' (Tuckey 1996, p. 82).

11.7 Government policy

Although the Government has opted for a freeze on tariffs at 15 per cent until 2004, it will legislate for a tariff reduction to 10 per cent in 2005.

The PMV industry might use high investment and low-profitability claims to prevent further tariff reductions below 10 per cent, since this was a successful tactic in forcing the Government's decision to freeze tariffs at 15 per cent for 5 years. However, caution is increasingly being taken with respect to both investment and PMV profitability claims. There are several reasons for such caution. First, recorded profits are net of payments by the companies to parent firms (and others) for knowledge and technical expertise (IC 1997a, p. 61). Second, payments for royalties, licence fees and physical products from Australian subsidiaries to their foreign parents raise the prospect that transfer pricing is distorting observed profitability. This is plausible because recent decisions to substantially expand capacity have been made in the light of claimed low profits and planned tariff reductions. The Australian Tax Office is known to have launched a major investigation into possible transfer pricing by PMV producers. Successful prosecutions of foreign PMV producers for transfer pricing have already occurred in Japan and the United States. On investment, Conlon (1997) finds discrepancies

between anticipated investment (investment plans announced by PMV producers) and actual investment (amounts committed). Table 11.10 shows that between 1986 and 1996, anticipated investments continually outweighed actual investments — for six of the seven periods, the discrepancies ranged from 13 to 35 per cent. These issues deserve further investigation, but it seems likely that future arguments for continued protection on the grounds of low profitability or investment will fall on increasingly deaf ears.

Apart from deciding future tariff levels, the Government must also make a decision with respect to its fleet purchases. Government has favoured locally-produced vehicles when it comes to choosing its fleet cars — there has been an unwritten 'buy Australian' policy and each locally-produced manufacturer relies heavily on Government fleet purchases. It is unclear whether this will continue, as the 'big-sixes' find it increasingly difficult to compete with cheaper imports. However, it would seem unlikely that Governments would make a complete about-face in coming years when choosing fleets. This is particularly true for the State Governments of Victoria and South Australia, where the PMV industry is heavily concentrated.

Finally, it should be noted that the Automotive Competitiveness and Investment Scheme (ACIS) will take effect on 1 January 2001 for a period of five years. The ACIS will provide duty free imports of PMVs or componentry equivalent to 25 per cent of the value of a company's total Australian production. The value of assistance will be capped at five per cent of the previous year's sales.

The ACIS will replace the previous Export Facilitation Scheme (EFS), which provided duty free imports of PMVs and componentry equivalent to the value of exports. The EFS was expected to be the subject of a challenge under the rules of the World Trade Organisation. Whereas the EFS provided, in essence, an export bounty equivalent in value to the tariff rate, the ACIS provides a subsidy based on the value of total production. The cap on the level of assistance at five per cent of previous year's sales under ACIS will certainly *bind*, so the scheme will have no net expansionary effect on either exports or the level of domestic output — it is a pure handout that boosts the profits of local PMV manufacturers. The only real effect it could conceivably have is to offset losses that would otherwise encourage local assemblers (and in particular the South Australian-based Mitsubishi) from exiting the market.

11.8 Conclusions

Future prospects for the Australian car industry look sound. Increasing profits, capacity and investment provide the basis for what can be a successful expansion period for the industry. A proven expertise in the design and production of large cars has been demonstrated. However, local producers must develop appropriate domestic and international strategies if they are to survive in an increasingly competitive, global environment.

In the light of recent and possible future developments for the international automotive industry, firms should (i) increase plant volumes to benefit from scale economies; (ii) develop export markets for their products; and (iii) consider producing more than one model to take advantage of economies of scope and greater production flexibility with respect to uncertain tastes.

Given the small domestic market, continual reliance upon the upper-medium vehicle segment is risky if producers maintain a domestically-oriented approach. Thus, exports are the key to future viability of local assembly, with future export success primarily dependent on successful market-niche targeting. The importance of developing feasible export markets is clear. Despite the growing problem of global excess capacity, the Australian industry is anticipating overseas sales of 150 000 vehicles by 2005, up from the 1996 volume of 40 000 cars (Mellor and McDonald 1997, p. 2).

The case for exports as a means of boosting scale economies is strong. The risk of sacrificing a domestic design niche is reduced if export markets have similar tastes to domestic markets, thereby allowing local firms to pursue overseas market niches without the need for major product redesigns. The targeting of appropriate export markets is a key influence on the competitiveness of the local industry. As Australian firms generally specialise in larger six-cylinder vehicles, identifying appropriate overseas demands is essential.

Almost surprisingly, it is predicted that most of the growth in Australian PMV exports to 2005 will come from sales to the mature, low-growth markets of Japan, North America and Europe, markets for which consumer demands are broadly similar to those in Australia (IC 1997a, p. 2). Asia, the Middle East and South Africa are also emerging as significant potential markets. With transport costs being relatively insignificant, it makes little sense to target only near neighbours at the expense of more distant markets.

Among the four local car makers, Mitsubishi's position seems in the greatest danger. However, the withdrawal of one vehicle assembler would not signal the industry's demise. Mitsubishi is the smallest local producer with production of

around 45 000 units annually. So, if Mitsubishi exited, the impact on remaining Australian producers or component producers would be negligible or possibly positive, given that demand for the Magna would then flow to other locally-produced vehicles.

The Australian PMV industry has the potential to develop into a highly sophisticated, technologically advanced and viable manufacturing industry. Such potential will be realised by those firms that are able to cut costs and skilfully target markets.

Appendix 11A Labour costs and productivity

It is often claimed that Australian PMVs cannot compete with imports because of high local labour costs. IC (1996) data suggest this argument is false. Table 11.1 presents labour cost data for Australia, Japan, South Korea and the United States. Average labour costs in Australia are clearly quite low and have converged toward Korean levels. Australia is a low labour cost producer relative to the United States and Japan with its cost advantages increasing over the period 1988 to 1993. Table 11.2 shows relative labour productivity in these countries.

Table 11.1 **Automotive industry average labour costs, 1988 and 1993**
(A$ per hour)

	1988	*1993*
Australia	13	17.5
USA	24	36
Japan	23.5	39
Korea	5	13

Source: IC (1996).

Table 11.2 **Labour productivity, 1988 and 1993**
(labour hours per standard vehicle)

Location of producer	*1988*	*1993*
Australia	40.9	35.5
Japanese (in Japan)	16.8	16.5
Japanese (in North America)	21.8	17.4
United States of America	25.1	21.9
Europe	36.2	25.3
Newly-industrialised countries[a]	38.8	29.7

[a] Includes Korea, Taiwan and Malaysia.

Source: IC(1996).

Given these wage and productivity rates, the implied total labour costs in 1988 were $530 per standard vehicle in Australia, $195 in Korea, $610 in the United States and $400 in Japan (table 11.3). By 1993, total labour costs were $620 per standard vehicle in Australia, $780 in the United States and $635 in Japan. Korean costs doubled to $390 per vehicle.

Table 11.3 **Total labour costs to produce a standard vehicle, 1988 and 1993**
(A$ per vehicle)

	1988	*1993*
Australia	530	620
United States of America	610	780
Japan	400	635
Korea	195	390

Source: IC (1996).

The difference in total labour costs between Australia and Korea in 1993 was $230 per vehicle, which is of the order of just 1 per cent of a new vehicle's price. Furthermore, relative capital costs are higher in Korea because of higher interest rates there.

Although both United States and Japanese plants operate at higher volume, thus generating substantial scale economies and labour productivity, the high hourly labour costs in both these countries mean that the total labour costs of producing a standard vehicle are lower in Australia than in either Japan or the United States. With recent major redundancies and the shift towards lean production, skill levels of Australian PMV workers are much higher than a decade ago. These skills in design and engineering exceed those of workers in LDC plants. Indicative of this was GM's recent decision to involve Holden Australia in training workers (primarily managers) for future positions in GM's new assembly plant in Thailand.

Although these arguments disregard labour costs associated with component manufacture, such components are traded internationally, which prevents Australia from necessarily experiencing relative cost disadvantages in using such components.

Appendix 11B Evidence of economies of scale

Vehicle assembly involves substantial fixed costs of capital. These arise because of development, tooling and general plant and equipment costs. At the plant level, long-run average cost declines as output expands, enabling scale economies (IC 1997a, p. 92).

Fuss and Waverman (1992) verify that PMV production in the United States, Japan, Germany and Canada enjoys increasing returns. The estimated elasticity of scale is between 1.10 and 1.17 so an output doubling reduces costs by 10 to 17 per cent. Using a translog cost function for the Australian industry, Truett and Truett (1996) also conclude economies of scale exist. Their estimated elasticity of cost with respect to output is 0.76, implying an elasticity of scale of 0.24 and returns to scale of 1.32.

While there have been various estimates of minimum plant volumes required to achieve optimum scale economies in assembly (anywhere from 55 000 to 200 000 units per year), it is generally argued that minimum production runs of 100 000 units are necessary.[13] Industry Commission estimates suggest unit costs decrease by 5 per cent as volumes increase from 75 000 to 100 000 units per plant (IC 1997a, p. 94). These cost reductions are small given the output increase.[14] The cost-importance of increasing volumes may be relatively low as most Australian producers currently generate close to 100 000 units annually.

Their larger production runs suggest both Ford and Holden are benefiting more from scale economies than Mitsubishi or Toyota. This view is supported by profit comparisons (tables 11.4–11.10). However, this might be an incorrect inference. Ford invests most heavily in local design and development and so requires a larger production run to break even.

[13] IC estimates suggest that scale economies were realised between 55 000 and 110 000 units annually in 1997 (IC 1997a, p. 92), in contrast with 200 000 units in 1990 (IC 1990, p. 31).

[14] A similar conclusion was reached by Mitsubishi: 'increasing production from 50 000 units per annum to 100 000 units per annum … would reduce the total assembly cost … by 7.5 per cent — just 1 per cent of the total cost of the car' (IC 1990, p. 200). Note assembly costs are only 20 per cent of total vehicle costs — costs of materials and components account for the remainder.

Holden also feels that the incremental savings on expansion are limited: 'while there are further potential cost savings to be made by increasing output from 150 000 to say 200 000 per year these are relatively modest, especially when the investment required is taken into account. Maintaining a consistently high level of capacity utilisation is far more important in car manufacture than increasing capacity — at least when volumes in excess of 150 000 per annum are concerned' (IC 1997a, p. 93).

Appendix 11C International plant capacity comparison

Table 11.4 International plant capacity comparison, 1996

Australian plant			*Overseas plant*		
Location	*Vehicles produced*	*Capacity*	*Location*	*Vehicles produced*	*Capacity*
Altona, Vic.	Toyota Corolla/Camry	100 000	Cambridge, Canada	Toyota Corolla	100 000 (rising to 250 000)
Broadmeadows, Vic.	Ford Falcon	120 000	Georgetown, USA	Toyota Camry/Avalon	400 000
			Chicago, USA	Ford Taurus/ Mercury Sable	300 000

Source: IC (1997b).

Table 11.5 Tariff rates on PMVs, selected countries, 1997

(per cent)

Country	*Tariff*
Canada	8
United States of America[a]	2.5
Japan	0
Indonesia[b]	125+75
Malaysia[c]	140 to 200
China	100

[a] The United States applies a separate 25 per cent tariff on imported light trucks. [b] Indonesia applies a 125 per cent tariff on imported PMVs plus an extra 75 per cent surcharge. [c] Malaysia's tariff on imported PMVs recently increased to 300 per cent in light of growing currency problems.

