AGENDA
A Journal of Policy Analysis and Reform
Volume 19, Number 1, 2012

Contents

ANALYSIS

Australia's 2009 ATM Reforms: Transparency for Transparency's Sake

Hugh Green[1]

Abstract

This paper reviews the effectiveness of the reforms to the Australian ATM of early 2009. Data indicate that consumers have acted on the more transparent display of fees by shifting their transactions towards fee-free ATMs provided by their own financial institution, thereby reducing the fees they pay, at the cost of added travel costs. But although consumer surplus has increased, the overall result has been a less-efficient use of the ATM network: banks have lost the whole fall in fee revenue, while consumers have gained only the fee saving, less the extra travel costs. Further, the promise of lower fees did not eventuate, and there was no significant increase in availability of ATMs. Nonetheless, the reforms have been hailed as a success because of the increased pricing transparency in the market. This paper questions the worth of increased transparency when prices are inefficient and welfare loss is the consequence.

Introduction

In March 2009 the Reserve Bank of Australia (RBA) implemented reforms aiming 'to improve competition and efficiency in the ATM system'.[2] The reforms consisted of two main changes. First, interchange fees between institutions were replaced with a direct charge on ATM customers by the machine's owner. Second, whenever an ATM transaction would result in a fee being charged, this fee was to be displayed on the screen and customers asked whether they wished to proceed with the transaction.

Prior to the reforms, customers of an ATM that did not belong to their own bank's network would typically pay a 'foreign fee' for the use of the ATM's services. This fee was charged to the customer by their bank which would in turn pay an interchange fee to the ATM owner. The foreign fee was usually around double that of the interchange fee, with all major banks charging a $2 foreign

1 The Australian National University; hugh.green@anu.edu.au
2 RBA 2010.

fee and paying an interchange fee of about $1.[3] The RBA was also concerned that interchange fees, which were set bilaterally, could lead to collusive behaviour and generally reduce the flexibility of ATM pricing.

Several studies had estimated the marginal cost of providing ATMs and concluded that this cost was substantially less than the foreign fee,[4] implying that the high fees being charged were leading to an inefficient use of the existing network. That is, the level of foreign fees at the time were encouraging an inefficiently high use of own ATMs over foreign machines.

By removing interchange fees and allowing ATM owners to charge customers directly, it was thought that this would stimulate competition among competing machines and reduce fees towards cost. This would be aided by the mandatory displaying of fees to customers who were not previously aware of the magnitude of the foreign fees they were being charged.

At the time it was understood that there were two conflicting forces that would govern behavioural responses to the reforms. The first was the (anticipated) reduction in fees that would drive down the cost of foreign ATM use, and therefore reduce the proportion of transactions carried out at own ATMs. The second was the transparency aspect of the reform which may have driven greater use of own ATMs by consumers who were initially unaware of the foreign fee. This latter effect would be smaller the more successful the reforms were at promoting competition between ATMs. Although the relative magnitude of these competing effects was unknown at the time, it was thought that Bertrand-style competition would emerge between ATMs, which essentially provide a homogenous service from the point of view of consumers, and that prices would therefore fall towards cost.

Now, three years on from the reforms, we are in a position to assess whether they have achieved their goals. For its part, the RBA claims that the reforms have been largely successful, saying, 'the Bank's assessment is that the key objectives of the reforms are being met. In particular, cardholders' reaction to the increased transparency of fees has resulted in a significant reduction in the fees paid in aggregate on ATM transactions.' The RBA estimates that consumers paid $120 million less in foreign fees in the year following the reforms.

This saving to consumers has been the result of reducing the number of transactions on foreign machines, not from a fall in fees themselves. Customers typically face the same fee now as they did prior to the reform, with most

3 Some smaller banks charged fees as low as $1.50, while others charged up to $2.10. In addition, some smaller deposit-taking institutions chose to absorb the interchange fee, charging no foreign fee, in order to attract customers to their other banking services.

4 In 2007, the RBA estimated this cost to be $0.74, while Schwartz *et al*. 2007 estimated the resource cost to be $0.75.

transactions attracting an unchanged fee of $2 when using a foreign machine. In fact, average fees have increased slightly since the reform, indicating that not all consumers may have gained.

This paper argues that the reforms have, in fact, not achieved their goals and the result has been a reduction in net welfare. While consumer surplus is likely to have increased as a result of the increased transparency, the end result has been an increase in the net costs of accessing cash as consumers travel further afield to reduce their ATM fees. The reduction in fees paid merely represents a transfer from ATM owners to consumers, which nets out in a welfare analysis.

Nonetheless, the RBA and some other commentators[5] have championed the reforms on the basis of increased transparency alone. However, in the absence of significant fee reductions, the effect of increased transparency appears to have been to awaken consumers to the fact that the fees are inefficiently high and lead them to act accordingly. Transparency is not an objective in its own right. Where the underlying price signals are efficient, increased transparency allows these signals to work effectively and improve the allocation of resources, but where price signals are not efficient there is no reason to believe that increased transparency would lead to a more efficient outcome.

What would efficiency look like?

ATM pricing is complex and, amongst other things, may be set strategically by banks to increase their market share of deposits. On the demand side, consumer desire for cash depends on a number of factors, including the relative attractiveness of alternative payment instruments such as debit or credit cards. Nonetheless, for a given level of demand for cash, we can ask what the most efficient way of granting access to it through the ATM network would be.

Efficiency is defined in this paper as the sum of consumer and producer surplus. Consumer surplus has three components: the benefit obtained from making the cash withdrawal, the fee paid, and the cost of travelling to the ATM. Producer surplus is the difference between the fee revenue and the cost of providing the ATM. Clearly the fee nets out in this analysis;[6] however, it is important in determining total travel costs through the price signal it sends to consumers. Assuming that consumers receive the same benefit regardless of the machine used, the focus of efficiency then becomes the transport cost and marginal production cost terms. There are two broad aspects to determining the

5 See, for example, King 2011.

6 Naturally, the regulator may attach a higher weight to consumer surplus than to producer surplus; however, this analysis abstracts from these equity considerations.

efficient use of the ATM network. First, given the number of ATMs available to consumers, what would be the most efficient choice of machine for processing a transaction? Second, given the cost of installing and running ATMs, what is the optimal number and distribution of machines to be supplied to the market?

With regard to the first question, it would be efficient for a consumer demanding cash to use the ATM that involves the lowest travel cost, so long as the difference in processing costs between the closest ATM and a 'domestic' ATM is less than the difference in travel cost between that ATM and the closest domestic ATM.[7] In reality, the switching costs in processing a foreign transaction are a few cents per transaction,[8] meaning that it will be efficient for consumers to minimise travel costs in all cases except for when the difference in distance between the foreign and domestic ATM is a matter of metres.

There will, of course, be some optimal level of travel costs, which will depend on the demand for cash and size of the ATM network. But given this demand, it cannot be optimal for consumers to make their withdrawals at less-convenient machines. Essentially, this implies that optimal pricing would not distinguish between own and foreign customers. As we see below, this distinction is made for strategic reasons and does not have a basis in cost differences.

Prior to the reforms, around 53 per cent of ATM transactions were conducted at customers' own-bank ATMs. Clearly, there was a strong preference for own-bank ATM transactions to avoid paying a foreign fee. The reforms intended to improve efficiency by promoting competition between ATM owners, reducing the fee charged to customers on foreign transactions and therefore reducing the incentive to use domestic ATMs over more-convenient foreign machines.

The optimal distribution of machines is perhaps more complex. Loosely, it will be optimal to supply a new ATM in an area so long as the total incremental convenience to consumers outweighs the cost of installing and maintaining a machine. A secondary objective of the reforms was to ensure that ATM owners continued to have sufficient incentives to continue deploying machines. With the average marginal cost of providing an ATM being around \$0.75 per transaction, and expected to rise at a rate faster than inflation, there was concern that a \$1 interchange fee may not provide sufficient revenues. A move to direct charging would ensure that ATM deployment continued to be profitable. Direct charging would also facilitate provision in areas where demand was relatively low (such as areas of low population density or in specialty venues such as stadiums or

7 Clearly, if the closest ATM to the consumer is a domestic ATM it will be efficient for them to use that machine.

8 RBA 2000 and Schwartz *et al*. 2007.

some clubs). In such places, the per-transaction cost of providing an ATM may exceed the interchange fee that was received. Indeed since the reforms, direct charges as high as $4 have been observed in some specialist venues.[9]

The paper now briefly considers the relevant literature on ATM fees and usage before turning to the post-reform evidence to assess whether the goals of competition and efficiency have been achieved.

The academic literature

There is a small but informative academic literature on ATM pricing and provision. The literature typically considers consumers who desire both deposit-taking services and access to ATMs. By choosing a particular bank, consumers in effect choose which ATMs are domestic, with the remainder of the network becoming foreign machines. Banks can use their ATM network size and level of foreign fees to encourage consumers to deposit with them. ATM pricing is therefore based on strategic motivations and need not be cost-based. Two key papers, Massoud and Bernhardt (2002) and Donze and Dubec (2009), show that banks can use ATM deployment and pricing to boost their market share of deposits.

In the direct-charging model of Massoud and Bernhardt, equilibrium results when banks charge their own customers zero ATM fees, instead recovering costs through high account fees, while charging foreign customers a fee higher than that which would maximise ATM revenue. This is because a high fee increases the cost of banking with competing institutions and therefore increases the ATM-owning bank's share of the deposit market. Indeed, Massoud and Bernhardt show that the strategic motivation behind ATM pricing may be sufficiently large that it is optimal for banks to make a loss on ATM provision.[10]

Donze and Dubec (2009) consider a similar model and compare pricing and deployment strategies under three different systems: the first, where ATM transactions are provided free of charge; the second, where an interchange fee is paid between the cardholder's bank and the ATM owner; and the third, where a direct-charge system is in place. They show that ATM provision is highest with direct charging, again as a result of the strategic effect on deposit shares. They also replicate the result that banks may make losses on ATM transactions under

9 RBA 2010.

10 This result is not just an academic curiosity. The 2006 ATM Deployer Study by Boston-based Dove Consulting found that in the US the average on-premises ATM made revenues of $1104 per month and incurred costs of $1444. For off-premises machines these figures were $1013 and $1450, respectively. Clearly motivations other than profit are behind a bank's provision of ATMs.

a direct-charging system. Foreign ATM fees are highest under direct charging but consumer surplus may be higher than under an interchange system if they place a large enough value on the convenience of the additional ATMs.

Importantly, these papers reveal that removing interchange fees in favour of direct charging should not be expected to result in Bertrand-style competition between ATM owners. The strategic interactions between ATM services and the deposit market imply that high fees on foreign transactions were likely to remain after the reforms. As we see below, this has been borne out in the Australian experience.

Evidence from the reforms

By definition, displaying fees on foreign ATM transactions will increase transparency. The question of interest to this paper is whether the move to direct charging, in conjunction with this transparency, has increased the net welfare of Australians.

There are a few key indicators as to the success of the reforms. First, the competition objective aimed at reducing fees on foreign ATM transactions. Second, a more efficient use of the ATM network would show a reduction in domestic ATM transactions. Finally, assuming that the market was not saturated, there would be an increase in ATM availability and use.

Turning first to foreign fees, there has been very little impact on the fees charged to customers of foreign ATMs. As of May 2010, just over a year after the reforms were implemented, 88 per cent of ATMs charged a foreign fee of between $1.50 and $2.[11] Of the remainder, 5.2 per cent charged a fee greater than $2, while 6.5 per cent charged a fee of between $1 and $1.50. There is some evidence that foreign fees have, in fact, risen further since then. By December 2010, the proportion of ATMs charging a fee between $1.50 and $2 had fallen by 11 per cent, to 77 per cent, while the proportion charging a fee greater than $2 had risen by 11.5 per cent, to 16.7 per cent. As of December 2010, overall, the average foreign fee increased slightly following reform, to $2.04. Clearly, the extent of fee competition envisioned has not eventuated.

Prior to the reforms, some institutions with smaller ATM networks did not charge their depositors foreign fees, choosing instead to absorb the interchange fee in order to remain competitive in the deposit market. There is evidence that this practice is continuing in some instances. A year after the reforms, two

11 RBA 2010.

institutions (Bankwest[12] and ING) had a policy of rebating direct charges their members faced on foreign ATM transactions. In addition, smaller institutions have the option of forming fee-free sub-networks among their own ATMs or reaching agreements with larger ATM providers to reduce the fee for their customers. There is some evidence that this is occurring, increasing the availability of zero-foreign-fee ATMs for customers of smaller institutions (RBA 2010). In contrast, some providers who initially chose to charge zero foreign fees prior to the reforms currently only offer to rebate direct fees from a single institution. Nonetheless, it is possible that while there has been no significant competition at the fee level, competition may have been realised at the institution level by increasing the number of fee-free ATMs available.

Of course, the ultimate test of whether competition has worked is the impact on the pattern of ATM transactions between domestic and foreign machines. As previously stated, a more efficient use would see a reduction in the domestic share of ATM transactions as consumers use more-convenient ATMs. Figure 1 shows the impact of the reforms on domestic ATM use.

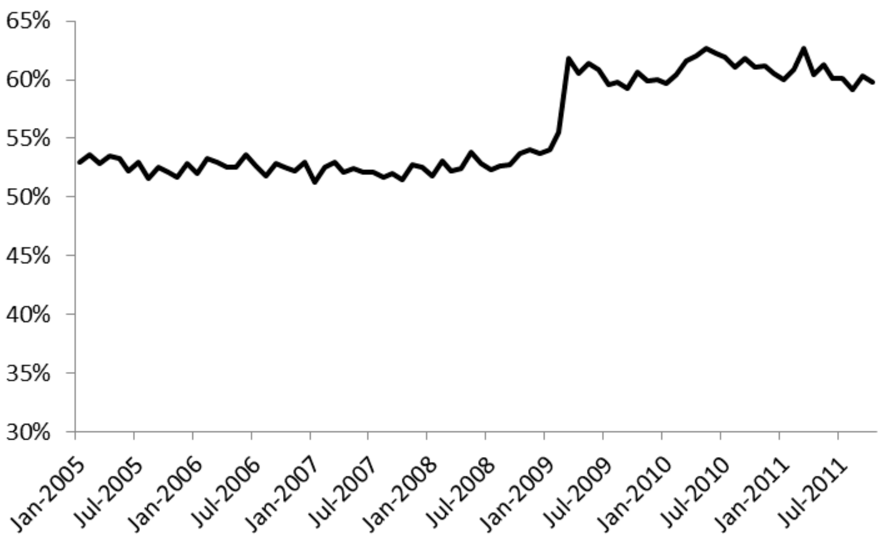

Figure 1: Share of domestic ATM transactions

Source: RBA.

Rather than reducing the proportion of transactions carried out at domestic ATMs, the move to direct charging and displaying fees at the time of the transaction has actually caused a shift towards the use of own ATMs. With the reduction in foreign fees not being realised, it is no surprise that the transparency effect

12 Bankwest offered accounts which charged zero foreign fees prior to the reforms.

has dominated and an increased share of domestic transactions has resulted. Since the reforms were implemented in March 2009, the share of domestic ATM transactions has risen from around 53 per cent to 61 per cent. The RBA calculates that this has resulted in savings to consumers of around $120 million. Clearly, armed with better information about prices, some consumers have found it in their interests to travel to domestic ATMs to avoid paying the foreign fee. The reduction in fees paid also represents a reduction in revenue received by ATM owners (and therefore their shareholders), with the net result being an increase in total costs and reduced efficiency.

To make these ideas more concrete, consider an individual who was initially unaware of the foreign fee they were being charged. They use an ATM that is most convenient to them and unknowingly pay a fee of $2. Now, since the reforms, they are aware of the foreign fee and incur a travel cost of $1 to visit the closest own ATM (of course, their own ATM may have a larger travel cost associated with it, in which case we would observe no change in behaviour). They are now net $1 better off following the reforms, while the ATM owner has lost $2 in revenue.

The RBA reports that consumers have saved $2 but does not mention travel costs (thus overstating the extent to which they are better off). I claim not only this, but that the fee is simply a transfer and that from an efficiency perspective we are *only* interested in the travel costs. Of course, consumers as a whole are better off, as with no change in fee they are simply benefitting from the additional information and changing their ATM transactions only if it is in their interest. However, this transparency clearly comes at a cost.

Other costly behavioural changes have also been observed. For example, the average value of foreign withdrawals has increased since the reforms, implying that consumers are withdrawing higher amounts of cash than originally desired in order to reduce the per-dollar cost of foreign fees. There has also been a slight shift towards other payment methods to avoid the need to withdraw cash at foreign ATMs.

The higher travel costs resulting from an increase in domestic ATM usage may be offset somewhat if the reforms have led to a significant increase in ATM provision. This appears not to be the case on aggregate. Figure 2 shows the total number of ATM transactions prior to and after the reforms.

There has been remarkably little change in total ATM usage over this period. If provision had increased in areas where ATMs were needed, we would expect the convenience to result in an increase in the number of ATM transactions. This appears not to be the result.