Source: IC (1997a).

Table 11.6 Shares of Australian PMV market, selected PMV segments

(per cent)

Year	*Small/light*	*Medium*	*Upper-medium*
1993	32.6	17.6	38.0
1994	31.6	16.3	39.6
1995	35.0	13.2	39.8
1996	37.3	10.6	40.4

Source: DIST (1995).

Table 11.7 **Profitability of Australian PMV producers, 1989 to 1996**

($ million)

Year	*Ford*		*GM Holden*		*Mitsubishi*		*Toyota*	
	Pre-tax	Post-tax	Pre-tax	Post-tax	Pre-tax	Post-tax	Pre-tax	Post-tax
1989	221	124	85[a]	66[a]	70	41	20	25
1990	(134)	(83)	222	151	26	13	(22)	(23)
1991	(165)	(114)	83	53	(35)	(24)	(70)	(19)
1992	(37)	(38)	45	37	2	1	(26)	(32)
1993	76	(45)	98	64	(18)	(48)	(19)	(32)
1994	180	146	242	163	17	11	13	(27)
1995	(212)	202	360	260	(44)	(44)	(136)	(145)
1996	na	217	na	243	(66)	(66)	na	(6)
1997	na	179	na	172	na	106	na	60

a Figure for half year to June. Holden figures are for financial years, except for 1995,1996 which are calendar years. **b** 18 month figures to December. Other Toyota data are for financial years.

Source: FCAI (1997), FMCA (1991) and Lynch (1998b).

Table 11.8 **Local PMV sales, by purchaser, 1994 to 1996**

(per cent)

	1994			*1995*			*1996*		
	Prvte	*Govt fleet*	*Non-govt fleet*	*Prvte*	*Govt fleet*	*Non-govt fleet*	*Prvte*	*Govt fleet*	*Non-govt fleet*
Falcon	30.22	22.07	47.71	23.02	28.68	48.30	23.21	28.59	48.20
Commodore/Lexcen	27.95	23.66	48.39	27.00	27.35	45.65	21.89	32.82	45.29
Camry/Apollo	39.70	18.68	41.61	42.48	20.12	37.40	44.68	15.36	39.96
Corolla/Nova	44.98	19.57	35.45	39.84	26.56	33.60	48.96	23.38	27.65
Magna/Verada	32.51	22.47	45.02	28.06	23.44	48.50	30.87	14.30	54.83

Prvte: private; Govt: Government.

Source: IC (1997b) and DIST (1995).

Table 11.9 **Car running costs over three years, 1995–97[a]**

	Ford Falcon GLI auto, A/C	*Holden Commodore Executive auto, A/C*	*Mitsubishi Magna Executive V6 auto, A/C*	*Toyota Camry CSI auto, A/C*	*Toyota Corolla CSI auto, A/C*	*Hyundai Excel Sprint man, A/C*	*Ford Festiva GLI man, A/C*
List price	$32,442	$31,740	$31,635	$29,925	$25,665	$15,890	$18,540
Cost per km (cents)							
Operating	38.8	37.8	37.96	36.07	36.57	20.68	24.01
Standing	16	14.7	13.8	14.2	12.4	12	11.5
Average total running costs							
Cents/km	54.8	52.5	51.76	50.27	48.97	32.68	35.51
$/week	316	303	299	290	254	189	205

a Figures are for business vehicle running costs based on travelling 30 000 km a year for three years. Calculations also factor in the buying price (with air-conditioning), depreciation, interest on finance, fringe benefits tax, registration, insurance, fuel, tyres and service.

Source: RACV cited in Tuckey (1997b).

Table 11.10 **Anticipated and actual investment, 1986–89 to 1993–96**
(1993$ billion)

Period	*Anticipated Investment*	*Actual Investment*
1986–89	2.96	2.30
1988–91	2.01	1.75
1989–92	2.36	1.75
1990–93	2.93	1.89
1991–94	2.53	1.76
1992–95	2.30	1.82
1993–96	2.09	2.08

Source: Conlon (1997).

References

Colebatch, T. 1996, '$525m gamble on Magna', *The Age*, 4 April.

Conlon, R.M. 1997, The Automotive industry in Australia: policies and prospects, School of Economics, University of New South Wales, unpublished.

DIST (Department of Industry, Science and Tourism) (various years), *State of the Automotive Industry*, AGPS, Canberra.

Doak, A. and Lynch, M. 1997, 'Ford plans export push with new Falcon', *The Age*, 28 November.

FCAI (Federal Chamber of Automotive Industries) 1997, *Automotive Industry Financial Performance*, Federal Chamber of Automotive Industries, Canberra.

FMCA (Ford Motor Company of Australia) 1991, *Ford: The Australian Story*, Ford Motor Company of Australia, Campbellfield, Victoria.

Fulcrum-Analysis 1997a, 'Going global at GMH', *The Age*, 26 August.

—— 1997b, 'Geelong on top of the world, says Ford boss', *The Age*, 28 November.

Fuss, M.A. and Waverman, L. 1992, *Costs and Productivity in Automobile Production — The Challenge of Japanese Efficiency*, Cambridge University Press, Cambridge, United Kingdom.

Gover, P. 1998, 'Holden export hint', *Herald Sun*, 3 April.

IC (Industry Commission) 1990, *The Automotive Industry*, Report no. 5, AGPS, Canberra, December.

—— 1996, *Automotive Case Study: Micro Reform — Survey of Impacts on Firms*, Report 96/17, AGPS, Canberra.

—— 1997a, *The Automotive Industry*, *Volume I: Report*, Report no. 58, AGPS, Canberra, May.

—— 1997b, *The Automotive Industry*, *Volume II: Appendices*, Report no. 58, AGPS, Canberra, May.

Kable, M. 1998 'GM holds out promise of wide international role for Holden', *The Australian*, 30 March.

Keller, M. 1996, 'Excess capacity — A past, present and future constraint', *Power Report*, July, pp. 4–16.

Linder, S.B. 1961, *An Essay on Trade and Transformation*, Almqvist and Wiksell, Stockholm.

Lynch, M. 1997a, 'Why the 'rustbucket' state now has a shiny new set of wheels', *The Age*, 23 August.

—— 1997b, 'Driving around the global village', *The Age*, 23 August.

—— 1998a, 'Mitsubishi reverses its fortunes', *The Age*, 2 April.

—— 1998b, 'Toyota battles back into the black?', *The Age*, 14 May.

Lynch, M. and Saunders, D. 1998, 'Ford stays on top as market surges', *The Age*, 9 January.

McCormick, J. 1997, 'Commodore 'too big' for GM's world car', *The Age*, 16 October.

McDonald, N. 1998, 'Falcon grilled to be the frontrunner', *The Australian*, 13 August.

Markusen, J.R., Melvin, J.R., Kaempfer, W.H. and Maskus, K.E. 1995, *International Trade: Theory and Evidence*, McGraw-Hill, New York.

Mellor, J. and McDonald, N. 1997, 'Makers set export goal to top $6bn', *The Australian*, 14 June.

Moore, The Hon. J., MP 1997, *Media Release: Toyota to Increase Exports*, Canberra, 5 August.

RACV (Royal Automobile Club of Victoria) 1997, 'Cars: The good news — and what it means', *Royalauto*, September.

TMCA (Toyota Motor Corporation Australia) 1996, *Toyota's Altona Plant a Showpiece*, Toyota Motor Corporation Australia, Port Melbourne.

Truett, L.J. and Truett, D.B. 1996, 'A cost function analysis of the Australian transportation equipment industry', *Australian Economic Review*, vol. 116, no. 4, pp. 367–84.

Tuckey, B. 1996, 'Showrooms brace for impact of buy-backs', *Business Review Weekly*, 26 August.

—— 1997a, 'The battle has started, but who is going to win?', *The Age*, 10 April.

—— 1997b, 'An auto epic in four acts', *Business Review Weekly*, 2 June.

VFACTS 1995, *Vehicle Retail Sales*, Federal Chamber of Automotive Industries, Melbourne.

12 How does antitrust affect business behaviour? Lessons from Australian corporate experience

Grant Fleming

The impact on business behaviour of new laws governing restrictive practices is often difficult to gauge unless behaviour is revealed in subsequent legal proceedings. One solution to this difficulty is to investigate past experience of regulatory change and draw upon this data to inform current policy. This paper provides before-and-after case studies of Australian business responses to antitrust legislation using a sample of six firms operating during the early years of the twentieth century. From these studies, this paper describes the legal game established by the new legislation, defines avoidance and defence strategies and outlines the strategic behaviour undertaken by firms.

12.1 Introduction

The enactment, implementation and enforcement of early Australian competition policy provides industrial economists with a natural experiment from which to examine the ways in which firms alter behaviour under new antitrust laws. While there has been much interest in American regulatory change and the impact of these changes on business behaviour, less attention has been given to Australian experience.[1] This is more so the earlier history of competition policy developments that began with Federation. Study of this earlier time period yields many benefits for researchers who are presented with a regime change from an 'unregulated' to a 'regulated' jurisdiction. Using the enactment of new Australian antitrust legislation in 1906, this paper presents results of a 'before-and-after' study of business responses to regulatory change.

[1] See, for example, Joskow and Rose (1989) for a summary of work on economic regulation in the United States.

The paper has five sections. Section 12.2 reviews prior research on early Australian competition policy focusing on the 1906 Act. The sample used for the explanatory case studies is described in section 12.3, and results and interpretations are presented in section 12.4. Some conclusions follow.

12.2 Prior research

Research on the history of Australian antitrust laws and cartel behaviour can be organised into two broad subject areas. Legal and social historians have focused on the Federation debates, enactment and operation of the *Australian Industries Preservation Act 1906* and on the first prosecution collectively known as the Coal Vend cases. Writers such as Stalley (1958), Castles (1959), Hunter (1961), Hopkins (1978a, b) and Ransom (1981) describe the sequence of legislative changes that comprised Australian competition laws prior to the first Trade Practices Act and the economic arguments considered in the first antitrust cases involving the New South Wales coal cartel and shipping companies between 1910 and 1913. Less detailed examinations of Australian antitrust experience include a number of recent articles examining the twenty years of operation of the Trade Practices Act 1974, which provide introductory comments on early antitrust laws (eg Round 1994).[2] Few works attempt to examine the *impact* of the antitrust laws on business structure and subsequent behaviour, although Stalley (1958, p. 288) notes that 'indirectly it [the 1906 Act] may have induced more caution on the part of some businessmen and a tendency to achieve monopoly by other and safer means than by formal agreement and open collusion'. The added value of this paper is to investigate the means by which 'caution' and 'safer means' manifested themselves.

The second theme in the literature relates to the operation of trusts and combines during the 'unregulated' period prior to the 1906 Act and subsequent to the Act. Wilkinson's (1914) classic account of cartel structures provides us with a comprehensive catalogue of cartels, although there is little analysis of these agreements in an economic framework. Recent business history studies also provide some discussion of individual cartels, but do not direct attention to the impact of the 1906 Act (Terwiel 1996; Fleming and Terwiel 1997; Merrett and Ville 1998; Ville 1998). Finally, several studies in the 1960s examined the structure of the Australian economy that resulted from a lack of antitrust enforcement after 1913 (Barwick 1963; Karmel and Brunt 1963; Sheridan 1968). Most of these studies reinforced the findings of Karmel and Brunt (1963, pp. 77–92) who showed that the market concentration of Australian industries was high relative to other developed

[2] See also texts on Australian restrictive trade practices legislation such as Donald and Heydon (1978), Norman (1976) and Pengilley (1979).

countries, such as the United States and England, and that high market concentration was *partly* due to a lack of antitrust enforcement. Few of these studies present a detailed analysis of cartel behaviour under conditions of regulatory switching, although several provide excellent industry studies (eg Hunter 1963).