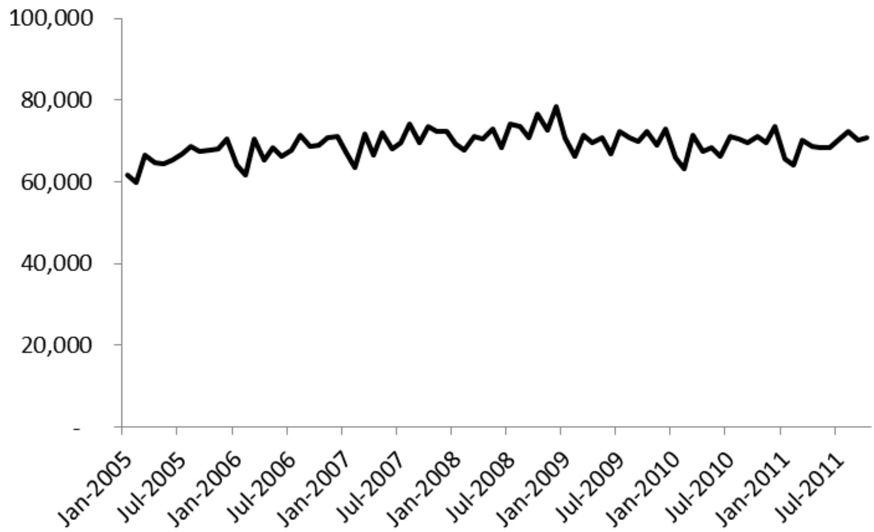

Figure 2: Total number of ATM transactions

Source: RBA.

Writing a year after the reforms, the RBA claimed that ‘the year-to-year volatility in provision of machines makes it difficult to draw strong conclusions about post-reform ATM deployment relative to earlier periods. Nonetheless, the evidence suggests that the reforms have been positive for the availability of ATMs.’[13] Figure 3 shows the total number of ATMs over the last decade.

Prior to the reforms, provision had been increasing steadily and this continued over the years straddling and immediately after the reforms, with network growth of 5.3 per cent in the year to March 2009, 3.1 per cent in the year to March 2010, and a 4.5 per cent growth in the year to March 2011. A formal test for a structural break[14] in the series after the reforms showed no evidence of an increase in provision above the pre-reform trend.

13 RBA 2010.

14 The quarterly series of ATMs (number of ATMs) was modelled as an AR(1) process with a trend. A dummy variable was created which took a value of 0 prior to the reform and 1 thereafter. The estimated coefficient on this variable was -304, with a standard error of 332, indicating that it is not significant at any conventional level.

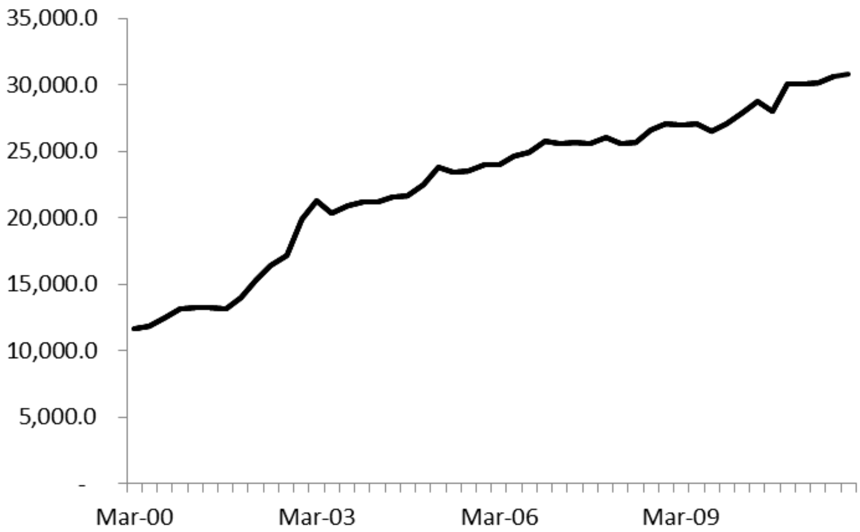

Figure 3: Total number of ATMs

Source: Australian Payments Clearing Association.

Overall it appears that the reforms have had negative effects on efficiency. The direct result has been a shift towards greater usage of domestic ATMs, increasing the total travel costs incurred by consumers. While transparency has been touted as a benefit of the reforms, it is precisely this greater transparency which has driven these inefficiencies.

Of course, the clearer price signals may have led to a net improvement in welfare if the competition aspect of the reforms had been more successful. It is interesting that, and not altogether obvious why, competition on fees did not result. The academic literature suggests strategic reasons for why banks may not respond to the reforms by reducing fees. However, around half of Australia's ATMs are owned by independent ATM deployers for which these strategic motivations are not present (the largest being Customers Limited, Cashcard and RediATM, which together account for just under one-half of the total number of ATMs). It may be that each ATM has sufficient local monopoly power that Bertrand-style competition may not result.

It may also be that the form in which fees are displayed at ATMs may not be conducive to active competition. Because fees are displayed only once the transaction is under way, consumers may be unwilling to cancel a transaction and shop around for a cheaper machine. This has led to suggestions that machines could have their fee published on the exterior of the machine so that consumers are informed of the fee prior to commencing the transaction.

However, the models in the literature assume that consumers are fully informed of foreign fees, so if strategic motivations are at play it is unlikely that clearer price signals are the answer and may even reinforce the existing situation.

Conclusion

ATMs provide great convenience to consumers who would otherwise need to visit bank branches to access cash. Given the ATM network in place, the most convenient usage of ATMs would involve using those machines which minimise a consumer's travel costs. Since there is very little cost difference in processing a transaction at a domestic and foreign ATM, the most convenient usage pattern is therefore also the most efficient. Pricing which encourages consumers to use their own ATMs over foreign ATMs will therefore distort efficient decisions and reduce welfare.

It is apparent that prior to the 2009 reforms a significant number of consumers were not aware of the prices they were charged on foreign ATM transactions, or at least acted in a way that indicated that fees were not being taken into account in their purchasing decisions. This had the interesting result that many consumers acted as though they were part of a competitive market, even though prices were significantly higher than the competitive level.

Increased transparency was easily achieved by displaying fees at the time of a transaction and, as is the goal of any transparency regulation, consumers responded to the clearer pricing signals. The problem was that with the competition objective not achieving its goals, the pricing signals remained inefficient. The end result is likely to have been a fall in welfare.

The fact that this result was not anticipated is to some extent not surprising. There was little information at the time the reforms were implemented to inform decision-makers about the likely reactions of consumers or ATM owners. While the academic literature does highlight the strategic motivations behind ATM pricing, this literature was not well established at the time. Indeed, over the last decade Australia has been a testing-ground for several issues in payments-system regulation, with these reforms often preceding the academic literature.[15]

However, despite the unanticipated outcomes and resulting welfare losses, the RBA and others have claimed that the reforms have been successful on the basis of achieving transparency alone. But surely the goal of transparency should be to allow efficient price signals to be communicated to consumers. When

15 Particularly developments in the theory of two-sided markets and the efficiency implications of interchange fee regulation.

these signals are inefficient, it is unclear what benefit communicating them to consumers would have. What value should we place on transparency itself when the end result is a net destruction of welfare?

References

Donze, J. and Dubec, I. 2009, 'Paying for ATM usage: Good for consumers, bad for banks?' *Journal of Industrial Economics* 59: 583–612.

King, Stephen P. 2011, 'Competition Regulation in Australia: A Report Card', the 2011 Colin Clark Lecture.

Massoud, N. and Bernhardt, D. 2002, 'Rip-Off ATM Surcharges', *Rand Journal of Economics* 33(1): 96–115.

RBA and ACCC 2000, *Debit and Credit Card Schemes in Australia: A Study of Interchange Fees and Access*.

RBA 2009, *The Introduction of Direct Charging for ATMs*.

——— 2010, *Reform of the ATM System — One Year On*.

Schwartz, C., Fabo, J., Bailey, O. and Carter, L. 2007, 'Payment Costs in Australia', prepared for the *Payments Systems Review Conference* at the Reserve Bank of Australia, November 2007.

Citations as a Measure of the Research Outputs of New Zealand's Economics Departments: The Problem of 'Long and Variable Lags'

John Tressler and David L. Anderson[1]

Abstract

The paper explores the merits of utilising citation counts to measure research output in economics in the context of a nationwide research evaluation scheme: the New Zealand Performance-Based Research Fund (PBRF). Citations were collected for all refereed papers produced by New Zealand's academic economists over the period 2000 to 2008, and used to estimate the time-lags in between publication and the flow of citations; to demonstrate the impact of alternative definitions of 'economics-relevant' journals on citation counts; and to assess the impact of citation measures on departmental and individual performance. We conclude that under certain scenarios around 60 per cent of papers received no citations over the period. Our findings suggest that the time-lags between publication and citation make it difficult to rely on citation counts to produce a meaningful measure of output in a PBRF-like research-evaluation framework, especially one based on the assessment of individual academics.

Introduction

The external evaluation of the research of university faculty has become an important part of academic life in many countries. The processes used to assess research differ significantly in the way in which evaluations are undertaken.[2] Peer-review panels or expert panels are used by a number of countries, but they are costly, both for the assessors and those being assessed. Advances in information technology and the quality of research databases have made citation analysis a more attractive alternative or complement to other assessment

1 University of Waikato; tressler@waikato.ac.nz (corresponding author); Queen's University.

2 For an in-depth overview of processes used in the evaluation of research generally as part performance-based research funding for universities, see OECD (2010).

processes. The primary purpose of this paper is to explore the merits of utilising citations to measure research output in economics in the context of a nationwide research-evaluation scheme.

To provide a context for the evaluation of citation analysis we discuss its potential use in relation to the New Zealand government's Performance-Based Research Fund (PBRF).[3] To assess the quality of research under this program, individual academic staff must submit research portfolios providing a review of their research output, of their contribution to the research environment and of peer esteem over the previous six years. These portfolios are assessed and scored by peer-review panels formed for each subject area. Individual scores are aggregated to provide assessments of subject areas in an institution and the tertiary institution overall. Individual scores are reported to the academics concerned, and subject and institution results are made public. PBRF evaluations of the quality of research were undertaken in 2003 and 2006, and will be undertaken in 2012.[4] The evaluations of research quality determine 60 per cent of the PBRF funding allocations to tertiary institutions, with the remaining 40 per cent based on research degree completions and external research income. In 2010 the results were used to distribute NZ$268 million of research funding; this is equal to 18 per cent of government funding to universities and roughly 9 per cent of total system-wide revenue. Furthermore, and equally important, the PBRF results are aggressively used by the winners in their formal and informal promotional material.

Unlike the Australian ERA process, PBRF peer-review assessments are not supplemented by citation analysis. Over time it is likely that counting citations will play a part. It is our view that after the upcoming 2012 PBRF round, pressure may mount for the government to consider a shift, at least in part, to a metric-based system. If this were to occur, citation counting is likely to be at the heart of any such scheme, given its widespread acceptance as a reliable measure of performance in the physical and biological sciences. This view is based, in part, on the tendency of New Zealand tertiary education policy to follow that of the UK and, to a slightly lesser extent, Australia. In both of these countries, citation counting is now partially incorporated into their nationwide research-evaluation schemes.[5]

In this paper we will explore the merits of the concerns of many in the social science community over the use of citations as a measure of research output.

3 A discussion of the key elements of the scheme can be found in Goldfinch (2003) and Hodder and Hodder (2010). Additional information can be found on the official website: www.tec.govt.nz/Funding/Fund-finder/Performance-Based-Research-Fund-PBRF-/Resources/

4 The 2006 round was a partial round in which academics could carry forward their evaluation from 2003 or submit a new portfolio.

5 For details, see Research Excellence Framework (REF) (www.hefce.ac.uk/reserch/ref/) and Excellence in Research for Australia (ERA) (www.arc.gov.au/era).

In order to restrict the task to manageable proportions, we will focus on the discipline of economics. We have collected citation counts to all refereed papers produced by New Zealand's academic economists over the period 2000 to 2008 using the databases of the ISI/Web of Science (henceforth, the ISI) and, in a more limited fashion, Google Scholar. The data collected allows us to assess, among other things, the time-lags in economics between publication of an article and the flow of citations; the impact of alternative definitions of 'economics- relevant' journals; and a comparison of departmental and individual performance using both direct-citation measures and alternative schemes based on journal-specific weights. With respect to the latter, we utilise journal weights based indirectly on citation counts and on reputational surveys.

In general, our findings suggest that the conventional assessment period of six years may be acceptable from a departmental perspective due to averaging effects, but that this is too short a time period for individual assessment. Our work suggests that the output measures generated by citation counts are not highly correlated with traditional output measures based on journal impact factors.

Some doubts explored

Although citation counts have long been used, and generally accepted, in the physical and biological sciences (henceforth denoted as 'the sciences') as a proxy measure of research output, the applicability of this metric for estimating social-science research output is problematic.[6] Concerns have been expressed over purported differences in citation practices across the above-mentioned disciplines. This argument has at least two dimensions: major differences between the sciences and social sciences in the time-lag between the publication of an article and the commencement of a meaningful flow of citations to the article; and differences in the publication frequency and citing habits between the science and social-science disciplines that work to the disadvantage of the latter.

There is also a data-collection issue at play here. Historically, ISI focused on the sciences; it is only in the past few years that this organisation has started to aggressively expand the range of social-science journals for which it collects citations. This means that citation-based performance measures in the social sciences only capture a portion of all citations generated by researchers, especially those publishing in languages other than English and, of greater relevance to the New Zealand scene, to those publishing in regional journals

6 For example, see Centre for Science and Technology Studies (2007).

on regional issues. The latter is a major issue in small countries: governmental funding agencies generally wish to see a substantial degree of research performed on matters deemed to be of relevance to the nation-state. In the social sciences, this often results in articles that have greater interest to national or regional journals than to international journals. Therefore, if only the latter journals are in the database, researchers performing work with a regional focus will appear to be low or even non-publishers. Furthermore, even if the work is published in a journal included in the ISI list of recognised journals, papers discussing local issues are less likely to be cited than those addressing similar issues in a large-country setting.

Although the above issues are important, the primary problem with respect to the social sciences is likely to be the lengthy lag between the typical article's publication date and the commencement of a meaningful flow of citations. In order to demonstrate the importance of this matter, and to illustrate how it may arise, let us refer to the upcoming 2012 PBRF round. For all academic staff employed at the 2012 census date, the PBRF scheme will attempt to assess all research generated by these individuals over the period 1 January 2006 to 31 December 2011. If one were to introduce a measure designed to capture the number of citations generated by papers published over this six-year time period, it is quite apparent that the time-lag issue will be of great importance. If the lags are, say, on average two to three years, it means that much of the research performed over the six-year assessment period will be ignored by the PBRF scoring system; it also means that work published in the early years of the cycle will be deemed to be of greater value than work published at the end of the evaluation period (everything else being equal).

The lag issue creates a special problem for newly hired and newly minted PhDs. In addition to the time required to develop a research program, obtain necessary funding, prepare papers for submission to journals, and go through the review and publication process, we now have to add additional time to reflect the period between publication and a meaningful flow of citations. Even without the citation-lag issue, the PBRF scheme has been modified to treat new entrants (with limited prior experience) differently. In practice, it is widely recognised that institutions are shifting their hiring practices away from the inexperienced to those with a 'good' next-round PBRF-relevant publication record.[7]

All of the above is based on conjecture. We have not been able to find an empirical study of the citation practices in economics (or social science) that addresses the issues raised above. In this study we will attempt to shed some light on these matters, especially the time-lag issue. We employ data from New Zealand-

7 This statement is largely based on anecdotal evidence, but supporting evidence can be found in Cinlar and Dowse (2008).

based economists to generate estimates of the time pattern of citations based on alternative definitions of 'economics-relevant' papers, and, to a limited degree, on alternative citation-capturing schemes. We will also compare the output performance of economics departments and individual economists using direct citation counts and widely employed alternative measures. However, we will not attempt to compare citation practices in economics to those in the biological and physical sciences.

At this point we should mention that the economics literature on research-output measurement is dominated by the journal-based weighting approach (Macri and Sinha 2006).[8] The most common method for generating the desired journal weights is to count citations to each journal in the dataset, over a given time period, and then to divide the total by the number of articles contained in each journal over the same time period. This procedure yields an estimate that is commonly denoted as a journal's impact factor, and frequently assumed to be a measure of a journal's quality. This approach has been modified by a number of economists, through the use of iterative adjustment processes, to yield aggressive journal-weighting schemes that are widely used in the economics literature (Anderson and Tressler 2010). Alternatively, journal-based weighting schemes sometimes rely on 'expert opinion' such as that initially employed in the Australian government's Excellence in Research for Australia (ERA) scheme.[9] Regardless of the underlying approach, the resulting journal weights are applied to all articles in a given journal, and the resulting values are aggregated to arrive at departmental and, sometimes, individual output estimates. In the majority of cases, further adjustments are made to account for the length of the article in terms of the *American Economic Review* (*AER)* page-size equivalents, and to reflect each author's share in multiple-authored papers.