In summary, apart from analysis of the coal vend cases and the broad structural characteristics of Australian businesses in the pre-Trade Practices Act era, no work has attempt to (a) examine the structure of cartel agreements and collusive practices prior to the 1960s, or (b) assess the impact of new antitrust legislation on Australian business behaviour. The natural experiment that is presented in the enactment of the 1906 Act provides us with an opportunity to consider both these issues. The next section describes the sample used in this 'before and after' study.

12.3 Sample

The explanatory case studies undertaken in this paper derive from a sample of 23 industries examined by the Federal Attorney-General's Department between 1906 and 1911. The industries investigated by the Department covered State and interstate registered businesses and corporations, as well as several trade associations. The sample was reduced to six industries after allowing for the fact that the Federal legislation only applied to interstate collusive agreements and restrictive practices.

Summary statistics (where available) are provided for the six industries in table 12.1 below. In the tobacco, shipping, confectionery and dried fruit industries, cartel agreements covered all firms in the industry. In the confectionery and dried fruit cases, manufacturing trade associations comprised all manufacturers in the particular State, although the exact number of firms cannot be determined. Restrictive practices by the Colonial Oil Company were undertaken to restrict the entry and success of the British Imperial Oil Company. Two of the cases (tobacco and mineral oil) involved multinational corporations with head offices in United States and England. Pricing and/or price discounting agreement were the most popular form of restrictive practice. Output agreements were only found in two cases where output was relatively easy to monitor: coal, which had a long history of collusive agreements and a well-developed monitoring structure; and tobacco, where all firms held shares in a holding company which coordinated production and distribution. Overall, it is difficult make any generalisations about the length of each agreement. The coal and shipping industries had long histories of operating collusive agreements, but the joint coal-shipping cartel barely lasted four years

Table 12.1 **Summary statistics**

Industry	*Number of firms*	*Number of firms in collusive agreement*	*Types of restrictive practices*
mineral oil	2	1	D, F
tobacco	5	5	O, P, D, F
coal	–	40	O, P
shipping	7	7	D, P
confectionery	–	–	D, P
dried fruit	–	–	–

O=output restrictions, P=pricing agreements, D=discounting and F=full-line forcing.

between 1906 and 1910.[3] In the tobacco industry, the holding company facilitating trade dominated production and distribution until the late 1950s. By contrast, mineral oil production became a secondary product in the oil industry by the 1920s as rising consumer demand for petrol led mineral oil companies to switch production. A Royal Commission into the Petroleum industry in the late 1920s found that most mineral and petroleum companies had continued restrictive practices 'refined' in earlier years!

Sample bias, validity and reliability

Two features of the sample may lead to sample bias. First, while the sample covers all interstate cartels examined by the Department, it may be the case that other cartels escaped investigation. Thus, it is likely that the sample understates the number of cartels operating in Australia during this period. Second, a perverse survival bias may exist in the sample in that the sample does not capture any cartels that broke up due to the enactment of the new legislation. Thus, any results derived from these case studies will tend to underestimate the effect of antitrust legislation on business behaviour.

Validity and reliability issues relate here to the size of the sample, the validity of the evidence used in the case studies and the extent to which the business behaviour observed can be generalised. Given that the sample size is small, the data lends itself more to an event study using explanatory case studies rather than empirical analysis. It is hoped that further research can provide a more sound empirical analysis of cartel behaviour in Australia. In order to provide some test of the internal validity of the information gathered, it was possible to collect intra-firm and inter-firm correspondence from five out of six industries (thus providing a test for internal validity via triangulation). The use of multiple sources of evidence and the construction of a chain of evidence (establishing temporal sequentiality) is generally

[3] On the coal cartels of the late nineteenth century, see Fleming (1999). On the shipping cartels, see Broeze (1992).

regarded as an appropriate method in case study research to increase the internal validity of the study and avoid the occurrence of spurious association (Ryan, Scapens and Theobald 1992, pp. 113–29 and McGuire 1998). With regard to the external validity and reliability of results, it is recognised that tests for validity such as the replication of results are difficult to undertake. Nevertheless, the use of six industries ranging between basic raw materials (coal), intermediate goods (dried fruit) and final consumer goods (tobacco, confectionery, mineral oil) increases external validity and reliability.

12.4 Results and interpretation

The results of investigating the behaviour of our six industries before and after the enactment of new antitrust legislation is presented in terms of two common strategies: avoidance and defensive. Each of these strategies will be defined in more detail below, but first a brief description of the pivotal event, the enactment of the 1906 Act, will be provided.

Legislation

The *Australian Industries Preservation Act 1906* was modelled on the US *Sherman Act 1890*, although it contained several sections thought by the legislators to improve the US legislation. *Interstate* cartel behaviour was deemed illegal by sections 4(1)(a) and 7(1) of the Act, which stated that any person who enters a contract or engages in a combination that restrains (or monopolises) or *intends* to restrain (or monopolise) trade or commerce *to the detriment of the public* is deemed guilty of the offence.[4] These sections were based upon the Sherman Act, although both provided the courts with the task of determining intent of the prosecuted firms (Stalley 1958, pp. 263–4). In addition, the Act provided the courts with a rule of reason in that restrictive practices could be deemed legal if they were beneficial to the public. Section 4(1)(b) covered practices related to predatory behaviour by making illegal actions that were intended to destroy or injure by means of 'unfair competition' any Australian industries 'the preservation of which is advantageous to the Commonwealth'.

4 The new legislation initially covered firms operating intrastate if they were corporations. In Huddart Parker (8 CLR 330) the High Court found that the Commonwealth did not have jurisdiction to regulate intrastate business activities whatever the type of firm and, thus, from 1908 the legislation related solely to foreign and interstate trade and commerce (Stalley 1958, pp. 264–5).

Given the above offences, there were three defences open to defendants depending on which sections were used in prosecution. First, firms could argue that they did not intend to restrain or monopolise trade. Second, it could be argued that the cartel provided outcomes that were to the benefit of the public (using section 4(3)(a)). Related to this, the cartel could also argue that the restraint was 'reasonable' bearing in mind the price, quantity and quality, profit and wage outcomes facing consumers, producers and workers (section 4(3)(b)). Finally, it could be argued that the restrictive practices were 'fair' competition because the injured firm was 'reasonably efficient, effective and up-to-date' in its management, processes, plant and machinery (section 6(2)). The extent to which firms in the sample used these defenses is examined later in this section.

The new law established the following game for prosecution and defendants.[5] Let us use the example of a cartel facing prosecution under the Act for restraining trade via exclusive dealing and full-line forcing. In the first instance (D_1), the cartel decides to continue colluding and undertaking restrictive practices. If detection occurs, the next move is undertaken by the prosecution (P_1) which makes the decision whether or not to take proceedings against the cartel (the prosecution would most likely use sections 4(1)(a) and 7(1) in this case). The cartel then faces the choice (D_2) of arguing the legality of their actions using a set of statements outlining the defenses above, or admitting the offence. From this point, the two parties go to trial and receive an outcome from the courts. This game can be represented by a tree describing the sequence of events (see figure 12.1).

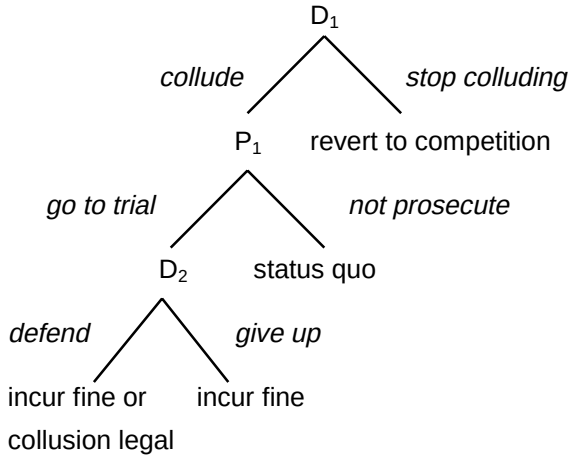
A second round of the game may arise if the initial prosecution was deemed unsatisfactory for either party (eg when a test case provides a precedent undesirable from the point of view of the regulator). Such a second round game took place in the coal vend cases.

Avoidance

The first decision in the game relates to the cartel's choice over whether to continue to collude and undertake an avoidance strategy or stop colluding. All firms in our sample continued to collude and some took avoidance strategies. Avoidance can be defined in the following way:

[5] Drawing upon Rasmusen (1994, pp. 98–106, 392–4) and Binmore (1992).

Figure 12.1 **Extensive form for the Australian antitrust Game**



Definition 1: An *avoidance strategy* is a course of action whereby a cartel C comprising members c_1, c_2, ...c_n continues to collude, but modifies the structure of the cartel agreement to lower the probability of detection.

A few points of definition are perhaps pertinent at this stage. Following Rasmusen (1992, p. 12), a 'strategy' is defined as a rule that tells the cartel which action to choose at any point of time given the information set of the cartel. The 'structure of the cartel agreement' is the terms and conditions of the cartel's agreement (which may include pricing, output, discounts and rebates, exclusive dealerships and penalties), as well as the governance structure of the cartel (such as regularity of meetings, recording of minutes, monitoring procedures and dispute resolution mechanisms).

Evidence indicates that avoidance strategies were undertaken by three out of the six cartels studied.

Avoidance Strategy 1: refrain from documenting the interstate cartel agreement. This strategy was undertaken by the coal-shipping, confectionery and mineral oil cartels. The first example is from coal-shipping cartel. The formal terms of the agreement regulating coal production were finalised in November 1906 and the agreement bound parties for seven years from 1 January 1907.[6] Difficulties pertaining to enforcement, product differentiation and price discrimination, and price setting were addressed. As in previous cartels, each colliery appointed a

6 An unsigned copy of the agreement has survived and can be found in Noel Butlin Archives Centre, Australian National University (hereafter NBAC/ANU), New South Wales Colliery Proprietors' Association Papers (NSW CPA), E207/29.1.

member to the Board which had responsibility for setting price levels and apportioning trade. Incentives to produce the allotted amount were provided, with funds flowing from over-producers to under-producers during each year. However, unlike other agreements, the 1907 agreement avoided establishing an explicit contract. The difficulties such a decision posed for the cartel were not lost on one of the key players, F. Livingstone-Learmonth (chairman of the Associated Northern Collieries), who stated that: '[w]e have certain unwritten rules which have verbally been agreed upon for our guidance, but beyond that we have nothing, and there is nothing to prevent any one member of our Association withdrawing therefrom whenever it suits him. Our common interests are our only bond'.[7] In addition to refraining from signing a formal agreement, the colliery and shipping companies were advised by lawyers in September 1906 to recall and cease recording minutes of association meetings.

The confectionery cartel maintained a State-based cartel agreement and, although sought to make explicit interstate restrictive practices, was warned against doing so by legal advisers. In addition, lawyers suggested a revision of the exclusive dealing clauses of the current agreements that covered Victorian middlemen to decrease the probability of a disgruntled middleman taking the association to court.[8] Meetings between Sydney and Adelaide manufacturers in November 1907 were also advised to restrict any agreements to 'understandings'.[9] In the mineral oil case, the Colonial Oil Company switched from a compulsory to a voluntary rebating system and from 1908 avoided using contracts between itself and retailers. Traditionally, the rebating system had been used to control the terms and conditions in the sale of mineral oil. The switch to voluntary rebating and the absence of contracts decreased the likelihood that regulators would identify the continuing restrictive practices. Certainly to the Crown Solicitor, such behaviour was evidence that the company had drawn upon experience under the Sherman Act to nullify antitrust proceedings in Australia: '[t]he company appears to have profited by experience gained by many prosecutions under the United States State and federal Antitrust and Anti Monopoly Laws, and are possibly refusing to enter into contracts binding them to give rebates'.[10]

---

7 Letter from Livingstone-Learmonth to Binnie, Secretary of the West Moreton Colliery Owners Association, 18 October 1906 in NBAC/ANU, AACo, 1/57/56 CC-BC C23.

8 Letter from lawyers Upton and Plante to F.L.W. Ashley (Chamber of Manufacturers), 13 August 1907 in Australian Archives (AA) A432/86 1929/2747.

9 See evidence to Crown Solicitor from M. Mendes, Wholesale Merchant in Confectionery 22 November 1907, and from Frank Taylor, Bennett and Taylor Wholesalers 22 November 1907 in AA A432/86 1929/2747. Similar sentiments were obtained by the Crown Solicitor a year later when interviewing F.J. Ransom of Morrow Ltd (Replies to Questions, 13 August 1908).

10 Opinion of Crown Solicitor on the Colonial Oil Company 27 July 1909 in AA A 432/86 1929/2749, p. 10.

Defence

The second type of strategy relates to the actions of the cartel once detected and involves mounting arguments to defend the cartel's behaviour. In the cases examined here, most cartels undertook counterfactual simulation studies to construct a set of defensive strategies to draw upon if the cartel was detected and faced prosecution. A defensive strategy is defined in the following way:

Definition 2: A *defensive strategy* is a course of action whereby a cartel constructs a legal argument outlined as statements s_1, s_2, ...s_k that collusion is permitted under the law.

The essence of this strategy is to construct statements based upon the statutorily provided defenses in the Act and use this defence when prosecuted by the regulator. The set of statements common to the cartels studied in this sample were (a) the cartel forms a beneficial monopoly (denoted S1); (b) the cartel has no interstate control (S2); and (c) the cartel did not intend to act to the detriment of the public (S3). Each strategy is examined in turn.

Defensive strategy statement S_1: the cartel forms a beneficial monopoly. The desire by legislators to permit beneficial monopolies and cartels to operate led to the inserting of the words 'to the detriment of the public' in the Act. The cartels' defence here was to argue that the restrictive practices benefited consumers by handing on gains obtained by economies of scale (as in the tobacco case) or that rebates paid to retailers for exclusive dealing were handed on to consumers (mineral oil). The British Tobacco Company Australia (BTCA) claimed that the holding company acted to the benefit of the public by permitting lower production and distribution costs.[11] Investigations into the Colonial Oil Company indicate that the Crown Solicitor weighted highly their defence of lower consumer prices — the policy of rebates may benefit the public as 'the greater the rebate allowed, the cheaper the public ought to get the goods and the greater the trade'.[12]

Defensive strategy statement S_2: the cartel has no interstate control. The constitutional limitations on the Commonwealth at that time meant that it could only prosecute cartels that operated at an interstate level. Various legal opinions provided to the dried fruit, confectionery, coal and shipping cartels indicate that cartel lawyers were well aware of the advantages to their clients of arguing that only *intrastate*

[11] The holding company BTCA was controlled by five major tobacco producers and distributors — the three largest domestic firms Kronheimers, W.D. and H.O. Wills, Dixons, and two Americans Geo.A. Cameron and American Tobacco Co. (Wilkinson 1914, pp. 38–40, 42–3, 47)

[12] Crown Solicitor to Attorney-General's Department, 20 November 1906 in Australian Archives (ACT) A432/86 1929/2749.

trade was controlled. In addition, in the tobacco industry, the holding company BTCA was constructed in such a way as to make the determination of control over the trade difficult to ascertain. The corporate veil structure involving three levels of holding companies was an ideal means to support a defence against antitrust litigation. Each firm *nominally* held no controlling interest in the interstate trust and yet each major firm held, via share-swaps and interlocking directorships, at least 10 per cent control of the interstate trade. The state-registered firms could not be prosecuted under the Federal legislation so long as they focused on intrastate trade. The BTCA could mount an argument that it did not control the interstate trade.

Defensive strategy statement S_3: the cartel did not intend to act to the detriment of the public. The inclusion of a *mens rea* (ie intent) requirement to be proved by the prosecution hampered early Attorney-General investigations. Lawyers for cartels exploited this requirement of proving intent. The Colonial Oil Company argued that its rebating system was intended to provide cheaper prices to the public. The confectionery combine argued that the intent of the State-based trade associations was to avoid the 'downfall of the industry' and subsequent loss of jobs. Both arguments were accepted by the Crown Solicitor during investigations. His opinion on the confectionery cartel, stating that 'a desire to keep the industry afloat, and to preserve many members of the parties to the arrangement, does not prove an intent to restrain trade or control the supply or price of commodities *to the detriment of the public*' (emphasis in original).[13]

Application: Avoidance and defensive strategies in the coal-shipping case

The celebrated coal vend cases provide an illustration of how the strategies described above were combined to successfully defend a cartel position.[14] The game took place over four years between 1910 and 1913 and can be summarised below:

Start: In 1906 the cartel entered a new agreement to collude secretly (using avoidance strategy 1). Given that both the regulator and the cartel were unsure about the probability of the outcome of the trial (given that this was a test case), moves 1 and 2 were made under the uncertainty inherent in the first judicial interpretation of a new piece of legislation.

[13] Crown Solicitor to Attorney-General's Department, 13 September 1909 p. 2 in A5522 M92. On the Colonial Oil Company see Crown Solicitor to Attorney-General's Department, 20 November 1906 in AA A432/86 1929/2749.

[14] It is of course rather ironic that by the time the cartel was found to have been legal under the Act, the cartel had disbanded and firms had reverted to competitive market strategies.

Move 1: The regulator investigated the coal-shipping cartel and collected evidence from disgruntled members. The Crown Solicitor advised that the regulator should go to trial.

Move 2: The cartel decides to defend the charge and utilises defensive strategies 1,2 and 3 in the case.

The trial took place in 1910 and was heard by Isaacs J. The judge found for the prosecution in the first instance, accepting the argument that the cartel had combined with intent to monopolise, to the detriment of the public, the interstate coal trade. The cartel was ordered to pay fines of five hundred pounds per defendant and an injunction prevented further collusive behaviour.

Move 3: The cartel decided to appeal the decision of Isaacs J. and was successful in arguing defensive strategy 3. The High Court and the Privy Council found that the cartel agreement was legal in that 'there was no evidence (at any rate no satisfactory evidence) of any sinister intention on the part of either colliery proprietors or shipping companies; and secondly, … there was no evidence … of injure to the public'.[15] While the detailed financial data from the cartel with which to quantify the legal costs and benefits are not available, it is most likely that the monetary penalties under the Act were insufficient to warrant an appeal. While Isaacs J. could have fined each defendant five hundred pounds per day for every day that the offence took place, he decided upon a lower fine of five hundred pounds per defendant. The appeal by the cartel was probably motivated by the cartel's desire to capture the discounted value of future reputation benefits and the use of a precedent, rather than by the immediate saving of penalties incurred, in first instance.

12.5 Conclusions

This paper has attempted to provide some insights into how businesses alter behaviour under an antitrust regime switch by using qualitative indicators extracted from intra-firm files, inter-firm correspondence and the files of regulators. *A priori* it might be expected that the responses of firms to a new regulated jurisdiction would range between continuation of collusion by 'secret' means or the emergence of tacit collusion, to reversion to more competitive practices (Posner 1976, p. 47). All too often, however, researchers lack the opportunity or are prevented by the currency of events to examine the reaction functions of firms facing changes in legal rules. The use of historical case studies, a before-and-after method in this instance,

[15] Lord Parker in Attorney-General of the Commonwealth of Australia v. The Adelaide Steamship Company Limited and Others (1913) AC 781 at p. 816.

provides evidence of the usually unobservable beliefs, intentions and strategic options examined by firms.

The avoidance and defensive strategies defined and described through the explanatory case studies show that firms sought to maintain their cartel advantages by structural alterations to agreements or by evading antitrust application using legal arguments. The tactic of avoidance created a dilemma for cartels. The move to secret collusion altered the incentive structure of these agreements, leaving firms with a new set of tradeoffs to calculate and perhaps exploit. In the case of the coal-shipping cartel, the move to a secret agreement was to speed up its demise. In the cases of tobacco and mineral oil, the restrictive practices continued despite the now regulated environment — perhaps not surprising given the small number of firms in both instances.[16] The peculiarities of Australian Federal-State relations, as manifested in antitrust jurisdiction, permitted cartels — such as the dried fruit and confectionery trade associations — to continue colluding, albeit at the interstate level by 'understandings'.

Once a cartel had been detected by the regulator, they switched resources to mount defensive strategies arguing for legality under the law. In the coal vend case, the decision to fight was most likely motivated by the expected future benefits of a successful defence (or appeal) rather than by the desire to avoid monetary penalties. In the other cases, the cartels undertook simulation studies to ensure that revised behaviour was consistent with a defensive position.

References

Barwick, G. 1963, *Australian Proposals for Legislation for the Control of Restrictive Trade Practices And Monopolies*, Government Printer, Canberra.

Binmore, K. 1992, *Fun and Games: A Text on Game Theory*, D.C. Heath, Lexington, Massachusetts.

Broeze, F. 1992, '"A great Frankenstein": The West Australian shipping association 1884–1906', in Broeze, F. (ed.) *Private Enterprise, Government and Society: Studies in Western Australian History*, Centre for Western Australian History, University of Western Australia, Perth, pp. 49–62.

Brunt, M. 1994, 'The Australian antitrust law after 20 years — A stocktake', in Round, D.K. (ed.), *The Australian Trade Practices Act 1974: Proscriptions and*

16 Indeed, Selten's (1973) argument that '4 are few and 6 are many' seems accurately to describe these industries, although with the tobacco industry the existence of five members certainly created incentives to cheat for member firms.

Prescriptions for a More Competitive Economy, Kluwer Academic, Dordrecht, pp. 25–68.

Castles, A.C. 1959, 'Australian anti-monopoly legislation', *American Journal of Comparative Law*, vol. 8, no. 1, pp. 82–8.

Donald, B.G. and Heydon, J.D. 1978, *Trade Practices Law: Restrictive Trade Practices, Deceptive Conduct and Consumer Protection*, vol. 1, Law Book Co., Sydney.

Fleming, G.A. 1999, 'Collusion and price wars in the Australian coal industry during the late nineteenth century', *Business History*, forthcoming.