The primary reason for favouring the journal-based approach is a variant of the time-lag argument presented above. Given the desire to generate an estimate of the probable long-term impact of an individual's relatively recent output (say, one to six years), it is necessary to resort to proxy measures. It is generally accepted by economists that the best proxy available is the impact or quality of the journal in which the paper is published. Rephrased, if citations are viewed as the principal indicator of research impact or quality, then the best indicator of the expected number of citations to a paper over the long run is best approximated by the relative importance (however measured) of the host journal.

---

8 Following convention, we have restricted research output to cover only refereed articles in journals listed in *EconLit*. Rephrased, academic work disseminated in books, conference papers, reports and non-refereed publications are ignored in this study.

9 For details, see ERA's website: www.arc.gov.au/era

However, the merits of journal-based weighting schemes have recently been called into question by Starbuck (2005), Oswald (2007), Wall (2010) and Chang, McAleer and Oxley (2011). For the purposes of this paper, their findings can be summarised as follows: good papers (lots of citations) can be found in lowly ranked journals (relatively few citations), and poor papers (very few citations) can be found in highly ranked journals (many citations). Indeed, Chang, McAleer and Oxley (2011) found that over a 25-year period, approximately 26 per cent of the papers published in the top 40 economics journals (based on ISI's two-year impact factors) did not receive a single citation. Furthermore, they found that 20 per cent of the selected journals had 'zero citation' rates in excess of 40 per cent. All of this work suggests that journal-based impact factors may not yield a good estimate of an individual paper's long-term impact. If this is correct, the search for a better proxy inevitably leads one to explore the use of a direct-citation measure — the counting of citations to a given paper, over a given time period.

Data

We assembled three basic datasets for this study.

- Dataset1: citations — be they in ISI or Google Scholar databases — collected over a 10-year period, for all papers published in 2000 and 2001 by New Zealand's academic economists on staff as at either or both 15 April 2007 or 15 April 2009.[10] More formally, we counted all citations to these papers using both databases; and the citations were collected in early January 2011.[11]

The collection of citations using Google Scholar is relatively straightforward. On the other hand, generating citations from the ISI database requires a number of adjustments and exclusions. First, we restricted our search to citations from ISI-listed journals (that is, we excluded citations from conference papers). Second, one faces an age-old problem in economics: which journals are economics journals? We handled this matter in two ways: we created a 'broad' definition of economics by assuming that all articles published by New Zealand's academic economists in both *EconLit* and ISI-listed journals are relevant to the discipline.[12] We refer to data based on this definition as *ISIB*. The alternative approach was labelled as a 'narrow' definition of economics and is based on the

10 These staff census dates were chosen for pragmatic reasons: we had previously collected publication records for all academic staff employed on these dates. More specifically, we collected data on all permanent staff with the rank of Lecturer, Senior Lecturer, Associate Professor and Professor. We should also note that we used both staff lists to maximise the size of the sample.

11 For papers published in 2000, we collected cites generated over the period 1 January 2000 to 31 December 2009. For 2001 papers, the time period shifted forward by one year.

12 This decision rule was used by Coupe (2003) in deriving his 'Impact' measure. At the time of his study (2000), some 800 journals were listed in *EconLit*; of these, 273 were listed in ISI/JCR. However, only approximately 170 of these journals were listed as economics journals by Journal Citation Reports (JCR).

restrictive practice of recognising only articles published in journals listed as 'economics' in the Journal Citation Reports (JCR).[13] We refer to data based on this definition as *ISIN*. In practice, under the broad definition of economics we include a number of journals in the areas of urban studies and finance that are excluded from the narrow definition list. The third restriction used in our collection exercise was to eliminate self-cites by authors.[14]

- Dataset2: ISI citations attributable to papers published between 2000 and 2008 by the same group of New Zealand economists as noted above. In this case, we counted citations up to the end of 2010. This means that we have a time series of citations ranging from 11 years, for papers published in 2000, to three years for those published in 2008. It should also be noted that in order to restrict the analysis to manageable proportions, we have limited our citation-collection exercise in Dataset2 (and in Dataset3 to follow) to our broad definition of economics (ISIB) rather than the narrow version (ISIN). Our rationale for selecting the broad over the narrow definition of economics for ISI counting purposes is based on our understanding of the current PBRF scheme — that work in boundary areas (such as finance and urban studies) is generally recognised as 'economics-relevant'.[15] Furthermore, our preference for ISIB over Google Scholar is based on the widespread view that ISI (narrowly or broadly defined) is the 'gold standard database' (Chang, McAleer and Oxley 2011).[16]
- Dataset3: constructed to allow us to compare rankings of departments and individual economists using various citation measures with those generated by more traditional measures. For all 135 economists employed by New Zealand's eight university economics departments as at 15 April 2009, we constructed a record of all articles published by these researchers in *EconLit*-recognised journals, over the period 2003–08.[17] In order to restrict the scope of the study, we have arbitrarily selected only two journal-based weighting

13 Most of the well-known journal-based ranking schemes in economics are based on this restrictive definition of economics-relevant journals albeit, in a few instances, with a couple of additions from the finance area. For example, see Liebowitz and Palmer (1984), Laband and Piette (1994), Kalaitzidakis, Mamuneas and Stengos (2003 and 2010), and for their 'economics' journal weights, Kodrzycki and Yu (2006).

14 Self-cites are eliminated to prevent game-playing tactics. Although this is not considered to be a problem in economics, or the social sciences, at this time, it is widely recognised as a problem in the biological and physical sciences. Although we are getting ahead of ourselves, we should note that of the total citations to papers published by New Zealand's academic economists over the period 2000–2008, approximately 15 per cent (441 of 2857) were 'self-cites' (for the citing period ending 31 December 2010). At this time we wish to stress that throughout the remainder of this paper, unless otherwise noted, all references to 'citations' will always be to the 'non-self' variety.

15 For a discussion of these issues, albeit from a Finance perspective, see Cosme and Teixeira (2010).

16 Although Google Scholar is rapidly gaining academic credibility, it has been criticised for lack of transparency in design and scope. For a New Zealand/PBRF-related assessment of Google Scholar, see Smith (2008).

17 Following convention, we allocated shares to individual authors based on the 1/n rule (for example, if a paper has three authors, each is granted a one-third share), and used the size-adjusted page (AER equivalent) as our unit of output (see Macri and Sinha 2006).

schemes for comparison purposes: *KMS2010* to represent an aggressive scheme based indirectly on citation counts;[18] and *ERAB*, the Australian government's journal-weighting scheme based on 'expert opinion' (that is, a perception-based system).[19]

In order to demonstrate the importance of the time-lag issue, we constructed three citation measures.

- *ISIB03–08* is based on a simple count of citations over the period 2003–08 to all papers published over this same time period. This time span corresponds to the timeframe adopted by the PBRF — a six-year period. Therefore, papers published in the first year of the evaluation period are able to generate citations over a six-year period, whereas those published in the last year of the cycle have, at the most, one year to capture citations.
- *ISIB01–06A* is based on papers published over the period 2001–06, with citations collected from 2001 to the end of 2008; therefore, the maximum citation collection period is eight years, and the minimum is two years.
- *ISIB01–06B* is also based on papers published over the 2001–06 period, but the citation-collection period now ends on 31 December 2010. The maximum and minimum period for capturing cites is now 10 and four years, respectively.

Findings

Ten-year citation patterns for articles published in 2000 and 2001

Our first task is to shed light on the time-lag between publication and the generation of a meaningful stream of citations for papers produced by New Zealand's economists. Before doing so, we should note that the nation's 156

---

18 This weighting scheme was developed by Kalaitzidakis, Mamuneas and Stengos (2010); it is an update of their prior work (Kalaitzidakis, Mamuneas and Stengos 2003). This is an aggressive weighting scheme in that the weights given to the 'top' journals are as large as 1000-times that assigned to lower-end journals. For example, the first-place journal, the *AER*, receives a score of 100.0, whereas the 50th- (*Labour Economics*), 100th- (*Journal of Economic Geography*), and 150th- (*Economic Geography*) placed journals receive scores of 3.06, 0.73, and 0.12, respectively. For a more rigorous discussion of the aggressive nature of this scheme, see Henrekson and Waldenstrom (2009) and Anderson and Tressler (2010).

19 We have adopted a broad version of the ERA scheme (hence the reason for denoting it as ERAB). That is, we recognise all journals listed in both the ERA and *EconLit* regardless of the category that they have been arbitrarily placed in. In practice, this means that a number of papers in finance and urban-studies journals receive a non-zero weighting. Recall that under the narrow definition of economics selection process, these papers would have received a zero weighting. It should also be noted that the ERA officially uses a four-point grading scale: A+, A, B and C. We have arbitrarily converted it to a five-point scale: 4, 3, 2, 1, and 0 (the last score for journals not covered by the ERA scheme but included in *EconLit*).

academics employed by its eight university-based economics departments in 2007 and/or 2009, published 167 papers in *EconLit*-listed journals in 2000 and 2001. As shown in Table 1, over a 10-year period, the average paper received 4.1 and 3.0 ISIB and ISIN citations, respectively; the corresponding number for Google is much larger — 15.9. This difference is not surprising given that Google Scholar collects citations from working papers, public reports, conference papers and books, whereas ISI citations are only collected for, and generated by, JCR-listed journals.

Let us now explore the time pattern of citations. In the first year after publication, very few citations were generated for ISIB and ISIN — less than one per cent of the 10-year total. Indeed, by the end of Year 3, the corresponding estimate is only 6.7 and 7.2 per cent, respectively. It is clear that a relatively steady, but growing, stream of citations does not commence until Year 4. Interestingly, the flow does not abate over the remainder of the collection period as Year 10 citations are higher than those in any preceding year.[20] Google exhibits a somewhat similar year-by-year time pattern; however, the flows in Year 1 and Year 2 are larger, but still below the levels generated in Year 3 and onwards. Not surprisingly, we found the ISI and Google Scholar year-by-year citation patterns to be highly correlated: the Pearson Correlation Coefficients for ISIB/Google and ISIN/Google are 0.951 and 0.906, respectively. However, if we explore the relationship on a per-paper basis, the correlation coefficients are much lower: ISIB/Google 0.262 and ISIN/Google 0.483.

Table 1: Non-self citations for papers published in 2000 and 2001

Citation Scheme	Total	Cites/ Paper	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10
ISI, Economics Broadly Defined	686	4.1	4	8	46	65	80	64	96	92	115	117
ISI, Economics Narrowly Defined	497	3.0	3	7	36	54	50	44	60	69	87	88
Google Scholar	2659	15.9	100	129	242	288	330	303	394	362	372	342

Note 1: Based on 167 refereed papers published in EconLit listed journals.

Note 2: The figure reported for Google Scholar does not reflect undated citations of which we found 203.

The above-noted discrepancy between the correlation coefficients generated by our broad and narrow definition of economics is undoubtedly largely attributable to a single paper. Gordon and McCann (2000), published in *Urban Studies,* generated 42.2 per cent of the total ISIB cites captured by all papers published in 2000, and 26.5 per cent of all such cites over the 2000 and 2001 period. For Google, the corresponding figures are 31.4 and 21.1 per cent, respectively. However, *Urban Studies* is not considered to be an 'economics' journal by ISI (it

20 As discussed later in the paper, this result may be attributable to a significant increase in the past few years in the number of journals eligible to generate and receive citations.

is deemed, rather, to be an urban studies publication), and hence generates zero ISIN cites. Although this is an extreme case, the distribution of citations across papers will subsequently be shown to be highly skewed.

As mentioned earlier, it is widely known that many papers in economics fail to receive a single citation over long periods of time (Chang, McAleer and Oxley 2011; Wall 2010; and Oswald 2007). We shall now explore this issue in the New Zealand context. Based on Dataset1, over a 10-year collection period only 40.1 per cent of papers received one or more ISIB citations (for ISIN, the estimate is 37.7 per cent). In contrast, the estimate for Google is almost double — 78.4 per cent. In large part, this discrepancy can be explained by differences in the scope of coverage of the citing and cited journals. Recall that ISIB is based solely on citations to and from JCR-listed journals, whereas Google Scholar sources citations from a much broader range of publications. In 2010, this resulted in only 64.7 per cent of papers in Dataset1 being eligible to receive ISIB citations, as opposed to 100 per cent for Google.[21] Therefore, of ISIB-eligible papers, 62.0 per cent were ultimately cited. (The corresponding numbers for ISIN are 61.1 and 61.7 per cent, respectively).

In the above discussion we have focused on the time pattern of citations per year to all journals in our sample. However, this is only one way of looking at the 'lag' issue. Another way of doing so is to explore the length of time it takes individual papers to receive their first citation. As shown in Table 2, three years after publication, only 16.2 per cent of papers had received one or more ISIB citations; after five years, 32.3 per cent of papers were in this category, and, as discussed earlier, after 10 years the number had increased to 40.1 per cent. On the other hand, the estimates for each time period under Google are dramatically higher: 52.7, 71.9 and 78.4 per cent, respectively.

Table 2: Non-self, non-zero citations per paper, various time periods based on cites to papers published in 2000 and 2001

	Percentage of Papers with Non-Zero Citations		
Citation Scheme	End Year 3	End Year 5	End Year 10
ISI, Economics Broadly Defined	16.2	32.3	40.1
ISI, Economics Narrowly Defined	15.6	31.1	37.7
Google Scholar	52.7	71.9	78.4

21 Economists in New Zealand face the regional-bias problem mentioned earlier in the paper. That is, the nation's only refereed economics journal, the *New Zealand Economics Papers* (*NZEP*), is not included in the ISI/JCR database. For obvious reasons, *NZEP* is the leading publication vehicle for New Zealand economists. If we arbitrarily drop papers in *NZEP* from the dataset, we find that 70.6 per cent of all remaining broadly defined economics papers were published in ISI/JCR-listed journals (the corresponding figure for narrowly defined economics papers is 66.7 per cent).

Despite the evidence of relatively long lags in the citation-generating process, especially for ISIB and ISIN, one can take some comfort from the information displayed in Table 3. For example, for all three measures of output, the correlation coefficients associated with three- and 10-year citation counts (on an individual-paper basis) range from 0.819 to 0.875. As expected, the estimates rise as we increase the collection period: for instance, the correlation coefficients for the five- versus 10-year citation period rise to 0.925 to 0.978 for all three output measures; the corresponding estimates for the seven- versus 10-year citation period range from 0.973 to 0.995. This suggests that if a 10-year collection period is considered to be an ideal time period for generating estimates of citation-based research output, then the use of, say, a five-year collection period could result in acceptable proxy estimates.

Table 3: Correlation coefficients, non-self citations per paper, various time periods (based on cites to papers published in 2000 and 2001)

Cites Per Paper, Various Time Periods	ISI (Broad)	ISI (Narrow)	Google
Year 1-1: Year 1-10	0.523	0.209	0.525
Year 1-2: Year 1-10	0.697	0.691	0.637
Year 1-3: Year 1-10	0.875	0.838	0.819
Year 1-4: Year 1-10	0.900	0.893	0.924
Year 1-5: Year 1-10	0.978	0.925	0.970
Year 1-6: Year 1-10	0.990	0.960	0.987
Year 1-7: Year 1-10	0.994	0.973	0.995
Year 1-8: Year 1-10	0.996	0.981	0.997
Year 1-9: Year 1-10	0.999	0.996	0.999

Citation patterns for all articles published between 2000 and 2008

Let us now move to an analysis based on Dataset2. Recall that the distinguishing feature of this dataset is that we have expanded the publication period from 2000–01 to 2000–08; however, the research group remains the same as in Dataset1. Over this nine-year period, NZ's 156 economists published 871 articles in *EconLit*-listed journals, and by the end of 2010 these publications had received a total of 2470 ISIB citations. The distribution of these by year is shown in Table 4. Note that, with one exception, the citation pattern is similar to that discussed earlier when we explored the 10-year pattern for papers released in 2000 and 2001. Now the collection period ranges from 11 years (for papers published in 2000) to three years (for papers published in 2008). The one exception relates to 2008 publications: it would appear that more citations are generated in Years 2 and 3 than expected. This might be related to the nature of the papers published and their topicality, but it might also be related to the rapid expansion of ISI's journal coverage in economics and, more geneally, the social sciences. This issue will be discussed later in the paper.

Table 4: ISI non-self citations, papers published 2000–2008, broad definition of economics: distribution of citations per year; citations collected up to 31 December 2010

Year Published	# of Papers	Total # of Non- Self Cites	Average # of Non-Self Cites/ Paper	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10	Year 11
2000	73	478	6.55	2	4	22	35	51	40	61	61	75	74	53
2001	94	262	2.79	2	4	24	30	29	24	35	31	40	43	
2002	101	347	3.44	3	11	23	33	43	55	56	72	51		
2003	95	447	4.71	2	19	50	67	81	68	70	90			
2004	91	144	1.58	1	15	25	24	27	20	32				
2005	101	277	2.74	3	20	48	65	57	84					
2006	115	220	1.91	9	17	40	66	88						
2007	94	171	1.82	6	18	63	84							
2008	107	124	1.16	5	43	76								

Table 5: ISI non-self citations, papers published 2000–2008, broad definition of economics for papers ultimately cited, percentage with non-zero citations at various year-ends

Year Published	# of Papers	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10	Year 11
2000	73	6.5	16.1	48.4	67.7	80.6	83.9	90.3	96.8	100.0	100.0	100.0
2001	94	5.6	11.1	33.3	55.6	80.6	86.1	94.4	94.4	97.2	100.0	
2002	101	6.7	20.0	22.2	60.0	80.0	91.1	91.1	93.3	100.0		

Although NZ's economists published 871 papers over the period 2000–08, only 41.1 per cent of them received one or more ISIB cites. However, this figure is misleading in two respects. First, only 540 papers were published in currently listed JCR journals. After making this adjustment, we find that 66.3 per cent of eligible papers received one or more ISIB cites. Second, 50 papers in our sample were published in journals that at time of publication were not covered by ISI (that is, these journals were added at a later date). Therefore, if we restrict the sample to papers eligible for citation counting by ISIB*,* we find that 73.1 per cent of them received one or more citations.

It is interesting to note that of the papers eventually receiving one or more ISIB cites, the vast majority reached that status by the end of Year 5. This can be seen by reference to Table 5. More specifically, for papers with a nine-or-more-year citation-collection period (papers published over the period 2000 to 2002), approximately 80 per cent of papers that were eventually cited had reached that status by the end of Year 5.

Let us now turn to an examination of the distribution of citations across papers. In Table 6 we display the percentage distribution of ISIB cites over various groupings for three different collection periods: 10, eight and five years. As previously noted, approximately 40 per cent of papers receive at least one cite over our catchment period (up to 10 years). However, the number of papers receiving multiple cites drops off rather quickly. For example, across our five-, eight- and 10-year collection periods, only 19.1, 22.6 and 20.4 per cent of papers received five or more cites; the corresponding figures for 10 or more cites are 9.9, 12.7 and 12.0 per cent. It is clear from the data that few papers receive 20 or more cites; even 10 years after publication, only 4.8 per cent of papers published in 2000 and 2001 reached this status.

Citation patterns compared with other measures of research output

We concluded our empirical work by calculating departmental and individual researcher output using various citation measures, and we compared the results with those generated by competing schemes. As noted in the discussion of Dataset3, we constructed three citation measures for academic staff employed as at 15 April 2009: ISIB03–08, ISB01–06A, and ISIB01–06B. It is important to recall that these schemes differ in two ways. First, ISIB03–08 is based on publications over the period 2003–08, our hypothetical PBRF timeframe. By contrast, ISIB01–06A and ISIB01–06B count citations to papers published over the 2001–06 period. Second, each scheme differs with respect to the lag time between the last year of publication and the final year of citation counting. More explicitly, the time-lags are zero, two and four years for ISIB03–08, ISIB01–06A and ISIB01–06B, respectively.

Table 6: ISI non-self citations, broad definition of economics, papers published 2000–2008; cumulative percentage distribution of papers receiving a given number of citations over various time periods

			Percentage of Papers with denoted number of Citations								
Years of Citation Coverage	# of Papers	Total # of Non- Self Cites	Papers with Zero Cites	Papers with > = 1 Cite	Papers with > = 2 Cites	Papers with > = 3 Cites	Papers with > = 4 Cites	Papers with > = 5 Cites	Papers with > = 10 Cites	Papers with > = 15 Cites	Papers with > = 20 Cites
10 Years (Publ. 2000 - 2001)	167	686	59.3	40.7	32.9	28.1	22.8	20.4	12.0	7.8	4.8
8 Years (Publ. 2000- 2003)	363	1480	57.3	42.7	35.8	31.1	25.3	22.6	12.7	7.4	4.4
5 Years (Publ. 2000- 2006)	670	2121	59.3	40.6	32.4	27.3	22.2	19.1	9.9	5.4	3.0

Note: Citations collected from date of publication to 31 December 2010.

For comparison purposes, we derived output estimates for three competing output schemes: KMS2010, ERAB and *EQUAL*. The first two weighting schemes were discussed above, but EQUAL appears for the first time. This metric represents the number of share-adjusted pages of qualifying research (contained in journals listed in *EconLit*); in other words, a 20-page article in the *AER* is deemed to be equivalent to a 20-page article in an obscure regional journal. EQUAL is really a representation of quantity, not quality, but serves as a useful reference point when one is trying to judge the aggressiveness of alternative weighting schemes.

In Table 7 we reveal the relationship between our various measures of departmental output. It is clear that our three citation-based measures are very weakly correlated with KMS2010 (ISIB03–08: 0.10; ISIB01–06A: 0.06; and ISIB01–06B: 0.03).[22] Recall that KMS2010 is an updated version of a widely accepted, aggressive journal-based weighting scheme. On the other hand, the correlation coefficients for ERAB (the Australian government's research-evaluation scheme) and our various citation measures range from 0.88 to 0.91. Perhaps more surprising is the nature of the relationship between EQUAL and ISIB03–08, ISIB01–06A, and ISIB01–06B — they range from 0.93 to 0.94. This result might be explained by the fact that once a journal has been listed by ISI, all citations are deemed to be of equal value; and with respect to New Zealand's economists, papers published in lower-ranked journals appear to be as successful in capturing cites as those published in higher-ranked journals.

Table 7: Pairwise correlation coefficients, departmental output, weighted pages and citations per capita (2003–2008)

	EQUAL	KMS2010	ERAB	ISIB03-08	ISIB01-06A	ISIB01-06B
EQUAL	1.00	0.01	0.93	0.94	0.93	0.93
KMS2010		1.00	0.25	0.10	0.06	0.03
ERAB			1.00	0.88	0.91	0.91
ISIB03-08				1.00	0.96	0.96
ISIB01-06A					1.00	0.99
ISIB01-06B						1.00

Let us now turn our attention to individual economists. Given that the PBRF scheme evaluates individual performance, a movement away from the current peer-evaluation system to a more mechanistic scheme would undoubtedly produce many winners and losers. Although we are not able to generate proxy PBRF results, we are able to capture the nature of the relationship between our various output schemes. As shown in Table 8, we present the pairwise correlation coefficients between our three citation-based schemes (ISIB03–08, ISIB01–06A,

22 These correlations are not significantly different from zero.

and ISIB01–06B), and three alternative schemes (KMS2010, ERAB and EQUAL). For illustration purposes, our sample is restricted to output estimates for the top 30 researchers as ranked by EQUAL. We have done so since highly ranked producers by any measure have more to lose in the adoption of an alternative measure and because many economists in our sample have generated zero output under KMS2010 and all of our citation-based schemes.

Table 8: Pairwise correlation coefficients, individual output, top 30 (ranked by EQUAL), weighted pages and citations per capita, 2003–2008

	EQUAL	ERAB	KMS2010	ISIB03-08	ISIB01-06A	ISIB01-06B
EQUAL	1.00	0.74	0.06	0.43	0.51	0.55
ERAB		1.00	0.47	0.67	0.77	0.78
KMS2010			1.00	0.19	0.34	0.38
ISIB03-08				1.00	0.84	0.80
ISIB01-06A					1.00	0.99
ISIB01-06B						1.00

It is apparent that our three citation schemes are weakly correlated with KMS2010 — with correlation coefficients ranging from 0.19 to 0.38. On the other hand, the perception-based ERAB scheme yields much higher estimates: ISIB03–08 (0.67), ISIB01–06A (0.77) and ISIB01–06B (0.78). Of interest is the fact that, as opposed to the departmental situation, the relationship between EQUAL and the ISI-based measures is only of moderate strength: 0.43 to 0.55. It is clear that as far as individuals are concerned, an evaluation system based on citation counts yields different results from one based on journal weights.

Policy implications

Our findings suggest that the time-lags between publication and citing are such that it would be difficult to rely on citation counts as a meaningful measure of output in a PBRF-like research evaluation framework, especially one based explicitly on individual assessment and a six-year timeframe. Nationwide evaluation schemes such as the UK's Research Excellence Framework (REF),[23] Australia's Excellence in Research for Australia (ERA)[24] and New Zealand's PBRF attempt to provide an indication of recent research productivity. This is evidenced by the fact that they use a stock measure of output; that is, they select a census date that is as close to the portfolio submission date as possible, and then they assess each institution's research activity over the preceding six years.

23 For details, see www.hefce.ac.uk/research/ref The Research Excellence Framework (REF) will carry out its first nationwide evaluation in 2014; it replaces the Research Assessment Exercise (RAE) (www.rae.ac.uk) that, in many ways, served as a model for the PBRF scheme.

24 For details, see www.arc.gov.au/era

Hence, the average paper (assuming a relatively stable publication flow) is only in print for three years prior to the end of the assessment period. As we have shown above, three years after publication the vast majority of papers had not received a single ISI cite; and for those cited, early citation patterns can deviate substantially from those exhibited over a longer time period. This problem is much more severe at the individual level than at the departmental level (due to the effects of averaging). We found numerous cases wherein individual papers did not receive any cites until Year Eight or later, with some as late as 10 years.

On the other hand, an argument can be made that citation counts provide additional information that could be used in a multi-criteria evaluation system. Our work suggests that the output measures generated by citation counts are not highly correlated with traditional output measures based on journal-impact factors. This follows primarily from the fact that some papers in lower-ranked journals generate a relatively large number of citations, and some in highly ranked journals receive few, if any, cites. Therefore, especially if collected over a longer time period than the six-year window currently used by PBRF, citations could provide evaluation committees with useful information. However, if the collection period were extended — say, from six years to eight — it would create even more of an incentive to hire productive, experienced staff rather than young, inexperienced researchers.

Earlier, we drew attention to the fact that the number of JCR-listed economics and social science journals has expanded rapidly over time. For example, when Liebowitz and Palmer (1984) undertook the research that led to their ground-breaking work in constructing adjusted-impact measures, they relied on an ISI/JCR database which, at that time, listed only 107 economics journals. By 1998, the JCR economics list had expanded to 159 journals and by 2003 the number of JCR economics journals had reached 169. However, in recent years the list has expanded dramatically: to 209 in 2008 and 247 in 2009 (the most recent list at the time of writing — March 2011).[25] A similar expansion has undoubtedly taken place in other social science disciplines.

An expanding journal list leads to two effects: first, the percentage of publications eligible for citation collection has increased and will continue to do so; and second, the number of citations per paper should also increase as the number of eligible citation-generating journals has grown (all journals in the ISI database). This has both positive and negative effects on the value of a citation-counting scheme. On the positive side, it will minimise the impact of the regional-journal issue (as more and more are incorporated into the database).

---

25 The dates 1998 and 2003 were chosen because they represent the journal selection dates adopted by two of the major papers in the journal-based weighting literature: Kalaitzidakis, Mamuneas and Stengos (2003), and Kodrzycki and Yu (2006).

It also helps departments and individuals working in new and emerging areas of the discipline, since journals with a focus on these areas are more likely to be included in the 'eligible list' than in the past.

On the other hand, the less discriminating the 'eligible list' becomes, the more pressure will arise to challenge the assumption that all citations are of equal value. One may find cries to weight citations by, say, the relevant JCR Impact Factor; however, this leads to problems similar to those arising from earlier efforts by economists to apply differential weights to cites in the development of adjusted-citation journal-weighting schemes, of which KMS2010 is a prime example. The primary argument against weighting is that it mixes individual performance (the number of citations to a given paper) with the average performance of others with papers in the same journal, and, indirectly, with the quality of the editorial staff at any point in time (the ability to pick winners!).

Summary and conclusion

In this paper we have attempted to assess the merits of using citation counts per researcher as part of a nationwide research-assessment exercise, with particular reference to the discipline of economics. Two issues gave rise to our interest in this subject. First, the growing interest in using bibliographic techniques in research-assessment exercises driven, in part, by advances in information technology; and second, the concerns expressed by many social scientists over the merits of using citations to measure performance, especially with respect to the nature of the time-lag between publication and the generation of a meaningful flow of citations in their disciplines.

We explored these issues in the context of a single discipline — economics — and a single nation — New Zealand. Our findings, based on a 10-year collection period, suggest that cites are, indeed, initially slow to develop; for example, the proportion of cites collected over a 10-year period that are generated within the first three years of publication is in the order of 10 per cent. This estimate rises to roughly 30 per cent in Year 5. We also found that roughly 40 per cent of papers received one or more citations, 20 per cent five or more citations, and slightly less than 5 per cent received 20 or more citations. However, we must stress that many papers in our sample were not eligible for ISI citation collection. After adjusting for this fact, we found that slightly over 73 per cent of eligible papers were eventually cited within the period of our analysis.

In general, our findings suggest that the conventional assessment period of six years may be acceptable from a departmental perspective due to averaging effects, but that this is too short a time period for individual assessment.[26] This arises from the fact that the average paper will have only three years to collect citations. Although this problem can be addressed, in part, by expanding the citation collection period, doing so provides an additional incentive for departments to, in effect, buy 'CVs' rather than hire young, inexperienced researchers. Overall, we agree with the view expressed on the REF's website: 'The pilot exercise showed that citation information is not sufficiently robust to be used formulaically or as a primary indicator of quality; but there is considerable scope for it to inform and enhance the process of expert review.'[27]

Although this study is set in a New Zealand context, our findings should be of interest to academics and higher-education policy analysts in other domains who are considering the use of citations as a research-assessment measure in the social sciences. In addition to demonstrating that the lag between publication and the generation of a meaningful flow of citations is problematic in the context of the time span normally employed in a nationwide research-assessment scheme (five or six years), our findings raise other policy-relevant issues. First, our work demonstrated that the choice of a citation collection scheme is not a trivial matter. For instance, we found that Google Scholar not only generates more cites, but does so much sooner after publication than does the ISI. However, the former engine is much less discriminating, and its use in a nationwide program will undoubtedly give rise to cries for 'quality' adjustment — something that is very difficult to implement in a real-world setting. Second, the widespread concerns that young academics are disadvantaged by journal-based evaluation schemes will only be reinforced by citation-based schemes that increase the time between initiation of a research program and the generation of measurable outputs. A third issue relates to the generality of our findings: is the economics discipline representative of the social sciences, and is the research performance of New Zealand's economists similar to that found in many nation-states? We must leave the answer to these questions to others.

26 Note that individuals receive notification of their score, and regardless of confidentiality rules, outcomes are widely known within departments, and perceived to be used in promotion and merit-pay assessments. Hence, the generation of individual scores may have long-term career implications.

27 www.hefce.ac.uk/reserch/ref/Biblio/ (Accessed 25 March 2011).

References

Anderson, D. L. and Tressler, J. 2010, 'The Merits of Using Citation-Based Journal Weighting Schemes to Measure Research Performance in Economics: The Case of New Zealand', Working Paper 10/3, Department of Economics, University of Waikato, Hamilton, New Zealand.

Centre for Science and Technology Studies 2007, *Scoping Study on the use of bibliometric analysis to measure the quality of research in UK higher education institutions*, Leiden University, U.K.

Chang, C. L., McAleer, M. and Oxley, L. 2011, 'What Makes a Great Journal Great in Economics? The Singer not the Song', *Journal of Economic Surveys* 25: 326–61.

Cinlar, N. and Dowse, J. 2008, *Human Resource Trends in the Tertiary Academic Workforce*, Tertiary Education Commission, Wellington, New Zealand.

Cosme, P. and Teixeira, A. 2010, 'Are finance, management and marketing autonomous fields of scientific research? An analysis based on journal citations', *Scientometrics* 85: 627–48.

Coupe, T. 2003, 'Revealed Performances: Worldwide Rankings of Economists and Economics Departments, 1990–2000', *Journal of the European Economic Association* 1(6): 1309–45.

Goldfinch, S. 2003, 'Investing in Excellence? The Performance-based Research Fund and its Implications for Political Science Departments in New Zealand', *Political Science* 55(1): 39–53.

Gordon, I. R. and McCann, P. 2000, 'Industrial Clusters, Agglomeration and/or social Networks?', *Urban Studies* 37(3): 513–32.

Henrekson, M. and Waldenstrom, D. 2011, 'How Should Research Performance be Measured? A Study of Swedish Economists', *Manchester School* 79(6): 1139–56.

Hodder, A. P. W. and Hodder, C. 2010, 'Research culture and New Zealand's performance-based research fund: some insights from bibliographic compilations of research outputs', *Scientometrics* 84: 887–901.

Kalaitzidakis, P., Mamuneas, T. and Stengos, T. 2003, 'Rankings of Academic Journals and Institutions in Economics', *Journal of the European Economic Association* 1(6): 1346–66.

——— 2010, 'An Updated Ranking of Academic Journals in Economics', Working Paper 9/2010, Economics Department, University of Guelph, Guelph, Canada.

Kodrzycki, Y. K. and Yu, P. 2006, 'New Approaches to Ranking economics Journals', *B.E. Journal of Economic Analysis and Policy: Contributions to Economic Analysis and Policy* 5(1), Article 24.

Laband, D. and Piette, M. 1994, 'The Relative Impact of Economics Journals', *Journal of Economic Literature* 32(2): 640–66.

Liebowitz, S. J. and Palmer, J. P. 1984', 'Assessing the Relative Impact of Economics Journals', *Journal of Economic Literature* 22(1): 77–88.