Fleming, G.A. and Terwiel, D. 1997, 'How successful was early Australian antitrust legislation? Lessons from the Associated Northern Collieries 1906–11', in *Business, Public Policy and Performance*, Canberra, pp. 98–117.

Hopkins, A. 1978a, *Crime, Law and Business: The Sociological Sources of Australian Monopoly Law*, Australian Institute of Criminology, Canberra.

—— 1978b, 'Antitrust and the bourgeoisie: 1906 and 1965', Chapter 3 in Wheelwright, E.L. and Buckley, K. (eds), *Essays in the Political Economy of Australian Capitalism*, Australia and New Zealand Book Company, Sydney, pp. 87–109.

Hunter, A. 1961, 'Restrictive practices and monopolies in Australia', *Economic Record*, vol. 37, no. 77, pp. 25–52.

—— (ed.) 1963, *The Economics of Australian Industry: Studies in Environment and Structure*, Melbourne University Print, Melbourne.

Joskow, P.L. and Rose, N.L. 1989, 'The effects of economic regulation', in Schmalensee, R. and Willig, R.D. (eds), *Handbook of Industrial Organization*, vol. 2, North Holland, Amsterdam, pp. 1451–506.

Karmel, P.H. and Brunt, M. 1963, *The Structure of the Australian Economy*, Cheshire, Melbourne.

McGuire, L. 1998, 'Case studies: Storytelling of scientific research method?', in Browne, M., Steane, P. and Forster J. (eds), *Cases in Strategic Management*, Macmillan Education Australia, South Yarra, pp. 41–61.

Merrett, D. and Ville, S. 1998, 'The development of large scale enterprise in Australia, 1910–1964', *Business History*, forthcoming.

Norman, N.R. 1976, *Trade Practices Regulation in Australia: An Analysis*, Australian Industries Development Association, Canberra.

—— 1994, 'Progress under pressure: The evolution of antitrust policy in Australia', in Round, D.K. (ed.), *The Australian Trade Practices Act 1974: Proscriptions and Prescriptions for a More Competitive Economy*, Dordrecht, pp. 69–87.

Pengilley, W. 1979, *Concept, History and Enforcement of Australian Laws Against Restrictive Trade Practices*, Leo Cussen Institute for Continuing Legal Education, Melbourne.

Posner, R. 1976, *Antitrust Law: An Economic Perspective*, University of Chicago Press, Chicago.

Ransom, A. 1981, 'Monopolisation and price fixing — Good cases making bad law' *Federal Law Review*, vol. 12, no. 4, pp. 308–43.

Rasmusen, E. 1994, *Games and Information: An Introduction to Game Theory*, Cambridge, Massachusetts.

Round, D.K. (ed.) 1994, *The Australian Trade Practices Act 1974: Proscriptions and Prescriptions for a more Competitive Economy*, Kluwer Academic, Dordrecht.

Ryan, B., Scapens, R.W. and Theobald, M. 1992, *Research Method and Methodology in Finance and Accounting*, London.

Selten, R. 1973, 'A simple model of imperfect competition where four are few and six are many', *International Journal of Game Theory*, vol. 2, pp. 141–201.

Sheridan, K. 1968, 'An estimate of the business concentration of Australian manufacturing industries', *Economic Record*, vol. 44, no. 105, pp. 26–41.

Stalley, D.J. 1958, 'Federal control of monopoly in Australia', *University of Queensland Law Journal*, vol. 3, pp. 258–89.

Terwiel, D. 1996, The 1906 Australian Industries Preservation Act and the Newcastle vend case: A case study in the early development of restrictive practices, BEc Hons thesis, Australian National University, unpublished.

Ville, S. 1998, 'Business development in colonial Australia', *Australian Economic History Review*, vol. 38, no. 1, pp. 16–41.

Wilkinson, H.L. 1914, *The Trust Movement in Australia*, Melbourne.

13 Market value, R&D and intellectual property in large Australian firms

*Mark Rogers**

This paper considers the R&D and intellectual property activity of large firms in Australia. R&D data are available for the 1994 to 1996 period and data on patent, trade mark and design applications for 1996. An overview of the extent and nature of these activities is provided. The second half of the paper analyses the link between the share market value of the firm and the extent of R&D and intellectual property activity. The findings suggest that R&D, patent and trade mark activity can all be positively and significantly associated with market value. However, the significance of the results on R&D and trade mark does not hold under some specifications.

13.1 Introduction

This paper is concerned with assessing the role of innovation in firm performance. There is widespread popular support for the general assertion that innovation is important in determining the performance of a firm (eg Business Council of Australia 1993, Australian Manufacturing Council 1994, Mortimer 1997). However, quantifying the extent of innovation in a firm, a necessary prerequisite for formally testing this assertion, is a difficult task. There are many different types of innovative activities which vary across firms and across time. In quantifying innovation, many authors use research and development expenditures (R&D) as a proxy. Also, the use

* The research behind this paper was funded from two sources. The first was the 'Innovation Scoreboard' project which has received support from BRW Media, IBIS Business Information Pty Ltd and the IR&D Board. The second was a collaborative research program entitled 'The Performance of Australian Enterprises: Innovation, Productivity and Profitability' which was generously supported by the Australian Research Council and the following collaborative partners: Australia Tax Office, Commonwealth Office of Small Business, IBIS Business Information Pty Ltd, the Industry Commission (now Productivity Commission), and Victorian Department of State Development. The views expressed in this paper represent those of the author and not necessarily the views of the collaborative partners.

of patent data is common in the literature, which can be thought of as either an output of the innovation process or an input.[1] Although R&D and patents are likely to be major aspects some firms' innovation activities, firms may also devote substantial resources to areas such as managerial and organisational change, the purchase of latest technology, training and marketing. Data on such areas of innovation are difficult to obtain, although the situation in Australia has been much improved by the ABS *Innovation Survey* and the ABS *Growth and Performance Survey* (ABS 1995, Phillips 1997).

The innovation proxies used in this paper include not only R&D and patents, but also trade marks and designs. A trade mark is a symbol that distinguishes the firm's goods or services from other goods. For example, the symbol can be a letter, word, phrase, sound or logo that allows customers to identify a product or service. Although there has been little empirical economic research into trade marks or brand names (Wilkins 1992), the theoretical industrial organisation literature does discuss the role of brand names. Trade mark activity is likely to be positively linked to the launch of new or improved products or services and expenditure on marketing. Moreover, whereas R&D and patents tend to be used by manufacturing firms, trade mark usage does extend to service sector firms (Rogers 1998a). In contrast, designs — which are used to protect the visual appearance of manufactured products — are dominated by manufacturing firms (see Rogers 1998b for more discussion of innovation proxies and the use of trade marks and designs).

Conducting R&D or applying for intellectual property rights obviously involves an investment decision by the firm. Since such activities are relatively widespread (see below), a natural assumption is that the firm will, on average, experience a net increase in profitability or other performance measure. This paper sets out to test this hypothesis using the share market value of the firm as a measure of firm performance. The market value of a firm should, in theory, represent the market's valuation of the expected future stream of profits, which in turn is based upon an assessment of the return that can be generated from a firm's tangible and intangible assets. The level of intangible assets includes various dimensions of the firm's activities, including past innovative activities and human capital. Since the innovation data used are relatively recent, the use of share market value is vital (since there is not a long time series of profit data following the innovation data).

The structure of the paper is as follows. Section 13.2 provides an overview of the R&D and intellectual property data. Section 13.3 considers the link between R&D

[1] Griliches (1990) states that Jacob Schmookler — who was the first economist to intensively study patent data — started his research considering patents as an output measure, but by the end of his research decided that 'patents became an index of inventive 'activity', primarily an input rather than an output index' (Griliches 1990, p. 1670).

and market value using three years of R&D data (1994 to 1996). Section 13.4 extends the analysis to include patent, trade mark and design activity, although data limitations mean this is only available for 1996. The final section concludes the paper.

13.2 Overview of R&D and intellectual property data

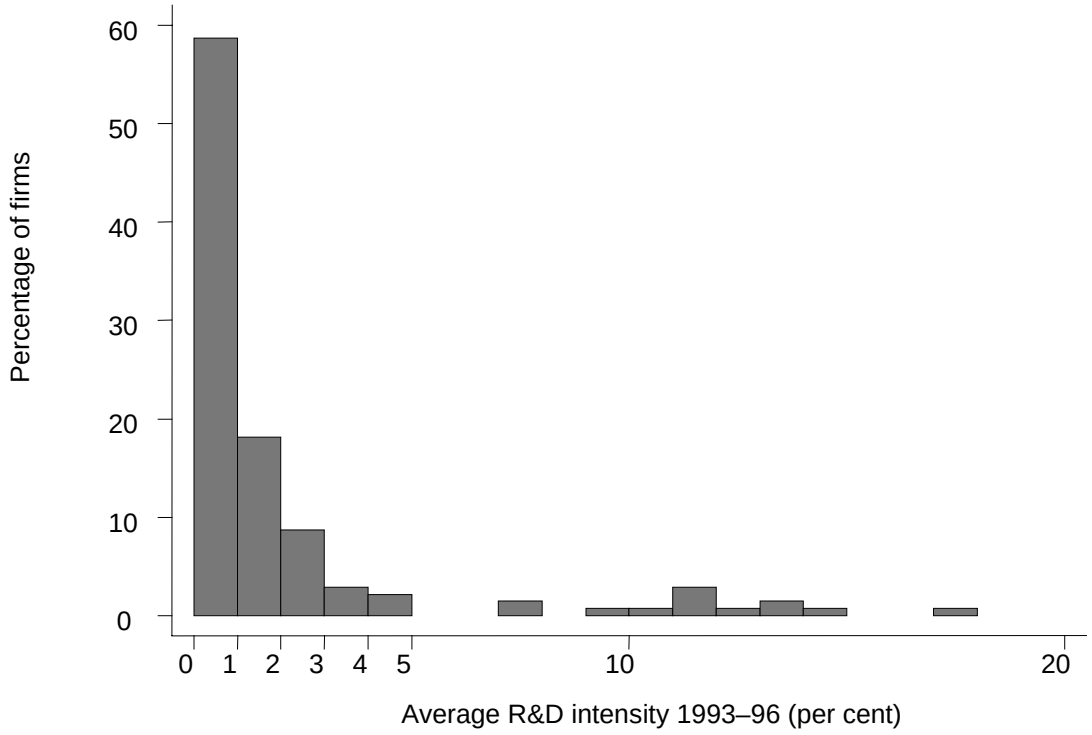
This section provides an overview of the R&D and intellectual property (IP) data. This provides background for the subsequent sections. The sample of firms considered in this section is not the same as those in subsequent sections. This is because the market value sections can only consider firms that are listed on the Australia stock exchange, which is a small sub-set of firms.

R&D data

The R&D data come from the IBIS Business Information Pty large firm data base, which contains data on around 2 800 firms. Data in this data base come from a variety of sources including published accounts, the ASX and direct surveys of firms by IBIS. It is important to note that there is no requirement for firms to report R&D expenditures. This means that the IBIS data base is unlikely to provide an unbiased sample — firms that wish their R&D expenditures to remain secret will not appear. Since such firms may be highly innovative, this is of concern, but is an issue that cannot be easily rectified. To provide an overview of the R&D data, a sample of non-public firms that have R&D data in the each of the years 1993, 1994, 1995 and 1996 is used. This four year balanced panel contains 138 firms. These firms in 1996 have in total $1.5 billion of R&D. In comparison, the ABS estimates that a total of $4 billion was spent on R&D by the Australian private sector in 1995-96 (ABS 8104.0). Since the IBIS data base has a much greater coverage of large firms, a more appropriate comparison would be the ABS estimate of R&D expenditure by large firms. For firms with over 1000 employees, the ABS estimates R&D expenditures of $1.5 billion. The 138 firm sample has total R&D expenditure of $1.2 billion. These comparisons show that, although the sample size may appear small, it contains a high proportion of the Australian total R&D for large firms.