Macri, J. and Sinha, D. 2006, 'Rankings Methodology for International Comparisons of Institutions and Individuals: An Application Economics in Australia and New Zealand', *Journal of Economic Surveys* 20(1): 111–56.

OECD 2010, *Performance-based Funding for Public Research in Tertiary Institutions: Workshop Proceedings*, OECD Publishing.

Oswald, A. J. 2007, 'An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-Makers', *Economica* 74(1): 21–31.

Smith, A. G. 2008, 'Benchmarking Google Scholar with the New Zealand PBRF research assessment exercise', *Scientometrics* 74(2): 309–16.

Starbuck, W. H. 2005, 'How much Better are the Most-Prestigious Journals? The Statistics of Academic Publication', *Organization Science* 16(2): 180–200.

Wall, H. J. 2009, 'Don't Get Skewed Over by Journal Rankings', *B.E. Journal of Economic Analysis and Policy: Topics* 9(1): Article 34.

ARGUMENT

Paved With Good Intentions: *The Road Home* and the Irreducible Minimum of Homelessness in Australia

Gary Johns[1]

Abstract

Despite public interest and public inquiries as long ago as the 1989 Human Rights Commission and the consequent increase in funding by the Hawke and subsequent governments, the 2006 rate of homelessness in Australia was 32 per 10 000, only marginally better than the 34.8 per 10 000 of 2001. This article explores the reasons for the apparent failure to fulfil past promises and the weaknesses in current strategy. It argues for an explicit policy framework containing the assumption that there is an irreducible level of homelessness, and that there should be an assessment of the cost-effectiveness, probability of success and timeliness of each known remedy, in order to achieve the irreducible level.

Introduction

In its 2008 White Paper *The Road Home*, the Australian Government promised to halve by 2020 overall homelessness and offer supported accommodation to all rough sleepers who were deemed to need it. The then Prime Minister, Kevin Rudd, contended in the same document, that 'In a country as prosperous as Australia, no-one should be homeless' (FaHCSIA 2008a: Foreword). The apparent difference between Rudd's contention and his promise suggests a lack of political will. Or was it recognition that homelessness can be reduced, but is not solvable? The Rudd goal was no doubt a political aspiration, constrained by, on the one hand, the wish not to repeat Prime Minister Bob Hawke's promise of 1987 ('no Australian child will be living in poverty'[2]) and, on the other, a need to make sufficient 'announcement' impact. Indeed, the stated goal may amount to an implicit admission that, despite best endeavours, there is an irreducible level of homelessness.

1 Australian Catholic University; gary.johns@acu.edu.au

2 1987 Federal election campaign launch.

This is not to ignore the exhortation to 'do better' that a goal implies, but an admission that an irreducible level of homelessness should, if proved, be made an explicit part of policy. Indeed, the goal of such policy may well be to reach the irreducible level of homelessness. Whether such level is 'half' the present level — or less, or more — is unknown. This paper proposes a way to discover the irreducible level of homelessness. The proposed method is to distinguish causes of homelessness and, for each cause, to assess the cost-effectiveness, probability of success and timeliness of each present remedy, and alternative remedies. At the same time, account is taken of measures that will reduce the stock of homeless and distinguish these from those that reduce the flow of homeless.

Some history

In 1989 the Australian Human Rights Commissioner, Brian Burdekin, reported on homelessness among the young in Australia in *Our Homeless Children* (AHRC 1989) and made homelessness a headline issue. He was appointed by a sympathetic government and he reported to a sympathetic Minister in Brian Howe, a member of a pro-intervention cabinet. The responses to Burdekin's recommendations were exhaustive and continued in subsequent administrations. Contributions under the Commonwealth State Housing Agreement rose from $1.27 billion to $1.51 billion in the period 1990–95 (HAA 2002: 105); funding for Supported Accommodation Assistance Program (SAAP) agencies[3] increased to $400.4 million by 2008, a 28 per cent increase in real terms from 1996, and the number of SAAP agencies increased from 500 to 1562 in the period 1985–2008 (AIHW 2009, 7: 297).

In 1993 the Australian Human Rights Commission conducted a national inquiry into the human rights of people with mental illness (AHRC 1993), which highlighted mental illness as a major factor in homelessness. Although the responses to the second inquiry were not as comprehensive as those that followed the first, there were nevertheless considerable changes to the treatment of those with mental-health problems (Ozdowski 2001). The new inquiries and the government responses unfortunately appear to have been insufficient, as the rate of homelessness recorded in the 2006 Census was 32 per 10 000 of the population, down only marginally from 34.8 in 2001 (ABS 2011a: 78 and 90).[4] These estimates have been subject to major adjustment as a result of substantial measurement issues, discussed below. This modest gain occurred in circumstances where employment rose 13 per cent, unemployment fell 22 per

3 Not-for-profit organisations that deliver support to the homeless.

4 Comparable data are not available in earlier censuses.

cent, and disposable household incomes of low-income households rose 38 per cent in nominal dollar terms. By contrast, rents rose only 17 per cent during this period and rental vacancy rates remained above the 3 per cent 'neutral' level until at least the end of 2004. Reported domestic violence, which is the greatest single cause of homelessness, fell 22 per cent and the divorce rate fell 12 per cent between 1996 and 2005 (ABS 2011a: 12).

For declining to endorse the Rudd goal of halving homelessness, the Leader of the Opposition, Tony Abbott, was roundly criticised (Dunkerley 2010). Abbott was, in effect, arguing that many causes of homelessness — broken relationships, broken minds, broken hearts, and broken bodies — were not preventable. While each of these may be 'mended', they were not preventable in a large population. To the extent that these problems cause homelessness, homelessness cannot be prevented. The amount of time spent homeless or the amount of 'cycling' through poor housing may be diminished, but it is most improbable that causes of homelessness can be abolished. The Rudd Government was attempting to bypass difficult policy and political issues; respectively to admit which aspects of homelessness were not solvable and the implications for achieving the stated goal of halving homelessness.

Nevertheless, there are clear gains to the individual and to society from some programs for the homeless. As a result of receiving support from the SAAP homelessness program, for example, the homeless may use fewer other government services. One study of SAAP services (Flatau *et al*. 2008: 10) has found that budget outlays for all other programs consumed by the homeless are at least twice as large as the annual SAAP program cost. In other words, there are significant potential net government cost savings from providing assistance. For example, the average cost per client for SAAP-Single Men clients is $4 625; the associated saving is more than twice that amount, at $10 212, resulting in net cost savings of $5 587 per person per year (*Ibid*: 12).

These gains, however, address the 'stock' or extant number of homeless (or, at least, those who are eligible for SAAP services); they do not necessarily address the flow of new entrants. A well-run policy may not lower homelessness at all; rather, it may deal efficiently and effectively with an ever-refreshing stock of homeless people, returning them to better circumstances at least cost. The positive outcomes to homelessness programs should be applauded but their limits in reducing it should be recognised. Housing assistance in all its forms, for example, is channelled to those most in need (AIHW 2011: viii). Services to the homeless should not be an exception. Social triage is obviously important in this context and the great work of those in the field must be acknowledged. Arguably, however, once the point has been reached at which homelessness is irreducible the policy goal should switch from reduction per se to reduction of turnaround times for those for whom such remedies are appropriate and, where not possible, to long-term provision.

The *Road Home* strategy — or lack of it

The very large investment in public and private inquiries, public policy and public expenditure (not to mention the considerable efforts of the charitable sector) over many years until 2008 has not had the desired effect of reducing homelessness by much, let alone eliminating it. Indeed, and perhaps reflecting some frustration with the apparently intractable nature of homelessness, shortly after the White Paper was published, representatives of the welfare sector completed its own inquiry, an 'independent' community inquiry funded by the Caledonia Foundation, the National Youth Commission *Inquiry into Youth Homelessness*. The report called for $1 billion of extra public funds over 10 years and, curiously, without offering its own, called for a government 'policy strategy' (NYC 2008: 374). Perhaps reflecting these apparent failures in strategy and funding, the Rudd Government announced a 55 per cent increase on the then current expenditure on homelessness, representing an additional $800 million over four years (FaHCSIA 2008a: Foreword), and three strategies to halve homelessness. In unison with previous efforts, however, no argument was advanced in the White Paper about the relationship between expenditure and result.

Instead, the White Paper announced the strategies:

- 'Turning off the tap': early intervention to prevent homelessness, especially in families
- Services to end homelessness through training and employment, and
- 'Breaking the cycle': moving people quickly into supported accommodation.

(FaHCSIA 2008a: chapter 2)

At face value, these strategies, coupled with a 55 per cent increase in funds, appear 'authoritative and rational' (Colebatch 2006: 42) and therefore constitute a solid foundation for public policy.

In fact, the goal, the funding and the strategies are less authoritative and rational than they appear. This is so because elements of the policy are assumed, not proved. The 50 per cent reduction goal is to be achieved by a combination of turning off the flow of the newly homeless; ending homelessness of the current stock more quickly; preventing the homeless from repeating the cycle; or, simply adding to the stock of accommodation and services for the homeless. Just how realistic each of these strategies is, or how much each of these distinct elements will contribute to the 50 per cent reduction or, indeed, how the funding of each will produce the best return is nowhere explained in the White Paper. Fifty per cent of new funding is to be devoted to prevention (FaHCSIA 2008a: chapter 2), for example, but it is unknown whether these funds will lessen homelessness by a proportionate amount. The services to end homelessness promise some new directions — for example, investing in individuals through the 'breaking the cycle' strategy — but the old directions — adding to the stock of accommodation and homeless services, which has been the 'standard strategy'[5] of the sector for many years — remain.

While the Rudd Government established a Council on Homelessness[6] and funded a national longitudinal survey of Australians experiencing homelessness or vulnerable to homelessness,[7] it did not offer a strategic assessment of where the numbers to reduce homelessness were to be found. The data available to the Inquiry on the cost of homeless services were limited (Pinkney and Ewing 2006; Flatau *et al*. 2008) but this should not have prevented an intention to establish a least-cost pathway to reach the goal. Knowing how much to spend on what programs would seem a necessary step in either achieving the goal or proving its validity. It appears, however, that the government had no idea which strategy would contribute to what amount of reduction, at what cost or, indeed, whether the goal was feasible, too ambitious or not ambitious enough. It should have been possible at least to conceive an analysis that would inform such a strategy.

The object of the remainder of the paper is to suggest such an analysis, commencing with a review of the number and characteristics of the homeless, an assessment of the causes of homelessness, the strengths and limitations of current programs, and an assessment of where and what reductions are likely. New goals for homelessness policy are recommended.

Measurement issues

The headline homeless figure at the time of the White Paper, drawn from the 2006 Census, was 105 000. This was composed of 16 000 actual homeless (those sleeping rough) and 89 000 potential homeless. The Australian Bureau of Statistics has, however, recently revised down massively its 2006 (and 2001)

5 From author's conversation with Lin Hatfield Dodds, Uniting Care Australia, during the development of the paper.
6 Communiqué from Meeting 9, Sydney, 12–13 May 2011.
7 The Melbourne Institute of Applied Economic and Social Research, 'Journeys Home: longitudinal study of factors affecting housing stability'.

Census estimates (ABS 2011a: xi). The 2006 revised amounts are 8 000 actual homeless and 55 000 potential homeless. (It seems, thus, that Prime Minister Rudd's promise has been reached before his programs started). The reasons for the revisions are many; but, in short, the ABS estimated homelessness using the Census when it did not include a homeless indicator. This meant that while some groups of homeless people were unobservable in the Census data, several other groups who were observable, but not homeless, were counted as homeless. For example, construction crews and owner builders, young professionals travelling overseas, and people following mobile employment opportunities were counted as homeless (ABS 2011a: 1). Making inferences about the observable — indeed, searching for the homeless in nonsensical categories — has brought the count into grave error. Accepting the revisions and, in the light of previous persistence of homelessness, there is, nevertheless, the question of how a government should proceed to achieve the goal of lowering homelessness. There are a number of relevant considerations.

The number of homeless reported at the 2001 and 2006 censuses (see Table 1) is estimated at one point in time and represents the stock of homeless. The number who experience at least one period of homelessness in a year is considerably higher and represents the flow of homeless (more than 300 000: ABS 2011c). The extent of recycling has implications for reducing homelessness; indeed, it needs to be borne in mind that there is a considerable amount of recycling in the flow numbers. In a study of homeless people in Melbourne, for example, 92 per cent had moved from one form of temporary accommodation to another — friends or relatives, boarding houses, supported accommodation — and had slept rough (Johnson and Chamberlain 2008: 342). As with the estimation of the stock of homeless, there is also the potential to overestimate the flow of homeless. In SAAP, for example, while a person who has multiple support periods in a single financial year is only counted as one client, they may record a client twice when his or her support period extends over two reporting periods (Flatau *et al*. 2008: 212). In relation to services for the homeless, the stock measure is generally most useful because these services are required on a daily basis, regardless of the flow through the services. Nevertheless, the length of time that a client requires a service or accommodation is crucial to overall resourcing.

The ABS was persuaded to adopt a cultural definition of homelessness. The cultural definition 'identifies shared community standards about the minimum housing that people have the right to expect' (ABS 2011a: 1). The accommodation categories — tertiary, secondary, primary[8] — are a valid way of gaining a sense of who is at risk of homelessness — those staying in boarding houses, or in supported accommodation (for example, a women's refuge), or with friends.

8 Primary being actual homeless, and secondary at considerable risk of, and tertiary, at some risk of homelessness.

While the concept is useful in enumerating the potential pipeline of homeless — that is, those who sleep rough — it runs the risk of overstating the case for the homeless. Others have cautioned against the tendency for 'a single big number … for advocacy purposes' (Pinkney and Ewing 2006: 12). Understanding the market in homelessness is fine; distinguishing the potential flow that may feed primary homelessness is legitimate; but building the numbers for advocacy is not helpful in as much as this may concentrate policy on accommodation and away from homeless services. Indeed, it may carry the implication that sufficient investment in accommodation, 'business as usual', will solve the homelessness problem.

The implication comes to grief in the first three categories of homelessness in Table 1. These define homelessness by their housing 'solution'. This mixes problems with solutions. A boarding house, SAAP, or staying with friends and family is an adequate solution for some people. There is an unstated assumption in a cultural definition of homelessness that there is some other remedy; for example, that it is realistic to expect a house to be readily available at short notice for the right price in the right area for those with short-term problems. The recent General Social Survey estimated that of those who had experienced homelessness, 13 per cent were homeless for less than a week, a further six per cent were homeless for less than two weeks, and another 12 per cent were homeless for less than four weeks (ABS 2011c). Some forms of accommodation presently relied on in government programs may well be all that is realistically available or, indeed, necessary.

Table 1: The homeless, defined by their accommodation

Homeless 'accommodation'	Homeless category	2001	2006
Boarding houses	Tertiary	23 748	16 828
SAAP	Secondary	13 420	17 331
Friends/relatives	Secondary	17 877	19 579
Persons in other temporary lodging	Primary	1395	1970
Rough sleepers, improvised dwellings	Primary	8943	7764
Total		65 384	63 472

Sources: ABS 2011a: tables 1.1 and 1.2, 4–5, and ABS 2011b. The figures are the result of census estimation, not a statistical estimate and thus are reported exactly, rather than rounded.

The situation for the long-term homeless is different. Twenty-two per cent spent six months or more without a permanent place to live (ABS 2011c), typically single men in boarding houses and women in supported accommodation (ABS 2008: 33). For these people the issues may be more difficult than for short-termers. The issue for the latter is how long someone can be supported without recourse to institutional care.

The cultural definition of homelessness also implies that there are solutions to forms of homelessness other than staying with friends/relatives. The inclusion of the friends/relatives category implies that a family-based solution is not acceptable and that there is no solution without government assistance. Staying with friends and family seems a legitimate course of action. In fact, the greatest overestimation in homelessness was in the category 'staying with friends and family', which in 2006 was 19 579, down from 46 856.

Similarly, it is implied that boarding houses are not a legitimate solution for some people, even long term, or that refuges are not essential solutions to those who need safety on a short-term basis. Non-specific or open-ended solutions implied by these categories are part of the difficulty of addressing homelessness. It is a contest between an assumed norm that does not exist and the reality of a very long commitment by a sector and governments to assist those who need assistance of varying types and kinds that, nevertheless, falls short of a solution.

Seeking the irreducible minimum of homelessness

The *reasons* why people are homeless, rather than their form of accommodation, may provide a far stronger clue to the necessary remedy to their problems. Indeed, the fact that an array of disparate people shares a similar problem may well cloud the nature of the remedies that need to apply. The causes of homelessness as established by the White Paper include a shortage of affordable accommodation, which may indicate a problem with the housing market, the remedies for which may be taxation or regulation (AHURI 2003; AHURI 2011). But this may have very little to do with remedies associated directly and specifically with the problems of other homeless people.

The problems of family breakdown are often entirely different in nature from those of affordability (although there will be affordability consequences in family breakdown) and exist in a realm where the remedies will be very different from the market-based remedies associated with affordability. Affordability issues may realistically be solvable; the problems of family breakdown, less so. Unemployment, like affordability, may be a market problem; then again, if unemployment is related to drug, alcohol and mental-health issues, finding someone a job will not be the solution to their housing problem. Those with drug, alcohol and mental-health problems encompass socio-cultural and medical dimensions, the remedies for which are different again from those of affordability and family breakdown. The homelessness of Aborigines is probably entirely different again from any of former, as the causes are both cultural and economic.

The diverse nature of those who are homeless, and the diverse nature of the remedies, will have a major impact on estimating the irreducible level of homelessness. Identifying the baseline level of homelessness expected under 'normal' conditions may be difficult but it should help to frame the goals of homelessness policy. With some concept of a baseline, and appreciating that feasible 'remedies' can be time- and place-specific according to what the society can afford or demands (Galston 2006: 545), the baseline feasibility criteria serve to underline that solutions may not be available for all categories of need. The most effective remedies applied to each distinct category of need should reveal the difficult cases and the limits to solutions. A differentiated approach will raise sensitive issues about priorities, but a non-differentiated approach leads policymakers to chase a goal that is not achievable.

The White Paper makes a number of naïve statements, claiming (hopefully), for example, that the policy 'provides a framework for preventing homelessness from occurring in the first place' (FaHCSIA 2008a: Foreword). It boasts strategies to increase support for victims of domestic violence to stay safely in their own home, for example, but it does not inspire confidence that such a strategy reduces homelessness. Is there housing support for the perpetrator, for example, who may become homeless as a result of a safety-first strategy? The policy of 'no exits into homelessness' from hospitals, mental-health and drug and alcohol services and statutory care sounds promising but, of course, it is predicated on there being no solution to the problems for which hospitals, jails and other institutions are the intermediary solution. The flow remains; the stock needs to be addressed.

In order to fulfil the promise of lowering homelessness, governments and the sector need to think through a number of links in the chain of strategies and remedies applied to homelessness. The suggested model consists of understanding the limitations (cost-effectiveness and acceptability) in each strategy — turning off the tap, quicker turnaround, better services — and the attendant remedies for each type (cause) of homelessness.

The irreducible level of homelessness and the reasons why it is irreducible needs to be identified. Where such admissions are left unstated there is a risk of wasted effort or adverse consequences. For example, 'full employment' is defined well above zero unemployment for various reasons to do with the mismatch of skills required and skills offered in a market. There is, for all practical purposes, an irreducible number below which unemployment is unlikely to drop. Further, efforts to reduce the number may have inflationary consequences once the level is reached. Indeed, there is a measure, the non-accelerating inflation rate of unemployment (NAIRU), which represents equilibrium between the state of the economy and the labour market. By analogy, if indeed there is an irreducible level of homelessness, it follows that the idea of a homelessness 'solution' may have to be redefined.

Attempts to lower homelessness below the irreducible level could constitute inducements to leave present situations with friends and family to access government services. The inducement effect of accommodation services has recently been observed in a review of SAAP services. The study concluded that 'an increase in capacity may have flow-on effects on the number of people seeking accommodation because, as supply increases, so too might the demand for that accommodation' (AIHW 2010: 62). In other words, the availability of more beds in the sector could result in an increase in the demand for accommodation. The figures in Table 1, for example, indicate an increase in those seeking SAAP services and a diminution of those seeking boarding-house services between 2001 and 2006. The shift may have been caused by the substantial lift in investment in SAAP in the period (AIHW 2009, 7: 9). An attractive set of services for the homeless may simply shift effort from individuals to the taxpayer and NGOs — for example, from friends and family to SAAP services, or between one form of government assistance to another — without solving anything.

Moving in the opposite direction, many of the places accommodating those with mental-health and substance-abuse problems have been closed in the last 20 or 30 years. The policy aspiration was that people should live in 'the community'. Unfortunately, the policy of de-institutionalisation sometimes resulted in people remaining uncared for and liable to end up on the street. The Mission Australia Centre and the Foyer model, discussed below, are examples of new institutions where people will receive their services in situ. The mentally ill and long-term drug dependent may well require new institutions. Reaching the goal of the irreducible level of homelessness may be impossible without revisiting institutions and their place in the suite of available remedies.

Similarly, if the proximate cause of homelessness among Aboriginal people is a lack of engagement in the economy, supplying public housing in the middle of an area where there are no jobs will not constitute a remedy. The best prospects for a solution may be in migrating to centres with job opportunities (Johns 2008). The policy suite itself may need to change in order to reach the goal of either halving homelessness or reducing the irreducible level of homelessness.

Assessing causes

Reaching the goal of irreducible homelessness requires a reasonable estimation of the number of people affected by each cause. There appears to be no consolidated account of the *causes* of homelessness sufficient to account for the entire cohort. The most plausible approach is to use surveys of those using SAAP services. Although SAAP services constitute less than 30 per cent of the total number of homeless and care must thus be taken in interpreting the figures as a representation of the homeless, they are, nevertheless, accurate in their own realm. As for other forms of homelessness and their causes, the data are unhelpful; there is only notional data on causation for those in boarding houses. Nevertheless, it is certain that most of those who pass through SAAP and boarding houses have also slept rough and/or stayed with friends. In other words, cycling is common and so the distribution of causes of homelessness as measured for SAAP may be a reasonable representation of causation throughout the sector (with the proviso that more single men attend boarding houses and more women and younger people access SAAP services). The categories are not exclusive; for example, 'those being unable to afford accommodation' may be unemployed. And 'previous accommodation ended' may be caused by financial hardship but may be from any other cause which led to the person being there in the first place.

Table 2: Estimating the proportion of homeless by cause

Homelessness causes (*The Road Home* 2008)	Homelessness causes of those accessing SAAP services	%
Shortage of affordable accommodation	Usual accommodation unavailable/ overcrowding	18
	Financial difficulty	15
	Eviction/previous accommodation ended	17
Emergency response to family violence Youth separation from family	Time out from family/other situation/ interpersonal conflict	12
	Relationship/family breakdown and domestic/family violence	32
	Physical/emotional abuse/sexual abuse	3
Unemployment	Recent arrival to area with no means of support	4
Long-term mental health	Mental health/other health	4
	Recently left institution	2
Substance abuse	Drug/alcohol/substance abuse	6

Sources: Australian Institute of Health and Welfare 2008: 32.

Table 2 contains estimates of people who access SAAP services, categorised by the cause of their homelessness. Understandably, people who need SAAP services because of a shortage of affordable accommodation are a prominent proportion of those surveyed. It is not clear, however, whether their 'financial difficulty' or the 'usual accommodation becoming unavailable' is the consequence of some crisis event. For example, domestic violence has consistently been the major reason people seek support from specialist homelessness services during the life of SAAP; as a consequence, 80 per cent of SAAP recipients are single parents. The categories of relationship and family breakdown and family violence represent more than 30 per cent of the reasons why people use its services. Moreover, there is evidence to indicate that a similar proportion use boarding houses for the same reason (Anderson *et al*. 2003: 4). Much has been made of the incidence of mental-health and substance-abuse problems as a cause of homelessness. The SAAP figures, however, indicate that these are not as great as implied in the White Paper.

The data on Indigenous homelessness in remote areas are confusing. The total number of homeless in remote areas constitutes six per cent of all causes. Although it is assumed that most would be Indigenous, there is no definitive figure. Aborigines constitute 18 per cent of all SAAP users, but the 2006 Census dataset classified Aborigines as 10 per cent of all homeless persons. However, Indigenous status is not reported for a significant proportion of those who were classified as homeless (14 per cent). If those people for whom Indigenous status was not provided are excluded from the calculation, Indigenous people make up 12 per cent of the reviewed estimates of homeless people (ABS 2011b).

Assessing remedies

The most common remedies presently available to address homelessness are income support and, as noted above, various forms of supported accommodation. Specific 'new' remedies, those being trialled to address particular groups of homelessness (often emphasising services in addition to shelter) place a great emphasis on prevention and rehabilitation. As noted above, 50 per cent of the new funding is being devoted to the prevention strategy. Although not well specified in the White Paper, these programs are designed to keep families together, concentrate on the prevention of domestic violence and to minimise risk to children in the home. Other strategies contribute to the most efficient and effective return of those who are homeless to a more permanent housing situation. Which remedy contributes to which strategy and by how much is not apparent from the White Paper. Nevertheless, a selection of major new remedies is set out below and briefly reviewed with respect to their strategic contribution and efficacy:

Family violence

A recent overview of women, domestic and family violence, and homelessness concluded that 'a substantial part of the solution in terms of strategies that prevent domestic and family violence related homelessness are those that prevent domestic and family violence from happening in the first place'

(FaHCSIA 2008b: 31). Unfortunately, the synthesis report did not quote a single instance of a successful program to prevent violence. While case studies were quoted, no proof was presented to indicate why, how or, indeed, if (other than by the author's word) these were successful. This study, if it is the best available, does not provide a single figure to indicate any program that may have reduced domestic violence and that may, in turn, have reduced homelessness. The COAG report *National Plan to Reduce Violence Against Women and Their Children* (COAG 2010), contains nothing more than targets and recitations of many of the same programs in the FaHCSIA report.

The NSW Auditor General found a reduction in family violence rates for Aborigines in the period 2002 to 2008 although no evidence was presented that proved the programs had anything to do with the result (NSWAG 2011: 17 and 30). The NSW Ombudsman reported that the *Safe Families* program in NSW, aimed at reducing sexual assault against children in Aboriginal communities, was operational in only two of the five proposed locations and that a total of six families in Wilcannia and two families in Lightning Ridge had received case management! The funding for *Safe Families* was $22.9 million, and is now entering its final year of operation (NSWO 2011: 21).

These two foundation documents and the NSW case studies leave empty the central strategy of the largest component of the White Paper homelessness strategy. On the basis of this evidence, these contributions to the 'turn off the tap' strategy are unlikely to yield much reduction in homelessness.

The 'Foyer' model

In the Foyer model each 'Foyer' is managed by a steering group made up of local members of the public, elected representatives, business people and professionals so as to assist homeless youth to prevent cyclical homelessness. The first Foyer model implemented in Australia is the *Live 'N' Learn Campus* (Miller Campus) based near Liverpool in NSW. The centre includes life-skills training — for example, skills to obtain and maintain a tenancy, healthy cooking and budgeting — as well as more formal skills acquisition. In strategy terms, it is appears to aim to 'turn off the tap' and 'end homelessness'. An evaluation of the *Live 'N' Learn Campus* (AHURI 2006: 4) indicates promising results. The model raises the question of the extent to which Foyer is an institution, and whether this will be the first of other attempts to reinvest in institutional care.

The *Reconnect* program

The *Reconnect* program (FaHCSIA 2003) funded by the Commonwealth government uses community-based early-intervention services to assist young people aged 12 to 18 years who are homeless, or at risk of homelessness, and their families. Young people using *Reconnect* have, for the most part, not yet left home or have only recently left home. This program fits squarely the 'turning off the tap' strategy. An evaluation of the program found that *Reconnect* has been successful in improving stability in young people's living situations (FaHCSIA 2003: 8). Young people living in temporary situations fell from 16.5 per cent at *Reconnect*'s initial intervention to 5 per cent at exit from the services. *Reconnect* intervention also increased the stability of young people's living situations in relation to parents — young people living with parents increased from 58 per cent at the start of support to 62 per cent after support (FaHCSIA 2003: 9). The evaluation contained no cost data, but the program nevertheless appeared to be mildly effective.

Outreach teams

One particular program, *Street To Home*, was established in 2004 and aims to find accommodation solutions for people sleeping rough in Inner Adelaide. It sees the provision of housing as a first step out of homelessness and therefore sits squarely in the 'breaking the cycle' strategy. The agency provides services over extended hours and staff work to locate, identify and engage with rough sleepers through a case-management response to help individuals make the transition from homelessness into secure housing. The available evidence is that, as of 2007, the service had 'assisted 40 of the inner city's most chronic rough sleepers into long-term sustainable housing. It ha[d] also helped 108 people into long term accommodation, with a 99.5 per cent retention rate; 177 have been referred to mental health services' (Newman 2007: 41). The referrals raise the prospect that long-term solutions for some of the mentally ill may require more permanent and formal institutional arrangements. Unfortunately, there is no cost-effectiveness analysis available to indicate if the program has succeeded.

Mission Australia

Mission Australia recently decided to use its funds to provide high-quality crisis accommodation to 40 people, where previously it provided accommodation to 110. Under the new model, clients receive 32 different professional services, including training, educational and other services, which are delivered to these 40 people on site (Mission Australia 2009). This switch in service from quantity to quality clearly required considerable strength of management because it denied a bed to 70 other men. Unfortunately, there was no cost-benefit analysis to prove that the sacrifice was worth it, the report simply stating that 'goals had been achieved' and 'programs delivered', although there were improved accommodation and employment outcomes after one year (*Ibid*: 22). None of the costs attributed to the program were assigned to the outcomes. There is a clear attempt to reduce the recurrence of homelessness, which fits the government's strategy to end homelessness. This remedy shows promise and may play some part in reducing not only the stock of homeless but that part of the flow of homeless attributable to 'cycling'.

National Partnership Remote Indigenous Housing

The National Partnership Remote Indigenous Housing funding of $5.5 billion over 10 years ($1.94 billion new funding plus previous funding allocated under the Australian Remote Indigenous Accommodation Program funding of $3.55 billion) will provide up to 4 200 new houses in remote Indigenous communities and upgrades to 4 800 existing houses through a program of major repairs and/or replacement (Macklin and Rudd 2008). The $5.5 billion will be spent in 26 priority settlements. All but two of these settlements are former missions or government ration stations and each is almost wholly or predominantly Aboriginal (Johns 2011: 255).

This program should, by increasing the amount of available and affordable housing, presumably satisfy strategy three, breaking the cycle. Unfortunately, this program is perhaps destined to be the largest and least-effective intervention. Economic prospects are slight in almost all of the 26 preferred settlements (*Ibid*: 255). The Government's criteria for selecting them were that there was significant concentration of population and that there were pre-existing shortfalls in government investment in infrastructure and services. The Government suggested that there was potential for economic development and employment, but no data to justify the selection accompanied the announcement.

It should be possible to use existing knowledge to assess the cost-effectiveness, probability of success and timeliness of known remedies for each sector of the homeless market in order to establish priorities and as a means of judging whether a goal is achievable. Unfortunately, the brief survey above suggests there is much to be done to inform policymakers. A recent major study aimed at developing costing of homelessness programs found that 'existing (or at least, publicly available) cost information is not adequate for the purposes of evaluation research' (Pinkney and Ewing 2006: viii). A recent review of SAAP services (AIHW 2010) suggests, however, that sufficient data are available for that service. Where data are inadequate, one possible means of investigating the strategic approach to estimating a pathway to the irreducible level of homelessness is to establish likely scenarios of different forms of homelessness (Kenway and Palmer 2003). While the cost data may not be precise, it should nevertheless be possible to assign nominal values to the effectiveness, probability and timeliness of remedies. The resulting matrix — causes and their remedies — may be a useful guide to policymakers to suggest where best returns on expenditure might be found.

Conclusion

It is possible, based on an assessment of the numbers and the remedies, and the review of evaluations immediately preceding, to derive some sense of where the gains may or may not come for each cause and remedy. Where the cause of homelessness is a lack of affordable accommodation, it may be that there are low-cost and fast remedies, and a high probability of success, but these may apply only to a few homeless people. By contrast, where the cause of homelessness is domestic violence, costs may be moderate if the turnaround is swift, and prospects of success with effective programs good, but the prospects for stemming the flow of new entrants may be low. Where the cause is mental-health problems, the gains will depend on the acceptability of the remedy, and the only real prospect for reducing homelessness may be in offering institutional accommodation on a long-term basis. Substance-abuse costs are likely to be high (especially for long-term addictions), probability of success low, and remedies slow. The cost of solving Aboriginal homelessness in remote areas is very high and the probability of success low; prospects may be improved under different remedies that were not tied to land.

This brief review of the major new remedies for homelessness, their cost (where possible), probability of success, timeliness, and the proportion of the homeless population to which they apply, suggests the prospects for reducing homelessness are limited. The few new remedies currently available, and those in prospect, are unlikely to yield large reductions. The greatest cause of homelessness is family breakdown and, as there is no known solution but readily available short-term assistance, it may be time for policymakers to reset their targets towards the efficient turnaround of the short-term needs and sufficient investment in long-term needs.

Regrettably, the Australian Government White Paper does not provide a convincing path to achieve its goal of halving homelessness. It is clear that a strategic analysis is needed, one that disaggregates the components of the homeless market, and reconsiders feasible remedies. Although the exercise on this occasion is but a brief overview, the technique does allow a policymaker to understand the likely gains in reducing homelessness and the likely areas which will be difficult to reduce. Governments may have to accept that efforts to reduce the stock of homeless will do no more than provide quicker short-term turnaround and more long-term places. Efforts to diminish the flow of homeless will depend on market conditions for housing and employment outside of the field of homelessness services. As for the large component of homelessness arising from family breakdown, the remedies are not obvious and may be beyond the scope of government programs.