A common way of analysing R&D expenditures is to consider the R&D intensity (R&D / total revenue). This controls for the fact that large firms will have much greater absolute R&D expenditures and hence gives a proxy for innovative effort. When considering the R&D intensity of the sample, there are two basic properties of the data that should be highlighted. First, the distribution of R&D intensities is skewed to the right. This can be seen in figure 13.1, which shows a histogram of the 138 firm sample (the distribution of the average R&D intensity over the period 1993

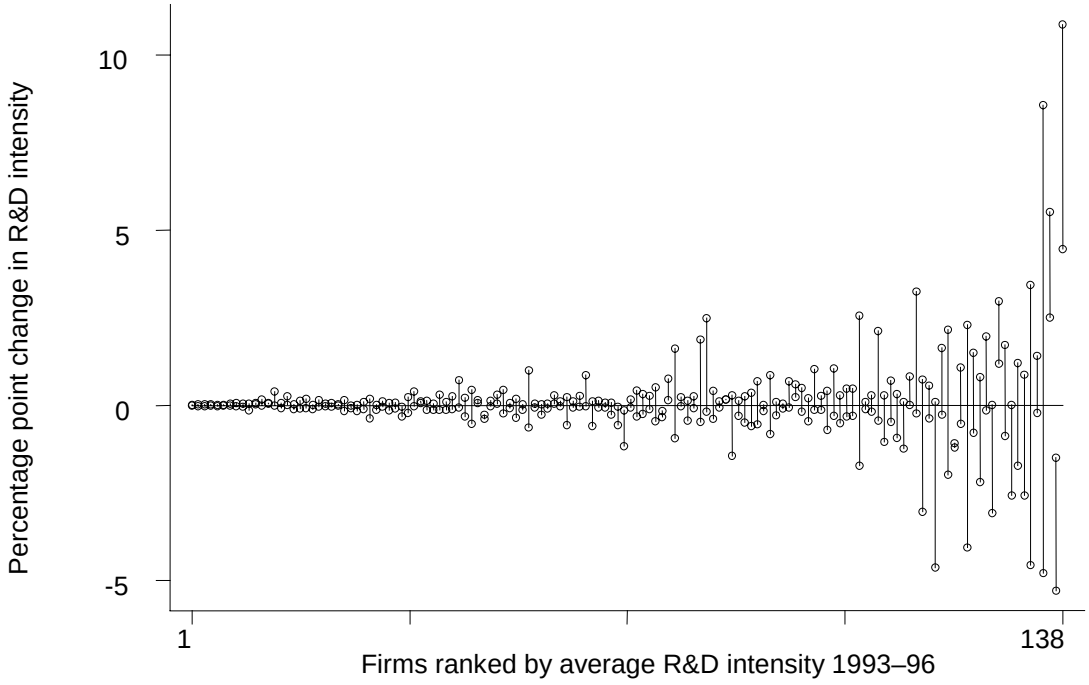
Figure 13.1 **Histogram of R&D intensity (138 firm sample)**



to 1996 is shown). Almost 60 per cent of firms have an R&D intensity below 1 per cent, but the range of intensities rises to close to 20 per cent.

A second important feature of the data is that the R&D intensities of firms vary over time, especially for firms with high R&D intensity. This would be expected if highly R&D intensive firms have 'lumpy' R&D projects, implying that expenditures need to vary from year to year. Alternatively, highly R&D intensive firms may experience cash flow problems that force expenditures to fall. The former reason is benign, the latter may be a cause for more concern if capital markets are for some reason not allocating funds to worthwhile R&D projects (see National Investment Council 1995 for a discussion of innovation financing). Figure 13.2 depicts one method of assessing the volatility of R&D intensities over time. Prior to graphing the figure, the change in year on year R&D intensity is calculated for each firm (ie if R&D intensity moved from 5 per cent in 1993 to 7 per cent in 1994, the change is 2 percentage points). Each vertical line represents a firm, with the firms ordered by average R&D intensity (1993–96) from left to right (with the highest average R&D intensity firm being represented by the far right vertical line). Each vertical line joins the maximum and minimum changes in R&D intensity for that particular firm. Hence, a long vertical line represents relatively large changes in intensity. If the line extends below zero on the y-axis then R&D intensities have fallen at some point. The figure shows the high volatility of R&D intensities, which increases with the

Figure 13.2 **Volatility of R&D intensity (1993–96)**



average level of R&D intensity. Note also that R&D intensities do fall relatively often (ie many of the vertical lines extent below the horizontal axis). Both the skewness of the distribution of R&D intensity and the year on year volatility suggest that influential variables could be a problem in regression analysis.

Intellectual property data

The data on intellectual property (patents, trade marks and designs) are for applications in the 1996 calendar year. These data are compiled by matching the names of firms in the IBIS data base against IP Australia's *Annual Record of Proceedings*, which is the complete list of all applications made or designated in Australia. The parent company name and the names of all majority owned subsidiaries were checked against the *Proceedings*. The use of applications, not grants (ie successful applications), can be justified for two reasons. First, there can be lag of a few years before an application is granted. Hence, the use of grants might provide an out of date assessment of a firm's current innovative activities. Second, since the use of the applications data is intended as a proxy for current innovative activities, and, given that innovation is normally defined as ideas that are new to the firm, the use of applications has some merit. This is because, even if the application is unsuccessful due to the idea existing somewhere else, there is still an implication that the firm is making efforts to innovate. Obviously, it would be preferable to have

data on both applications and grants to test such ideas, but creation of such data sets is a costly and time intensive process.

It is worthwhile taking a broad look at the intellectual property (IP) data since the analysis of IP data in Australia is rare. As stated, IP data were matched against firms in the IBIS data set. Table 13.1 shows the percentage of these firms who made at least one IP application in 1996. The proportion of firms who spent a positive amount on R&D is also shown for reference.

Table 13.1 shows that trade mark activity is the most common innovative activity (17.6 per cent), followed by R&D and patents. If the sample is restricted to large firms, all of the innovative activities become more common. It is also of interest to consider the subset of manufacturing firms, since these are thought to be the heaviest users of R&D and IP. Table 13.2 shows the equivalent percentages for the 581 manufacturing firms in the IBIS data base. All of the percentages are higher for manufacturing firms, with R&D now being the most common type of activity.

Table 13.1 **Extent of IP and R&D activity**
(per cent)

Innovation activity	*Enterprises in IBIS undertaking activity*	*Largest enterprises undertaking activity*
R&D	11.6	19.2
Patent applications	4.5	8.2
Trade mark applications	17.6	31.0
Design applications	2.1	4.3

Percentages based on 2 629 firms. Largest firms are those with total revenue exceeding $100 million.

Table 13.2 **Coverage of innovation indicators in manufacturing**
(per cent)

Innovation activity	*Enterprises undertaking activity*	*Largest enterprises undertaking activity*
R&D	30.0	43.3
Patent applications	11.5	18.7
Trade mark applications	27.4	44.3
Design applications	6.9	11.8

13.3 Market value and R&D

The analytical framework used for this section is based on the Tobin Q approach. This approach essentially considers that the market value of the firm is related to the

value of tangible and intangible assets. Griliches (1981) suggests that the market value (V) is given by:

$$V = q(A + K) \qquad [13.1]$$

where A is the stock of tangible assets of the firm, K is the stock of intangible assets and q is the 'current market valuation coefficient' of the firm's assets.[2] The latter reflects variations across firms in their risk and monopoly positions and is assumed to equal:

$$q_{it} = \exp(m_i + d_t + u_{it}) \qquad [13.2]$$

where m_i is a permanent firm effect, d_t is the market effect at time t and u_{it} is an independently distributed error term. Hence, the term q allows for the fact that the market valuation may vary across firms and time and there may also be 'noise' in such valuations.

Equations [13.1] and [13.2] can be rearranged to yield (using the approximation $\log(1+\varepsilon) \approx \varepsilon$)[3]

$$\log V_{it} = m_i + d_t + \log A_{it} + K_{it} / A_{it} + u_{it} \qquad [13.3]$$

Griliches (1981) takes the log$A$ term to the left hand side of [13.3] (and uses log(V/A) as a dependent variable). In this analysis, the coefficient on the log$A$ term is not constrained to be one, which follows Hall (1993) and Bosworth and Rogers (1998). In order to estimate [13.3], data are required on tangible assets, market value and intangible capital stocks. The data for tangible assets come directly from the IBIS data set. The market value is the sum of the share market valuation (obtained from the Australian Stock Exchange) and book value of debt (from IBIS data base). These represent the amount that a buyer would have to pay in order to own the assets of the firm.[4]

---

[2] It might appear that equation [13.1] is an identity. This would be the case if A and K were measured by their true value — in which case q would be one. However, the magnitudes of A and K are based on accounting methods, which do not necessarily correspond to economic value. Thus, the functional form of [13.1] may not be appropriate. For example, if for some reason the level of A and K, as measured by accountants, somehow interact, then the marginal products of A and K are not necessarily independent and equal (as they are in [13.1]). In such cases a Cobb-Douglas or other functional form might be more appropriate.

[3] The accuracy of this approximation depends on the value of ε — in this case K/A — being close to zero. This may not be the case for some firms and is, perhaps, increasingly less likely to hold as the role of intangible assets increases.

[4] Obtaining the share market valuation (or market capitalisation) is not straightforward. The ASX define market capitalisation as the share price times the number of normal shares on issue. This ignores some types of preference shares and other securities, although the latter are relatively

Estimation of [13.3] also requires data on the value of intangible assets (K). The IBIS data base does contain a book value of intangible assets for some firms (as reported in the annual accounts). The exact composition of intangibles will vary from firm to firm. A major component of intangible assets is likely to be a valuation for goodwill. However, capitalised past R&D, patent, trademark and licence valuations may also be included. Despite the availability of such a figure, it appears likely that accounting valuations may fail to incorporate the full extent of the firm's intangible assets. Therefore, previous studies have used proxies, such as R&D expenditures, advertising expenditures, recent cash flow, recent growth in output and patents (Hall 1993, Connolly and Hirschey 1988, Megna and Klock 1993). Such proxies are *flow* variables for what should a *stock* measure. Therefore, each of these proxies is normally entered separately and it is assumed that the flow is proportional to the stock. This is the method used here.

Estimation of equation [13.3] implies the use of a fixed effects or LSDV (least square dummy variable) model, due to the presence of the m_i term — the (firm) fixed effect. Before using a fixed effect model, some analysis of a data set containing the average values over the 3 year period is undertaken here (ie the data set is collapsed to 60 firm observations with the variables being an average over the 1994 to 1996 period). The advantage of this method is that any extreme values in a particular year will be smoothed. As noted above, the volatility over time of firm's R&D intensity is particularly severe for some firms.

Table 13.3 shows some regression results from the 60 firm (averaged) sample. The regressions have extremely high R^2 because of the 'log of tangible assets' explanatory variable. The coefficient on this term is just above 1, indicating that a 1 per cent rise in tangible assets increases market value by 1.05 per cent. The first regression (R1) simply includes the explanatory variables shown and a constant term. Both the coefficients on the R&D and intangible assets terms are positive and significant. The next two regressions (R2 and R3) both contain a set of industry dummies (2 digit level) to test the robustness of regression R1. Regression R3 also contains the percentage growth in revenue over the period 1994 to 1996. This variable was used by Hall (1993) as a rough proxy for intangible asset stocks. The addition of these variables does not greatly change the magnitude or significance of the coefficients on the R&D and intangible assets terms. A regression was also run

unimportant for the market as a whole (1 per cent of entire market). For 1995 and 1996, it was possible to obtain daily market capitalisation data (for the days the company trades) from the Securities Industry Research Centre of the Asia-Pacific at the University of Sydney. These data were averaged to find the average market capitalisation over the year. However, daily data were not available for all of the firms in 1994. Where daily data did not exist, the market capitalisation comes from the ASX Yearbook 1995 (for the single day, 28 February 1994) or the ASX Monthly Report (for 31 May 1994).

excluding those firms with an R&D/total revenue ratio of more than 10 per cent. In this regression, although the coefficient on the R&D term was positive, it was insignificant. This suggests that the high R&D intensity firms are important in establishing the significant association shown in table 13.3.

The coefficients on 'R&D/tangible assets' and 'intangible/tangible assets' are not straightforward to interpret as the log of tangible assets is also included in the regression. One method of assessing their economic impact is to assume that R&D rises, with the value of tangible assets held constant. If this is the case, a 1 per cent rise in the R&D/tangible assets ratio implies market value will rise by 2 per cent (as always, causation is not proved by such regression results).

Equation [13.3] suggested the inclusion of a firm specific effect (m_i) which allows the market's valuation of each firm to vary. This can be estimated by entering a dummy variable for each firm (the LSDV method) or by using the within, or fixed effect, estimator. Table 13.4 shows the results of using the LSDV estimation method. The first regression (R4) is run on the 180 firm sample (60 firms over 3 years) with dummy variables for 1994 and 1995 also added. The high R^2 for this regression is due to the inclusion of the 60 firm specific dummies. The estimated coefficients in regression R4 are dramatically different from the averaged sample regressions shown in table 13.3. The coefficient on the log of tangible assets is 0.36, the coefficient on the R&D term is negative and significant, and only the coefficient on the intangible asset term is remotely close to its value in regressions R1 to R3.

Table 13.3 **Regressions for three year averages**

Dependent variable: log of market value

Explanatory variable	*R1*	*R2*	*R3*
		Industry dummies	
log (tangible assets)	1.048	1.068	1.063
	(49.17)	(39.10)	(38.79)
R&D / tangible assets	2.077	2.268	2.073
	(2.54)	(2.68)	(2.31)
Intangible / tangible assets	0.73	0.827	0.787
	(4.36)	(7.42)	(6.64)
Growth in revenue 94 – 96			0.004
			(1.425)
R^2	0.977	0.986	0.987
No. of observations	60	60	60

t-statistics in brackets (based on White's robust standard error). Industry dummies at 2-digit ANZSIC level.

Such differences in results between the two models imply that there may be some difficulties with the specification of the models.

To interpret these differences it is important to remember that the coefficients in table 13.4 are estimated on the basis of changes in a firm's value over time. A first possibility is therefore that the volatility of R&D intensities over time, especially for some high R&D intensity firms, is leading to influential observation problems. A quick method of assessing this issue is to simply exclude high R&D intensity firms from the regression. Regression R5 excludes those observations with an R&D intensity over 10 per cent. Apart from the year dummies, the coefficients from this regression are different from those in the full sample regression. In particular, the coefficient on the R&D term is now positive (although not significant). The coefficient on the log of tangible assets has also increased. This suggests that influential observations may be driving the results in regression R4.

Although influential observations may account for some of the differences between tables 13.4 and 13.5, there is also the possibility that the model is misspecified. This could be due to (a) the functional form of equation 3, or (b) the omission of important explanatory variables. The latter case appears highly likely and the next section extends the analysis to include data on patents, trade marks and designs. In addition, innovation, and the intangible assets it creates, are due to a host of other

Table 13.4 **LSDV estimations**

Dependent variable: log of market value

Explanatory variable	*R4* *Full sample*	*R5* *R&D intensity < 10%*
log (tangible assets)	0.366	0.533
	(2.21)	(2.78)
R&D / tangible assets	-4.354	3.456
	(-2.01)	(0.84)
Intangible assets / tangible assets	1.063	0.784
	(2.56)	(1.83)
Dummy for 1994	0.049	0.058
	(0.96)	(1.12)
Dummy for 1995	-0.067	-0.070
	(-1.99)	(-2.12)
R^2	0.991	0.992
No. of observations	180	162

t-statistics in brackets (based on White's robust standard error). Industry dummies at 2-digit ANZSIC level.

variables for which there is little data. Concerning the functional form of [13.3], it was footnoted that Cobb-Douglas specifications could be investigated. Equally, it may be that the market valuation of firms is based on an equation such as:

$$y_{it} = \beta_1 \bar{x}_i + \beta_2 (x_{it} - \bar{x}_i)$$

Such an equation states that the impact of *x* on *y* is due to the mean level of *x* and also on the deviation of *x* from its mean value. A rationale for such an equation may be that the share market bases its valuations both on medium term averages and also short term deviations.

13.4 Market value, R&D and intellectual property

This section pursues one of the possible misspecifications noted above, namely the omission of variables on intellectual property (IP). Patents, trade marks and designs are assumed to be proxies for the stock of intangible assets in the same way that R&D expenditures were used above (eg the ratio of patent applications to tangible assets is used as a proxy for *K/A* in equation [13.3]). There is only one year of data on the applications of patents, trade marks and designs for firms in the IBIS data set (1996). Therefore, it is a major assumption to assume that a flow measure of IP activity in a single year is proportional to part of the intangible asset stock. The greater the variation in IP activity from year to year, the worse this assumption will be. As noted, R&D intensities can vary substantially over time and presumably the same is true for IP data. As more data become available this issue can be investigated, but for now such issues imply a lower likelihood of finding significant relationships between IP usage and market value.

Table 13.5 shows the results from some regressions that include intellectual property variables. The first regression (R6) yields significant and positive coefficients on 'log of tangible assets', the ratio of intangible to tangible assets and the ratio of patent applications to tangible assets. The magnitude of the coefficient on the patent term suggests that a 1 per cent rise in this ratio (holding tangible assets constant) is associated with a 7 per cent rise in market value. However, a 1 per cent rise in this ratio is twice the mean value for the sample (see appendix 13A), implying this would be a major shift for any firm. This regression also shows that the R&D term has lost its significance and that neither the trade mark or design terms have significant coefficients.

The second regression (R7) includes a set of industry level dummies (at the 2 digit level). As noted above, share market valuations are likely to vary across firms and industries because of risk and monopoly power differences, and the inclusion of industry dummies goes some way towards controlling for this. Regression R7 shows the significance on the patent term is maintained and, although the coefficients on both trade marks and designs are both positive, neither are significant. The coefficient on the R&D/tangible assets ratio is still statistically insignificant.

Table 13.5 **Regressions including intellectual property (1996)**

Dependent variable: log of market value

Explanatory variable	*R6*	*R7*	*R8*	*R9*	*R10*
		Industry dummies			
log (A)	1.074	1.065	1.063	1.031	1.033
	(39.25)	(31.11)	(32.05)	(23.21)	(27.79)
R&D / A	0.603	0.615		2.71	1.024
	(0.76)	(0.68)		(2.01)	(1.16)
Intangible /A	0.644	0.516	0.486	0.609	0.524
	(2.61)	(2.20)	(2.19)	(3.56)	(3.54)
Patents/A	7.137	6.795	7.348		6.502
	(5.18)	(5.84)	(6.83)		(5.93)
Trade marks /A	0.38	0.792	0.878		
	(0.60)	(1.09)	(1.23)		
Designs / A	-3.963	2.935	3.09		
	(-0.93)	(0.62)	(0.67)		
Patent dummy				0.096	
				(0.81)	
Trade mark dummy				0.258	0.243
				(2.26)	(2.58)
Design dummy				-0.111	
				(-1.00)	
R^2	0.942	0.967	0.967	0.965	0.969
No. of observations	120	120	120	120	120

t-statistics in brackets (based on White's robust standard error). Industry dummies at 2-digit ANZSIC level. A = tangible assets. Dummies for patents, trade marks and designs are 1 for those firms that made at least one application for the respective property right in 1996. Regressions R7 to R10 have 2-digit industry dummies included.

One explanation for the insignificant coefficient on the R&D term is that R&D and IP may be positively correlated, leading to multicollinearity and a failure to identify the separate effects of both R&D and IP. The appendix contains the correlation coefficients for this sample and shows that R&D is positively correlated with patents and trade marks, with values of between 0.35 and 0.55. This may account for the results. Regression R8 omits the R&D to tangible assets ratio. The results show that the coefficient on the patent term is still significant, with the trade mark coefficient positive and improving in significance (to the 22 per cent level).

Although these results point to a positive role for patents in generating intangible assets and thereby market value, they are disappointing with respect to R&D, trade marks and designs. One issue worth investigating is that the actual number of IP applications is not a good proxy for the importance of these activities. The fact that patent, trade mark and design applications will vary greatly in their value (some, of course, may never be granted as property rights) implies that use of the actual numbers may introduce a misspecification. It should be noted that this is not just a problem associated with IP data. For example, it is highly unlikely that a dollar's worth of R&D creates the equivalent benefit across firms and over time — some firms will be much better than others at R&D. To try and assess this situation, a dummy variable was created for whether the firm made at least one patent, trade mark or design application in 1996 (the appendix gives the number of firms in each case). Regression R9 enters these dummy variables (again 2 digit level industry dummies are also entered). The results show that the patent and design dummies are not significant. However, the trade mark dummy is positive and significant (at the 1 per cent level). The magnitude of the coefficient, however, appears implausibly large (suggesting firms that made at least one trade mark application have a market premium of 25 per cent). Therefore, it appears that the trade mark dummy is picking up the influence of a number of firm characteristics. Regression R9 also shows that the coefficient on R&D is now positive and significant. Regression R10 drops the patent and design dummies and enters the patent to tangible assets ratio. Again, the coefficient on the patent ratio is positive and highly significant, although this appears to have caused the R&D coefficient to become insignificant.

These results suggest that intellectual property activity, even when proxied by a single year's data, does have some power in explaining market value. The patent to tangible assets ratio is strongly significant. Trade mark activity has a positive and significant association with market value when it is entered as a dummy variable.

13.5 Conclusions

This paper has provided some initial analysis on R&D and IP, and specifically their association with market value. The underlying reason for such work is to try and understand more about the relationship between innovation and firm performance. Section 13.2 provided an overview of the R&D and IP data. As noted, the distribution of R&D is skewed to the right. Although the graphs are not presented above, this is also the case with the distributions of patent, trade mark and design intensities. Section 13.2 also noted that R&D intensity varies substantially over time, especially for firms with high R&D intensities. The most common activity out of those considered is filing a trade mark application (18 per cent of large Australia firms did so in 1996) followed by R&D (12 per cent of firms).