Above all, the concept of a 'single homeless policy' and a single aim to meet a specific reduction in the numbers of homeless is fraught, and probably obscures the important distinctions in the causes for homelessness among distinct groups of people. Homelessness can be reduced for some, but probably not for all where the boundaries of acceptable policy remain limited. The Government needs a more articulated response to an enduring problem. It should accept that in a country as prosperous as Australia, there is an irreducible level of homelessness.

References

Anderson, P., Rogers N., Stephenson, T. and Hume A. 2003, *It's No Palace — Boarding Houses: the sector, its clientele and its future*, Department of Human Services South Australia.

Australian Bureau of Statistics 2008, *Counting the Homeless 2006*, Australian Census Analytic Program.

——— 2011a, 'Methodological Review of Counting the Homeless 2006', Discussion Paper.

——— 2011b, 'Review of Counting the Homeless Methodology', Position Paper.

——— 2011c, *General Social Survey: Summary Results, Australia, 2010*.

Australian Institute of Health and Welfare 2008, 'Homeless people in SAAP', *SAAP National Data Collection Annual Report 2006–07 Australia*.

——— 2009, *Australia's Welfare 2009*.

——— 2010, *Demand for government-funded specialist homelessness accommodation 2008–09*, a report from the SAAP National Data Collection.

——— 2011, *Housing Assistance in Australia 2011*.

Australian Housing and Urban Research Institute 2003, 'Tax Concessions and Subsidies for Australian Homebuyers and Home Owners', *Research and Policy Bulletin* 27.

——— 2006, 'Youth Homelessness in Rural Australia', *Research and Policy Bulletin* 82.

——— 2011, 'How Do Planning Requirements Impact Housing Costs and the Development Process?' *Research and Policy Bulletin* 125.

Australian Human Rights Commission 1989, *Our Homeless Children*.

——— 1993, *The Human Rights of People with Mental Illness*.

Colebatch, H. K. 2006, 'Accounting for Public Policy in Australia', *Public Policy* 1(1).

Council of Australian Governments 2010, *National Plan to Reduce Violence Against Women and their Children*.

Department of Families, Housing, Community Services and Indigenous Affairs 2003, 'I'm looking at the future', *Evaluation Report of Reconnect*.

——— 2008a, *The Road Home: Homelessness White Paper*.

——— 2008b, *Women, Domestic and Family Violence and Homelessness: A Synthesis Report*.

——— 2010, *Evaluation of the Child Protection Scheme of Income Management and Voluntary Income Management Measures in Western Australia*.

Dunkerley, S. 2010, 'Abbott Rejects Rudd's Plan for Homeless', *Sydney Morning Herald* February 17.

Flatau, P., Zaretzky, K., Brady, M., Haigh, Y. and Martin, R. 2008, 'The Cost-Effectiveness of Homelessness Programs: A First Assessment', Volume 1 — main report Australian Housing and Urban Research Institute Western Australia.

Galston, W. 2006, 'Political feasibility: interests and power' in Moran, M., Rein, M. and Goodin, R. (eds.), *The Oxford Handbook of Public Policy*, Oxford University Press.

Housing Assistance Act 1996, 2002, *Annual Report 1999–2000*.

Johns, G. 2009, *No Job, No House: An Economic Strategic Approach to Aboriginal Housing*, Menzies Research Centre.

——— 2011, *Aboriginal Self-determination: the Whiteman's Dream*, Connor Court.

Johnson, G. and Chamberlain, C. 2008, 'Homelessness and Substance Abuse: Which Comes First?' *Australian Social Work* 61(4).

Kenway, P. and Palmer, G. 2003, *How many, how much? Single homelessness and the question of numbers and cost*, London: New Policy Institute.

Macklin, J. and Rudd, K. 2008, '$4 Billion to help close the gap for indigenous Australians', media release, 30 November 2008.

Mission Australia 2009, *An Icon for Exit: The Mission Australia Centre 2005–2008*.

National Youth Commission 2008, *Inquiry into Youth Homelessness*.

Newman, L., Biedrzycki, K., Patterson, J. and Baum, F. 2007, *A Rapid Appraisal Case Study of South Australia's Social Inclusion Initiative*, a report prepared for the Social Exclusion Knowledge Network of the Commission on Social Determinants of Health (established by the World Health Organisation).

New South Wales Auditor General 2011, 'Two Ways Together: NSW Aboriginal Affairs Plan', Performance Audit.

New South Wales Ombudsman 2011, *Addressing Aboriginal disadvantage: the need to do things differently*, a Special Report to Parliament under s 31 of the *Ombudsman Act 1974*.

Ozdowski, S. 2001, 'The Human Rights of Mentally Ill People: the HREOC Inquiry and After', Mental Health, Criminal Justice and Corrections Conference, Marrickville, 19 October 2001.

Pinkney, S. and Ewing, S. 2006, *Costs and Pathways of Homelessness: Developing Policy Relevant Economic Analyses for the Australian Homelessness Service System*.

REVIEW ARTICLE

Australia's Defence: A Review of the 'Reviews'

Henry Ergas[1]

The Australian Defence Force is held in high regard; the Department of Defence is not. Longstanding concerns about inefficiency, compounded by a succession of fiascos and bungles, have entrenched the perception that Defence is poorly managed. Earlier attempts at reform have yielded mixed, often disappointing, results (see Ergas and Thomson 2011), and the years since 2009 have seen a series of reviews aimed at improving performance, culminating in 22 defence-related reviews in 2011–12 alone.

Eight recent reviews deal primarily with the fallout from the so-called Skype incident at the Australian Defence Force Academy (ADFA), in which an ADFA cadet allegedly broadcast images of himself and a female cadet engaged in sex. While those reviews are of considerable interest and potentially significant consequence, they will not be discussed here; rather, my focus is on the reviews which go directly to the efficiency with which 'Defence' uses resources.[2]

The most far-ranging reviews are the Audit of the Defence Budget, undertaken by George Pappas with support from McKinsey and Company (Pappas 2009) and the Review of the Defence Accountability Framework, undertaken by Rufus Black (Black 2011). Others, such as the Collins Class Sustainment Review (Coles 2011) and the Plan to Reform Ship Repair and Management (Rizzo 2011), are more narrowly focused.

Despite the range there are some common themes. Central among these are deficiencies in Defence's management systems. Simply put, there are many plans, but no plan; myriad accountabilities, but no accountability. The result is a structure in which decisions are poorly integrated and in which individuals, while they know what they are intended to do, are not responsible for its being done. Moreover, while the structure generates torrents of data, the sheer scale and diffusion of information, and the lack of tight connection between decisions on the one hand and what is measured on the other, further undermine accountability. Coles, for example, notes an instance in which 'a junior officer [was] required to render a very detailed progress report every day which was

1 The University of Wollongong and Deloitte Access Economics; ergas.henry@gmail.com The author is as usual, greatly indebted to Mark Thomson for our many discussions of these issues. Responsibility for the opinions expressed lies entirely with the author.

2 Henceforth 'Defence' is used in the wider sense of both the Department and the ADF.

sent to over 100 recipients', with presumably few reading it, and even fewer feeling any sense of responsibility for the progress it supposedly tracked. Adding to the lack of accountability is the reluctance of senior management, and of the leadership in the services, to use what powers they have to hold individuals to account for poor performance. The unsurprising consequence is a chronic failure to exploit opportunities for improvement, accompanied by periodic instances of acute breakdown.

The Collins Class submarine program: Murphy was an optimist

Nothing better illustrates the chronic problems than the saga of the Collins Class submarines, now in its twenty-fifth year. Astutely analysed in *Lessons from Australia's Collins Submarine Program* (Schank *et al*. 2011), this was a project which showed Mr Murphy, of the eponymous law, to be an incorrigible optimist, at least as far as major defence ventures are concerned.

Yet it would be wrong to blame bad luck for the Collins' difficulties. Rather, from the start, almost everything that could *be done* wrong was *done* wrong. As the Coles review puts it, 'the problems originate from the very beginning of the program when, perhaps without fully appreciating the potential consequences, the Commonwealth embarked on the acquisition of a submarine which, for good reason, is quite unlike any other in the world'. Many years later, Coles concludes, we are still in a situation where 'despite the fact that virtually all senior people we spoke to were clear that the Collins Class capability is "strategic" for Australia, there is no clear or shared public understanding of why this is a strategic capability nor of the implications this has for sustainability.'

As for acute breakdowns, those too have been in abundant supply, with the most recent being in September 2010 when the Chief of Navy imposed an 'operational pause' on the seaworthiness of the 'amphibious landing platform' HMAS Manoora,[3] causing a collapse in Australia's amphibious ship capability. That collapse, which would have made it impossible for the ADF to respond promptly to a major disturbance in our archipelagic region, is thoroughly diagnosed in the review undertaken by Paul Rizzo. The causes, he finds, include 'poor whole-of-life asset management, organisational complexity and blurred accountabilities, inadequate risk management, poor compliance and assurance, a "hollowed-out" Navy engineering function, resource shortages in the System Program Office in DMO [the Defence Materiel Organisation], and a culture that places the short-term operational mission above the need for technical integrity.'

3 Since decommissioned.

Even more disturbingly, the Rizzo review notes that these problems were 'long-standing, well known to Defence and DMO, and the subject of many prior reports'. And there are indeed close parallels between the conclusions of the Rizzo review and those of the 2007 Board of Inquiry into the Sea King accident at Nias Island, which concluded that the tragedy causing the death of nine ADF members during the response to the Indian Ocean tsunami was not an 'isolated random event' but due to 'systemic failings across the ADF and parts of the Defence organization' (Royal Australian Navy 2007).

The Pappas review: Linking weapons systems selection to future military challenges

It is against the backdrop of those grim assessments that the reviews' recommendations must be seen. Some of those recommendations have been somewhat overtaken by events. This is especially so for the cost-reduction recommendations of the Pappas review, which formed the basis of a major search for savings known as the Strategic Reform Program (SRP), initiated in the 2009 Defence White Paper.

The SRP was somewhat misleadingly cast as aimed at achieving efficiencies. But rather than increases in defence output per unit of output, much of it, in reality, simply involved spending less to produce less. As Thomson (2012) shows, some of the savings targets were more credible than others. Moreover, while some of the targets (notably for sustainment spending) were achieved, others (notably for personnel costs) were not, statements by Defence to the contrary notwithstanding. Be that as it may, the SRP — described in the 2010–11 Budget as 'the Government's key initiative in 2010–11' — was effectively abandoned in the 2012 Budget, to be replaced by essentially arbitrary reductions in allowed outlays.

The Pappas review did, however, also recommend changes to Defence's planning and control framework, including a streamlined planning process and, especially, a tighter integration of strategic planning on the one hand and capability decisions on the other. Achieving such tighter integration between the identification of future military challenges and the selection of major weapons systems has been a central goal of Australian defence reformers since the days of Sir Arthur Tange (Secretary of the Department of Defence from 1970 to 1979).[4] Tange was determined to introduce into Defence's weapons-acquisition process

4 Tange's 1973 report, 'Australian Defence: Report on the Reorganisation of the Defence Group of Departments', laid the basis for the fusion in 1976 of the former service ministries into a single department headed by a 'dyarchy' of a civilian Secretary and a military Chief of the Defence Force (CDF).

the techniques of systems analysis and program budgeting pioneered by Robert McNamara during his tenure as US Secretary of Defense (1961–68). Those techniques were, in Tange's view, crucial to moving away from a 'requirements' approach, in which the services specified the weapons systems they intended to purchase with little or no regard to cost. Rather, Tange insisted, a 'disciplined' framework needed to be applied, which balanced system capabilities with costs and took full account of substitution possibilities between alternative ways of achieving a military objective. While some progress in this direction was made in the Tange years and in the subsequent decade, by the late 1990s those gains had been lost, and capabilities selection had returned to being largely based on 'wish lists' and logrolling between the services.

The Pappas report also highlighted the problems that still arise in the actual acquisition of capabilities, including cost overruns and serious delays. The formation of the Defence Materiel Organisation in 2000 as a separate, specialist acquirer of defence materiel and sustainment services on the services' behalf was intended to put acquisition on a commercial basis. Subsequent reviews, notably by Malcolm Kinnaird in 2003 (Kinnaird 2003) and by David Mortimer in 2008 (Mortimer 2008), brought new disciplines to the acquisition process, including a two-stage review process before major programs proceed. Nonetheless, Pappas shows, cost overruns and delays persist, with the greatest problems being in projects that involve systems especially built for, or extensively customized to, Australian conditions. In the light of those problems, Pappas, echoing an earlier recommendation by Mortimer, recommends that all 'customised' solutions be subjected to rigorous cost-benefit testing compared to a military off-the-shelf (MOTS) alternative.

The Black review: Accountability and contracts

The Black review takes Pappas' emphasis on improved planning and control and pushes it considerably further. While the Black review is littered with deplorable management jargon, at its heart is a sensible attempt to deal with three principal/agent problems that are of central importance to defence efficiency. These are the relations between the government as the principal and Defence as its agent; between the Department as the principal and its staff as the agent; and between each of the military services as the principal and the support services of the Department and DMO as their agent. In each of these, Black proposes ways to clarify expectations, improve the measurement of performance and enhance the incentives for success.

Specifically, Black notes that while Defence has myriad plans, and a complex, multi-tier planning process, it lacks a corporate plan. Rather, it has a Defence

Management Finance Plan, which, in practice, is an amalgam of a range of stand-alone plans, often prepared by individual fiefdoms within Defence, with targets for the current year and projections four to 10 years ahead. In contrast, Black proposes an overarching annual corporate plan that would set out succinctly the guidance for the Department, including the (preferably quantified) targets it would achieve. This would both allow better monitoring of Defence by the government and Parliament, and allow improved communications of priorities within Defence itself.

Such a plan, Black suggests, would be naturally linked to the objectives set for individual units within Defence, and hence would inform a cascading series of targets and associated performance indicators, reaching down to individuals. It would, in this way, help address principal/agent issues within Defence, though Black notes three additional prerequisites for enhanced accountability. First, the current structure of management by committee, which diffuses responsibility and excuses poor performance, must be replaced by a system in which major decisions are associated with a clearly identified individual decision-maker. Second, the practice of frequent rotation, which results in uniformed personnel shifting jobs every two years or so, needs to be replaced by an arrangement which provides sufficient stability that performance can be assessed, accompanied by opportunities to develop and exploit specialist skills in areas such as project costing and delivery. Last but not least, there must be real rewards and penalties associated with individual performance, though Black emphasises (and, in my view, somewhat exaggerates) the constraints imposed by the *Public Service Act 1999* and by the regulations made under the *Defence Act 1903* (which apply to members of the ADF).

Finally, with respect to issues between the three services and the procurement and support arms of the Department, Black proposes substantial reforms to the existing system of performance agreements. While these generally include targets and associated metrics, Black finds that the metrics rarely allow timely and effective monitoring of outcomes. Additionally, Black notes that the agreements neither include unit-cost information nor allow for trade-offs between unit cost and capability. Black recommends that the agreements be structured in a way that both informs and permits such trade-offs, presumably along with some mechanism for the sharing of gains. This would help move these agreements closer to customer/contractor relationships in the commercial world.

There is much to commend in these recommendations. But there are also inherent limitations. Black, for example, greatly overestimates the power of 'contract like' mechanisms within what remains a centrally planned economy, with no scope for contestability, few tools for price discovery, and far weaker incentives to seek gains than to avoid losses. Moreover, long experience in Defence itself shows that developing the kinds of contracts Black recommends is as costly and time-consuming as it is technically challenging, though that is not to say a move in that direction could not yield net gains.[5]

The greatest weakness, however, lies in the failure of the reviews to address this question: how it is that the difficulties have proven so persistent in the face of determined reform efforts. There is, in these reviews, a startling absence of historical perspective: each does a good job of addressing its remit; but none asks why previous, no less competent, reviews did not succeed.

The most natural response to this question would be in terms of the phenomenon Sigmund Freud, in his famous essay on 'Beyond the Pleasure Principle', termed 'repetition compulsion': the tendency in neurosis to repeat behaviour, even when it repeatedly leads to failure.

In the mind of the neurotic, Freud stressed, the mere fact of repetition gives behaviour a degree of legitimacy, an element of predictability and routine, which helps control the external world and is preferable to the pain and stress of breaking out of the cycle of failure. Like individuals, public-sector institutions suffer from behavioural pathologies — pathologies made all the more durable by the fact that, unlike firms in the private sector, those institutions are not subject to displacement by more effective competitors. Moreover, the fact that success and failure are so difficult to measure, and that assessments can be so readily fudged, only exacerbates the problems. The inherent tendency of bureaucracies to indulge in 'group think', to suppress the awareness of difficulties, to convert attitudes into hard and fast organisational cultures, and to treat dissent as disloyalty, then helps ensure that what has failed once will fail again and again.

There are certainly elements of this in Defence, perhaps to an even higher degree than in other areas of public administration. But Defence's difficulties in recent years stem less from those pathologies than from the evolution it has undergone.

'Things fall apart; the centre cannot hold'

Without trying to fill in the details, a sketch of the relevant story could run along the following lines.