The last half of this paper investigated the relationship between R&D and IP and market value (sections 13.3 and 13.4). In a sample of 60 listed firms (averaged over the 1994 to 1996 period), R&D was found to be positively and significantly linked to market value. The results imply an increase in R&D leading to a 1 per cent rise in the R&D to tangible assets ratio (holding tangible assets constant) is associated with a 2 per cent rise in market value. However, this result was not stable in the LSDV models (which enter a dummy variable for each of the 60 firms). For a sample of 120 listed firms for 1996, the patent to tangible asset ratio was positively and significantly associated with market value. This result appears stable to changes in the specification of the model. However, the role of R&D, trade marks and designs appears weak in the regressions (the coefficients are positive but usually insignificant). It appears that correlation between the various measures may be contributing to these results. When a dummy variable for trade mark activity is used, the coefficient is significant and positive, although its magnitude appears very large.

Overall, it appears that these activities are positively linked to market value, although the results are not very stable to changes in model specification. Given the data limitations, these results can be viewed as encouraging. It is important to remember that firms invest resources in undertaking such activities, so the natural assumption is that, on average, such investments payoff. Of course, the analysis also assumes that the share market can adequately assess (at least on average) the future payoffs of such investments, which may not be the case. These issues will be the subject of further research.

Appendix 13A

Abbreviations

lmv	Log (market value)
la	Log (tangible assets)
ita	Intangible / tangible assets
rta	R&D/ tangible assets
pa	Patent applications / tangible assets
ta	Trade mark applications / tangible assets
da	Design applications / tangible assets

Table 13.6 **Summary statistics for averaged (1994 to 1996) sample (n=60)**

	Mean	*Std Dev.*	*Min.*	*Max.*
lmv (average)	12.803	1.818	9.634	17.60
la (average)	12.684	1.767	9.567	17.23
Rta (average)	0.025	0.045	0.000	0.191
Ita (average)	0.104	0.191	0	1.296
growth in total revenue	12.895	14.87	-17.14	69.36

Table 13.7 **Summary statistics for LSDV (1994 to 1996) sample (n=180)**

	Mean	*Std Dev.*	*Min.*	*Max.*
la	12.685	1.763	9.454	17.366
Rta	0.025	0.046	0.000	0.212
Ita	0.104	0.193	0	1.479

Table 13.8 **Summary statistics for 1996 sample (n=120 firms)**

Variable	*Mean*	*Std Dev.*	*Min.*	*Max.*	*No. of obs>0*
lmv	12.610	1.850	8.593	17.702	120
la	12.528	1.714	9.679	17.366	120
ita	0.105	0.195	0.000	1.174	91
rta	0.022	0.044	0.000	0.228	118
pa	0.005	0.021	0.000	0.207	35
ta	0.017	0.049	0.000	0.337	61
da	0.001	0.005	0.000	0.054	19

Table 13.9 **Correlation coefficients for 1996 sample (n=120 firms)**

	lmv	*la*	*ita*	*rta*	*pa*	*ta*	*da*
lmv	1						
la	0.9647	1					
ita	-0.1324	-0.1941	1				
rta	-0.2612	-0.3193	-0.0992	1			
pa	-0.039	-0.1301	-0.0035	0.5494	1		
ta	-0.1768	-0.2088	0.0523	0.3464	0.1643	1	
da	0.0442	0.0163	0.5205	-0.053	0.0479	0.0882	1

References

ABS (Australia Bureau of Statistics) 1995, *Innovation in Australia Manufacturing*, Cat. no. 8116.0.

—— 1998, *Research and Experimental Development, Businesses, Australia*, Cat. no. 8104.0.

Australia Manufacturing Council 1994, *The Wealth of Ideas: How Linkages Help Sustain Innovation and Growth*, Australian Manufacturing Council, Melbourne.

Bosworth, D. and Rogers, M. 1998, *Research and Development, Intangible Assets and the Performance of Large Australia Companies*, Working Paper no. 2/98, Melbourne Institute of Applied Economic and Social Research, University of Melbourne, January.

Business Council of Australia 1993, *Managing the Innovating Enterprise*, Business Council of Australia, Melbourne.

Connolly, R. and Hirschey, M. 1988, 'Market value and patents: A bayesian approach', *Economics Letters*, vol. 27, no. 1, pp. 83–7.

Griliches, Z. 1981, 'Market value, R&D, and patents', *Economic Letters*, vol. 7, no. 2, pp. 183–7.

—— 1990, 'Patent statistics as economic indicators: A survey', *Journal of Economic Literature*, vol. 28, no. 4, pp. 1661–707.

Hall, B. 1993, 'The stock market valuation of R&D investment during the 1980s', *American Economic Review*, vol. 83, no. 2, pp. 259–64.

Megna, P. and Klock, M. 1993, 'The impact of intangible capital on Tobin's q in the semiconductor industry', *American Economic Review*, vol. 83, no. 2, pp. 265–9.

Mortimer 1997, *Going for Growth: Business Programs for Investment, Innovation and Export*, Department of Industry, Science and Technology, Canberra.

National Investment Council 1995, *Financing Growth: Policy Options to Improve the Flow of Capital to Australia's Small and Medium Enterprises*, Department of Industry, Science and Technology, Canberra.

Phillips, R. 1997, *Innovation and Firm Performance in Australian Manufacturing*, Industry Commission Staff Research Paper, AGPS, Canberra, September.

Rogers, M. 1998a, 'Innovation in large Australian enterprises', *Mercer-Melbourne Institute Quarterly Bulletin of Economic Trends*, no. 2.98, pp. 50–60.

—— 1998b, *The Definition and Measurement of Innovation*, Melbourne Institute Working Paper no. 10/98, Melbourne Institute of Applied Economic and Social Research, University of Melbourne, May.

Wilkins, M. 1992, 'The neglected intangible asset: The influence of the trade mark on the rise of the modern corporation', *Business History*, vol. 34, no. 1, pp. 66–95.

APPENDIXES

APPENDIXES

A Conference program

CONFERENCE THEME: PRIVATISATION, REGULATION AND REFORM

Monday, 6th July 1998

8.30 am Registration

9.15 am Introduction

Plenary session 1

9.30 am Professor Paul Joskow (Massachusetts Institute of Technology)
Electricity sector privatisation, competition and regulatory reform

11.00 am Morning tea

Contributed paper session 1

Session A: Industrial economics and theory I

11:30 am J. Jude Kline (ANU)
Carros and stick games

Steffan Ziss (University of Sydney)
Horizontal mergers and delegation

Dan Sasaki (University of Melbourne)
Endogenous co-leadership when demand is unknown

Session B: International comparisons

11.30 am Ray Trewin (ANU)
Analysis of Australian, Indonesian and Malaysian agricultural and food trade market shares

Nittaya Soonthonsiripong (University of Adelaide)
Regulatory reform and telecommunications structure in developing countries: A case study of local telephone network in Thailand

Jose Vargas (Carleton University)
The increasing economic relations between Mexico and Canada and its influences on some variables of Mexican organisational environment

1.00 pm Lunch

Panel session 1

2.00 pm Recent developments in economic theory

3.30 pm Afternoon tea

Contributed paper session 2

Session A: Regulation and investment

3.45 pm Joshua Gans (University of Melbourne)
A primer on access regulation and investment

Rohan Pitchford (ANU)
A simple theory of deregulation

Corrine Chaton (Toulousse)
Uncertainty and anticipated decommissioning or investment

Session B: Industry studies: Motor vehicles and roads

3.45 pm Harry Clarke (La Trobe University)
The future of the Australian motor vehicle industry

Neville Norman (University of Melbourne)
Can pausing agreed tariff reductions ever be rational? A rational industry economics approach

Barry Abrams (Productivity Commission)
A comparison of institutional arrangements for road provision

7.00 pm Conference Dinner

Tuesday, 7th July 1998

Contributed Paper Session 3

Session A: Research and development and patents

8.30 am Reiko Aoki (University of Auckland)
To disclose or not disclose: The impact of public disclosure of patent applications on firms' patenting and R&D behaviour

Mark Rogers (University of Melbourne)
Market value, R&D and intellectual property in large Australian firms

John King (Vanderbilt University)
Market value of patents and early patent disclosure: An event study

Session B: Industry studies and policy

8.30 am Loretta Mester (Federal Reserve Bank of Philadelphia)
Efficiency of and productivity changes in the US commercial banking industry

Tim Hazledine (University of Auckland)
Oligopoly behaviour in the trans-Tasman air travel market: The kiwi international story

Grant Fleming (ANU)
How does antitrust affect business behaviour? Lessons from Australian corporate experience

Plenary session 2

9.45 am Professor Robert Cooter (University of California, Berkeley)

11.15 am Morning tea

Panel session 2

11.30 am Privatisation

1.00 pm Lunch

Panel session 3

2.00 pm The Asian financial crisis

3.30 pm Afternoon tea

Contributed Paper Session 4

Session A: Industrial economics theory II

3.45 pm Heli Koski (ETLA - Research Institute of the Finnish Economy)
Ownership structure and adoption of multiple communication technologies

Martin Pietz (University of Alicante)
Intermediation can replace certification

Kieron Meagher (University of New South Wales)
The impact of hierarchies on wages

Craig Freedman (Macquarie University)
Do economic journals obey economic prescription? The case of academic referees

Session B: Australian policy issues

David Allen (Edith Cowan University)
A model of the speed of Australian interest rate adjustment featuring a time-series approach

Arif Syed (Department of Industry, Science and Tourism)
Identification of the factors affecting private investment in Australian regional development with specific reference to new growth theory

Jim Longmire (University of Southern Queensland)
Access to rail: A consideration of some issues relating to access pricing and economies of scale, scope and utilisation capacity

B Authors' affiliations

Mr Barry Abrams	Productivity Commission, Melbourne
Dr Harry Clarke	School of Business, La Trobe University, Bundoora
Mr Peter Cribbett	Productivity Commission, Melbourne
Dr Grant Fleming	Department of Commerce, Australian National University
Mr David Greig	Transport Reform Unit, Victorian Department of the Treasury and Finance
Professor Simon Grant	Department of Economics, Australian National University
Dr Don Gunasekera	Productivity Commission, Melbourne
Dr Midori Hirokawa	Department of Economics, Harvard University, Cambridge, Massachussits, 02138 USA; and Faculty of Economics, Hosei University, 4232 Aihara, Machida-shi, Tokyo, Japan
Dr Steven Kates	Australian Chamber of Commerce and Industry, Canberra
Professor Stephen King	Department of Economics, Melbourne University
Dr J. Jude Kline	Department of Economics, Australian National University
Dr Heli Koski	ETLA, The Research Institute of the Finnish Economy, Lönnrotinkatu 4 B, 00120 Helsinki, Finland
Dr Darcy McCormack	School of Business, La Trobe University, Bundoora
Dr Ross H. McLeod	Research School of Pacific Studies, Australian National University

Dr Rohan Pitchford	Centre for Economic Policy Research, Australian National University
Professor John Quiggin	Department of Economics, James Cook University
Dr Mark Rogers	Melbourne Institute of Applied Economic and Social Research, University of Melbourne
Dr Dan Sasaki	Department of Economics, University of Melbourne; and Department of Economics, University of Exeter, Exeter Devon EX4 4PU, England
Mr Joseph Sunderland	Productivity Commission, Melbourne
Dr Steffen Ziss	Department of Economics, University of Sydney