Subsequent to the Tange reforms of the 1970s, Defence entered into a long period of retrenchment, in which defence spending grew very slowly, if at all. Indeed, as Thomson (2012) points out, the defence budget, measured in real

5 Writing commercial contracts for the relevant services will usually be very difficult. Yet it remains desirable to try to travel in that direction, as it forces Defence to try as hard as it can to specify mutual expectations. But the journey is, perhaps, more valuable than the destination.

terms, was greater in 1985–86 than it was a decade later. The result was a force whose capabilities were extremely limited, as became painfully apparent in the Timor intervention of 1999–2000. That intervention acted as a wake-up call to the Howard government, which began a process of increased spending that accelerated with the terrorist attacks of 11 September 2001. The ADF expanded by about 20 per cent; defence spending in real terms grew by over 3 per cent annually, so that the defence budget (again, in real terms) increased from $17 billion at the end of the 1990s to $27 billion in 2009–10; and a slew of new, extremely sophisticated equipment was brought into service.

All this allowed substantial increases in capability. Indeed, it is difficult to believe the ADF could have sustained the range and tempo of operations in which it has since been involved without the post-1999 expansion. Yet those increases in capability were extremely uneven, with far greater improvements in the Army than in the other services. And the management processes in Defence itself failed to keep up with the greater scale on which Defence operates, the greater complexity and cost of the platforms it operates, and the greater demands of accountability being placed on government departments generally. Overall, the centre did not prove strong enough to control the greatly scaled-up parts, leading to a complete mismatch between promises and achievements.

The extensive reorganisation and outsourcing of tasks undertaken in the Defence Reform Process (DRP) of 1997–99 contributed materially to the severity of the problems that mismatch eventually created. The DRP had three consequences. First, it led to a significant loss in technical capability, as too little attention was paid to ensuring sufficient skills were retained within the ADF for it to fulfil its core responsibilities. Second, the DRP introduced a corporate model based on the centralised provision of core services, but without the processes needed to manage shared services in the highly complex matrix structure of Defence. And third, the DRP failed to put in place the management accounting systems that would allow cost control and could provide a basis for determining and monitoring outcomes.

What might otherwise have been mere weaknesses thereby became crippling structural flaws. An unduly ambitious procurement program added to the difficulties, as commitments to new platforms exceeded the capacity of DMO to manage, of industry to deliver and, ultimately, of the government to fund. To this unhappy mix, the election of the Rudd government brought new woes, as that government, with its expansive vision of Australia's role, was as willing to make bold commitments to ends as it was reluctant (and ultimately unable) to commit the requisite means. That gap, most manifest in the wake of the 2009 White Paper (Department of Defence 2009), led to an ever greater tension between goals and resources, which transformed long-running problems into crises. As these broke out, the government responded through a sequence of reform programs that any organisation would have found difficult to absorb and implement, much less one with as weak a corporate centre as Defence.

And that is the nub of the problem. Lasting reform is impossible unless it greatly strengthens Defence's corporate centre, in terms of its ability to challenge the services — for instance, on capability decisions — develop meaningful plans, and monitor and enforce outcomes. That was always a problem with the Department: unlike McNamara, who had greatly strengthened the Office of the Secretary of Defense (OSD), Tange managed through sheer intelligence and force of personality. The result was that while the US OSD was able to force the services to substantially lift their game, its Australian counterpart was far too weak to do so.

As a result, even by the 1980s, it was apparent that the corporate headquarters function in Defence was unable to achieve Tange's goal of imposing 'a disciplined relationship between strategy and force structure within the constraints of what is financially feasible' (Tange 2008: 22). And here too the DRP, in its quest to avoid duplication, made matters worse by removing that element of review and criticism of the services' capability wish lists that civilian analysts had previously provided.

But none of the proposals advanced in the reviews tackle head-on the need to significantly strengthen Defence's corporate centre, nor has the government shown any appreciation of its importance. It is therefore difficult to see how the sensible outcomes of these reviews could be successfully implemented in their substance; rather, merely being adhered to in their form (as has happened to so many reform proposals in the past).

There is every risk that the proposals advanced in these reviews will go that way. That would be a pity, for they reflect a mass of careful, considered work. But then again, so did their predecessors.

References

Black, R. 2011, *Review of the Defence Accountability Framework*, Department of Defence, Canberra, January.

Coles, J. 2011, *Collins Class Sustainment Review*, Department of Defence, November.

Department of Defence 2009, *Defending Australia in the Asia Pacific Century: Force 2030*, Defence White Paper, Canberra.

Ergas, H. and Thomson, M. 2011, 'More Guns Without Less Butter: Improving Australian Defence Efficiency', *Agenda* 18(3).

Freud, S. 1922, *Beyond the Pleasure Principle*, in The Standard Edition of the *Complete Psychological Works of Sigmund Freud*, W.W. Norton & Company, Inc, New York, NY.

Kinnaird, M. 2003, *Defence Procurement Review*, Department of Defence, Canberra.

Mortimer, D. 2008, *Going to the Next Level: The Report of the Defence Procurement and Sustainment Review*, Defence Materiel Organisation, Canberra.

Pappas, G. with McKinsey & Company 2009, *2008 Audit of the Defence Budget*, Department of Defence, Canberra.

Rizzo, P. J. 2011, *Plan to Reform Support Ship Repair and Management Practices*, Department of Defence, July.

Royal Australian Navy 2007, *Nias Island Sea King Accident: Board of Inquiry Report*, Canberra.

Schank, J. F., Ip, C., Kamarck, K. N., Murphy, R. E., Arena, M. V., Lacroix, F. W. and Lee, G. T. 2011, *Learning from Experience Vol IV: Lessons from Australia's Collins Submarine Program*, RAND Corporation (Prepared for Australia's Department of Defence), Santa Monica, CA.

Tange, A., 1973 *Australian Defence: Report on the Organisation of the Defence Group of Departments*, Australian Government Publishing Service, Canberra.

——— 2008, *Defence Policy-Making: A Close-Up View, 1950–1980*, Australian National University E Press, Canberra.

Thomson, M. 2012, *The Cost of Defence: ASPI Defence Budget Brief 2012–2013*, Australian Strategic Policy Institute, Canberra, May.

REVIEWS

Nicholas Wapshott, *Keynes Hayek. The Clash That Defined Modern Economics* (Scribe, 2011)

Reviewed by Selwyn Cornish[1]

Opinion about the role that government should play in the macro economy tends to conform to a rhythm, an ebb and flow, determined by circumstances and influenced by ideas. This book examines the ideas of John Maynard Keynes and Friedrich Hayek from the 1920s to the present and concludes that interest in their theories and policy proposals waxed and waned according to changes in economic conditions. It is at once a study of the development of modern macroeconomic theory and policy, and a review of competing political philosophies. The author is a Reuters contributing columnist and former senior editor at *The Times* of London. He writes with an engaging style and adopts a light touch, calculated no doubt to attract a wide audience. The book, to be sure, is a good read. But historians of economic thought — and economists in general — will be disappointed perhaps with the superficial nature of much of the argument. Yet the book may assist the general reader to understand the evolution of economic ideas, as long as it is recognised that substantial liberties have been taken to simplify the argument.

The author's intention may have been to write a book about Keynes. But there are now so many books about him, and about his economics, that it is difficult to conjure up new ways of presenting his ideas. Following the method adopted by the Greek essayist Plutarch, who wrote about parallel lives, Wapshott has come up with the idea of discussing the progression of Keynes's ideas by comparing them with those of one of his chief rivals, Friedrich Hayek. The discussion focuses in particular on the high point of their rivalry in the early 1930s. Keynes's major works in economics are first discussed and Wapshott then contrasts Keynes's conclusions with Hayek's ideas on monetary economics and capital theory. While Keynes's call for government expenditure on public works to alleviate unemployment began as early as 1924, it originally lacked an analytical foundation. As unemployment in Britain grew, Keynes spent the next 10 years trying to develop a theory to justify government expenditure on public works. On the other hand, Hayek's economics was influenced by the work of his Austrian forebears and by the hyper-inflation in Austria and adjoining countries

1 The Australian National University; Selwyn.Cornish@anu.edu.au

in the early 1920s. This led him to conclude that government intervention in the economy led inevitably to distortions in the underlying economic structure and to inflation.

In this battle of ideas and policies, Wapshott directs particular attention to the marked divisions in the 1930s between the economists in Cambridge led by Keynes and his associates, including Richard Kahn, Piero Sraffa and Joan Robinson, and Hayek's supporters at the London School of Economics, where his great champion was Professor Lionel Robbins. In fact, Wapshott asserts that Robbins brought Hayek to London from Vienna for the express purpose of halting the spread of Keynes's influence on contemporary economic thought and policy.

The highlight of the book for many readers will be those sections that cover the exchanges between the rival parties in the early 1930s. They pulled no punches when criticising the economics of the other side. But notwithstanding the ferocious attacks on each other's work, Keynes and Hayek were able to maintain a high personal regard for each other. Keynes, for example, went out of his way to take care of Hayek and his family when they were forced to vacate London during the war and move to Cambridge. Hayek, for his part, had no hesitation in adjudging that Keynes was 'the one really great man I ever knew'. This ability to retain a high personal regard for each other while conducting a heated public debate against each other's ideas and proposals, says a great deal about the two men; it recalls the earlier debate between two other great rivals, namely David Ricardo and Thomas Robert Malthus.

While discussion of their ideas and the intellectual and personal exchanges between Keynes and Hayek and their associates absorb a significant portion of the book, the central theme is the fluctuating fortunes of their ideas and related policy positions. Thus Keynes seemed initially to win the battle, dominating economic thought and policy from the 1930s to the early 1970s. In defeat, Hayek moved from economics to political philosophy, beginning with *The Road to Serfdom* and continuing with *The Constitution of Liberty*. These publications slowly created a new interest in Hayek, which accelerated with the stagflation of the 1970s and early 1980s. Now Hayek's early writing in economics won new adherents, and in 1974 he was the joint winner with Gunnar Myrdal of the Nobel Memorial Prize in Economic Science. The eclipse of Keynes, and the resurgence of Hayek, lasted until the Global Financial Crisis in 2008. In later chapters, Wapshott records the swing back to Keynes, and then gives brief attention to the subsequent questioning of the validity of Keynesian policies as sovereign-debt crises materialised and excessive monetary easing rekindled fears about the recurrence of inflation.

As an organising conception, the differences between Keynes and Hayek might seem to be a promising device to capture the general reader's attention. The problem is that such a framework leaves open the possibility of over-simplification. An example is the stress that is placed on the idea that Keynes's economics failed when stagflation arrived in the early 1970s, since it was thought incapable of handling the co-existence of inflation and unemployment. But one only has to read Chapter 21 of the *General Theory of Employment, Interest and Money* to know that Keynes well understood the possibility of inflation occurring when the economy operated at less than full employment, which for him meant an unemployment rate of 4–5 per cent. The idea that inflation and unemployment were necessarily incompatible was taken from the so-called Phillips Curve. The apparent trade-off between unemployment and inflation, however, has nothing to do with Keynes's economics, but was rather a concoction formulated by Samuelson and Solow decades after the publication of the *General Theory*. As to policy, there was nothing 'Keynesian' about the desire of American administrations in the second half of the 1960s and early 1970s to boost government expenditure and allow budget deficits to accelerate at a time of full employment and rapid inflation. Nor was there anything 'Keynesian' about the decisions of Australian governments in the early and mid 1970s permitting the continuation of a grossly overvalued exchange rate and massive increases in government expenditure when the foreign reserves quadrupled within the space of two years and unemployment was at near record lows and inflation at peacetime highs. Here are to be found the origins of the stagflation in the 1970s, rather than the work of Keynes. Further, there is nothing 'Keynesian' about the excessive stimulus measures in 2008 and 2009 when the Australian economy was scarcely touched by the GFC, or about the continuation of these measures into 2010 and the perpetuation of budget deficits in 2011 and 2012. Rather than being inspired by Keynes, all these developments were the result of political expediency and should be seen in that light.

What then was the essence of 'the clash' between Keynes and Hayek that 'defined modern economics'? Whereas Hayek believed that attempts to subvert the market mechanism, though they might be undertaken with the best of intentions, would lead inevitability to inefficiency, unemployed resources and ultimately to the loss of freedom, Keynes believed that the market system had a tendency to operate at below capacity and required fiscal and monetary measures to boost aggregate demand.

Keynes had two major problems with Hayek; one to do with the latter's economics, the other with his political philosophy. Like members of the 'classical school', Hayek's economics was based on the implicit assumption of full employment. Perhaps the most important point that Keynes made in the *General Theory* was that when the economy operated at less than full employment — and for him, this was the general case — there were spare resources that could be drawn upon by public authorities without having to divert resources from existing employment. In this sense, the structural distortions between capital-goods and consumer-goods industries that Hayek argued would occur when governments spent on public works, or when the monetary authorities allowed excessive monetary easing, would not occur. At full employment, Keynes conceded that the classical analysis was valid; in those circumstances, market forces should be allowed to do their job, governments should scale back their expenditures, and central banks should tighten monetary policy. While Arthur Pigou, for example, had been signing the same letters to newspapers as Keynes, calling as they did for increased government expenditure to stimulate activity, Keynes criticised Pigou's continued adherence to the economics of the classical school because it presupposed that there could not be any unemployment. As Keynes wrote in the Preface to the *General Theory*: '[I]f orthodox economics is at fault, the error is to be found not in the superstructure, which has been erected with great care for logical consistency, but in a lack of clearness and of generality in the premises' (Keynes 1936). He had made the same point in 1931 when criticising Hayek, saying that his analysis 'is an extraordinary example of how, starting with a mistake, a remorseless logician can end up in Bedlam' (Moggridge 1973).

The other issue that troubled Keynes — and many others, including Frank Knight and Jacob Viner — was an inconsistency that lay at the heart of *The Road to Serfdom*. Keynes saw great merit in the book's attack on excessive government intervention. Writing to Hayek in June 1944, he considered *The Road to Serfdom* to be 'a grand book. We have the greatest reason to be grateful to you for saying so well what needs to be said…morally and philosophically I find myself in agreement with virtually the whole of it; and not only in agreement with it, but in a deeply moved agreement' (Moggridge 1980). For many, however, this praise by Keynes for Hayek's most popular work has often been difficult to square with their differences on matters of economic theory and policy. But that was not the point, for when it came to politics it is often forgotten that Keynes and Hayek had much in common: both were liberals, and both saw a limited role for government in economic and social affairs. Unlike Marx, for example, the objective of both Keynes and Hayek was to preserve capitalism; that aim, according to Keynes, required the government to stabilise the economy, while Hayek supported government intervention in various areas of social policy (as did Adam Smith). Of *The Road to Serfdom*, Keynes told Hayek that his 'only serious criticism of the book' was 'the question of knowing where to draw the line' (on government intervention):

> You agree that the line has to be drawn somewhere, and that the logical extreme is not possible. But you give us no guidance whatever as to where to draw it. In a sense, this is shirking the practical issue. It is true

> that you and I would probably draw it in different places…But as soon as you admit that the extreme is not possible, and that a line has to be drawn, you are, on your own argument, done for, since you are trying to persuade us that so soon as one moves an inch in the planned direction you are necessarily launched on the slippery path which will lead you in due course over the precipice. (Moggridge 1980)

In 1971, after reviewing the debates of the 1930s, which to some degree he had initiated, Lionel Robbins — now Lord Robbins — wrote that the 'historian of the future, if he wishes to treat of the relations between London and Cambridge during this period, should be warned that any generalizations that he may wish to make must fit facts of considerable complexity if they are not seriously to misrepresent the situation' (Robbins 1971). Regrettably, Wapshott failed to heed this advice, since he has over-generalised and ignored the complexities involved. As to his part in the debates of the 1930s, Robbins later confessed that he 'was certainly an anti-expansionist where public expenditure was concerned, at a time when, as I now think, I should have been on the other side'. The problem was that the prevailing 'theory was inadequate to the facts'; he was prepared to admit 'my dispute with Keynes as the greatest mistake of my professional career'. Like Hayek, Robbins agreed that he 'would certainly regard Maynard Keynes as the most remarkable man I have ever met' (Robbins 1971).

One final point: there is an inordinate number of factual and bibliographical mistakes in this book. In the early 1980s the distinguished English novelist Graham Greene wrote a letter to *The Times* in which he took Wapshott to task for manifold errors in an article Wapshott had written for the paper. Rather than 'Mr Wapshott', Greene suggested, he should be described as 'Mr Badshott' (Hawtree 1989). It is a pity that Wapshott did not learn from his encounter with Greene.

References

Hawtree, Christopher 1989, *Graham Greene. Yours Etc. Letters to the Press 1945–89*, Reinhardt Press.

Keynes, John Maynard 1936, *The General Theory of Employment, Interest and Money*, Macmillan.

Moggridge, Donald 1973, *The Collected Writings of John Maynard Keynes: XIII, The General Theory and After, Part 1, Preparation*, Macmillan.

——— 1980, *The Collected Writings of John Maynard Keynes: XXVII, Activities 1940–1946, Shaping the Post-War World, Employment and Commodities*, Macmillan.

Robbins, Lord 1971, *Autobiography of an Economist*, Macmillan.